

2020

## Community-Driven Metadata Standards for Agricultural Microbiome Research

J. P. Dundore-Arias  
*California State University, Monterey Bay*

E. A. Eloë-Fadrosh  
*DOE Joint Genome Institute*

L. M. Schriml  
*University of Maryland School of Medicine*

Gwyn A. Beattie  
*Iowa State University, gbeattie@iastate.edu*

F. P. Brennan  
*Teagasc*

*See next page for additional authors*

Follow this and additional works at: [https://lib.dr.iastate.edu/plantpath\\_pubs](https://lib.dr.iastate.edu/plantpath_pubs)



Part of the [Agricultural Science Commons](#), [Agriculture Commons](#), [Cataloging and Metadata Commons](#), [Environmental Microbiology and Microbial Ecology Commons](#), and the [Genetics and Genomics Commons](#)

The complete bibliographic information for this item can be found at [https://lib.dr.iastate.edu/plantpath\\_pubs/287](https://lib.dr.iastate.edu/plantpath_pubs/287). For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

---

This Article is brought to you for free and open access by the Plant Pathology and Microbiology at Iowa State University Digital Repository. It has been accepted for inclusion in Plant Pathology and Microbiology Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

---

# Community-Driven Metadata Standards for Agricultural Microbiome Research

## Abstract

Accelerating the pace of microbiome science to enhance crop productivity and agroecosystem health will require transdisciplinary studies, comparisons among datasets, and synthetic analyses of research from diverse crop management contexts. However, despite the widespread availability of crop-associated microbiome data, variation in field sampling and laboratory processing methodologies, as well as metadata collection and reporting, significantly constrains the potential for integrative and comparative analyses. Here we discuss the need for agriculture-specific metadata standards for microbiome research, and propose a list of “required” and “desirable” metadata categories and ontologies essential to be included in a future minimum information metadata standards checklist for describing agricultural microbiome studies. We begin by briefly reviewing existing metadata standards relevant to agricultural microbiome research, and describe ongoing efforts to enhance the potential for integration of data across research studies. Our goal is not to delineate a fixed list of metadata requirements. Instead, we hope to advance the field by providing a starting point for discussion, and inspire researchers to adopt standardized procedures for collecting and reporting consistent and well-annotated metadata for agricultural microbiome research.

## Keywords

genomics, meta-analysis, metagenomics, microbiome, omics, ontologies, phytobiome, synthetic

## Disciplines

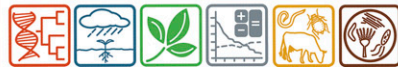
Agricultural Science | Agriculture | Cataloging and Metadata | Environmental Microbiology and Microbial Ecology | Genetics and Genomics

## Comments

This article is published as Dundore-Arias, J. P., Emiley A. Eloë-Fadrosh, Lynn M. Schriml, Gwyn A. Beattie, Fiona P. Brennan, Posy E. Busby, Rosalie B. Calderon et al. "Community-driven Metadata Standards for Agricultural Microbiome Research." *Phytobiomes Journal* (2020). doi: [10.1094/PBIOMES-09-19-0051-P](https://doi.org/10.1094/PBIOMES-09-19-0051-P).

## Authors

J. P. Dundore-Arias, E. A. Eloë-Fadrosh, L. M. Schriml, Gwyn A. Beattie, F. P. Brennan, P. E. Busby, R. B. Calderon, S. C. Castle, J. B. Emerson, S. E. Everhart, K. Eversole, K. E. Frost, J. R. Herr, A. I. Huerta, A. S. Iyer-Pascuzzi, A. K. Kalil, J. E. Leach, J. Leonard, J. E. Maul, B. Prithiviraj, M. Potrykus, N. R. Redekar, J. A. Rojas, K. A. T. Silverstein, D. J. Tomso, S. G. Tringe, B. A. Vinatzer, and L. L. Kinkel



## PERSPECTIVE

## Community-Driven Metadata Standards for Agricultural Microbiome Research

J. P. Dundore-Arias,<sup>1,†</sup> E. A. Elae-Fadros,<sup>2</sup> L. M. Schriml,<sup>3</sup> G. A. Beattie,<sup>4</sup> F. P. Brennan,<sup>5</sup> P. E. Busby,<sup>6</sup> R. B. Calderon,<sup>7</sup> S. C. Castle,<sup>8</sup> J. B. Emerson,<sup>9</sup> S. E. Everhart,<sup>10</sup> K. Eversole,<sup>11</sup> K. E. Frost,<sup>12</sup> J. R. Herr,<sup>13</sup> A. I. Huerta,<sup>14</sup> A. S. Iyer-Pascuzzi,<sup>15</sup> A. K. Kalil,<sup>16</sup> J. E. Leach,<sup>17</sup> J. Leonard,<sup>18</sup> J. E. Maul,<sup>19</sup> B. Prithiviraj,<sup>20</sup> M. Potrykus,<sup>21</sup> N. R. Redekar,<sup>22</sup> J. A. Rojas,<sup>23</sup> K. A. T. Silverstein,<sup>24</sup> D. J. Tomso,<sup>25</sup> S. G. Tringe,<sup>26</sup> B. A. Vinatzer,<sup>27</sup> and L. L. Kinkel<sup>28</sup>

<sup>1</sup> California State University Monterey Bay, Biology and Chemistry, 100 Campus Center, Seaside, CA 93955, U.S.A.

<sup>2</sup> Lawrence Berkeley National Laboratory, DOE Joint Genome Institute, Berkeley, CA, U.S.A.

<sup>3</sup> University of Maryland School of Medicine, Department of Epidemiology and Public Health, Institute for Genome Sciences, Baltimore, MD, U.S.A.

<sup>4</sup> Iowa State University, Department of Plant Pathology, 207 Science I, Ames, IA 50011-3211, U.S.A.

<sup>5</sup> Teagasc, Department of Environment Soils and Land-Use, Wexford, Johnstown Castle, Ireland

<sup>6</sup> Oregon State University, Department of Botany and Plant Pathology, Corvallis, OR, U.S.A.

<sup>7</sup> Louisiana State University, Department of Plant Pathology and Crop Physiology, Baton Rouge, LA, U.S.A.; and Benguet State University, Benguet, La Trinidad, Philippines

<sup>8</sup> U.S. Department of Agriculture-Agriculture Research Service, Plant Science Research Unit, Saint Paul, MN, U.S.A.

<sup>9</sup> University of California Davis, Department of Plant Pathology, Davis, CA, U.S.A.

<sup>10</sup> University of Nebraska, Plant Pathology, 406G PLSH 1875 N 38th St., Lincoln, NE 68510, U.S.A.

<sup>11</sup> International Alliance for Phytobiomes Research, Lee's Summit, MO, U.S.A.

<sup>12</sup> Oregon State University, Botany and Plant Pathology, 2121 S 1st St., Hermiston, OR 97838, U.S.A.

<sup>13</sup> University of Nebraska-Lincoln, Department of Plant Pathology and Center for Plant Science Innovation, 422 Plant Sciences Hall, Lincoln, NE 68588, U.S.A.

<sup>14</sup> North Carolina State University College of Agriculture and Life Sciences, Department of Entomology and Plant Pathology, Raleigh, NC, U.S.A.

<sup>15</sup> Purdue University, Department of Botany and Plant Pathology and Center for Plant Biology, West Lafayette, IN, U.S.A.

<sup>16</sup> Williston Research Extension Center, Williston, ND, U.S.A.

<sup>17</sup> Colorado State University, Bioagricultural Sciences and Pest Management, 1177 Campus Delivery, Fort Collins, CO 80523-1177, U.S.A.

<sup>18</sup> Louisiana State University, AgCenter and Department of Plant Pathology and Crop Physiology, Baton Rouge, LA, U.S.A.

<sup>19</sup> U.S. Department of Agriculture-Agricultural Research Service Administrative and Financial Management, Sustainable Agricultural Systems Laboratory, Beltsville, MD, U.S.A.

<sup>20</sup> Brooklyn College of the City University of New York, Department of Biology, Brooklyn, NY, U.S.A.

<sup>21</sup> Medical University of Gdansk, Department of Environmental Toxicology, Gdansk, Poland

<sup>22</sup> Oregon State University, Crop and Soil Science, 109 Crop Science Building, 3050 SW Campus Way, Corvallis, OR 97331-4501, U.S.A.

<sup>23</sup> University of Arkansas Fayetteville, Plant Pathology, 495 N. Campus Drive PTSC 217, Fayetteville, AR 72703, U.S.A.

<sup>24</sup> Minnesota Supercomputing Institute for Advanced Computational Research, Minneapolis, MN, U.S.A.

<sup>25</sup> AgBiome, 104 TW Alexander Drive Building 1, RTP, NC 27709, U.S.A.

<sup>26</sup> Lawrence Berkeley National Laboratory, DOE Joint Genome Institute, Berkeley, CA, U.S.A.

<sup>27</sup> Virginia Polytechnic Institute and State University, PPWS, Latham Hall AgQuad Lane Room 551, Blacksburg, VA 24061-0131, U.S.A.

<sup>28</sup> University of Minnesota, Plant Pathology, 495 Borlaug Hall 1991 Upper Buford Circle, Saint Paul, MN 55108, U.S.A.

Accepted for publication 6 January 2020.

<sup>†</sup>Corresponding author: J. P. Dundore-Arias; [jdundorearias@csmb.edu](mailto:jdundorearias@csmb.edu)

**Funding:** This material was supported in part by the National Science Foundation under grant number 1714276. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The author(s) declare no conflict of interest.

## ABSTRACT

Accelerating the pace of microbiome science to enhance crop productivity and agroecosystem health will require transdisciplinary studies, comparisons among datasets, and synthetic analyses of research from diverse crop management contexts. However, despite the widespread availability of crop-associated microbiome data, variation in field sampling and laboratory processing methodologies, as well as metadata collection and reporting, significantly constrains the potential for integrative and comparative analyses. Here we discuss the need for agriculture-specific metadata standards for microbiome research, and propose a list of “required” and “desirable” metadata categories and ontologies essential to be included in a future minimum information metadata

standards checklist for describing agricultural microbiome studies. We begin by briefly reviewing existing metadata standards relevant to agricultural microbiome research, and describe ongoing efforts to enhance the potential for integration of data across research studies. Our goal is not to delineate a fixed list of metadata requirements. Instead, we hope to advance the field by providing a starting point for discussion, and inspire researchers to adopt standardized procedures for collecting and reporting consistent and well-annotated metadata for agricultural microbiome research.

*Keywords:* genomics, meta-analysis, metagenomics, microbiome, omics, ontologies, phytobiome, synthetic

Advances in sequencing technologies coupled with declining sequencing costs have resulted in dramatic increases in the volume of sequence-based microbiome data generated for microbial communities across diverse environments. In agricultural and natural habitats, amplicon sequencing, metagenomics, and other omics datasets have revealed the extraordinary diversity of plant and soil microbiomes, and provided insight into variation in microbiome composition and function in relation to crop host, environment, and productivity across diverse settings. Collectively, these data have provided compelling illustrations of the critical roles that microbiomes play in plant and ecosystem health (Bodelier 2011; Finkel et al. 2017; Hooper et al. 2012; Jacoby et al. 2017; Sarkar et al. 2017; Vonaesch et al. 2018).

There is tremendous enthusiasm among agricultural researchers for integrating data and research findings across diverse cropping systems and management settings as a means for identifying generalizable principles of microbiome assembly and dynamics (Busby et al. 2017; Erlich and Narayanan 2014). In order to develop robust predictions of functional agronomic potential, comparative and synthetic (e.g., synthesis or integration of data across experiments or research systems) data analyses are needed. However, despite the widespread and expanding availability of crop microbiome data, variation in sampling, processing, and analytical methodologies, as well as in metadata collection and reporting constrains our capacity to integrate data across research laboratories and projects.

Variation among researchers in methodological techniques and study details has always represented a constraint to straightforward integration of data among research studies. Given the diversity of approaches for sample processing (e.g., DNA and RNA extraction and storage, sequencing technologies, and data processing algorithms) and the significant impacts of these methods on the resulting data, seamless integration of sequencing data across studies without a priori agreement on sampling, processing, and data analysis strategies is both challenging and inefficient. Moreover, attempting to establish a single methodology may in some cases limit the potential for researchers to optimize their design to test their primary hypotheses. However, the potential for comparative (rather than integrative) analyses across studies that may have variable processing methods is still possible with the establishment and use of consensus terminology and metadata standards in agricultural microbiome research.

Metadata are critical to providing the necessary context and description of the what, where, when, and how of the microbiome sample and play a critical role in enhancing the value of

the sequencing data (Cole et al. 2010; Huttenhower et al. 2014; Jones et al. 2006). Among microbiome scientists, agricultural researchers are especially likely to benefit from standardized metadata reporting a consensus sampling, data processing, and analytical pipelines. In particular, many agricultural researchers focus on a small number of important food crops across a wide variety of geographic locations, habitats, and management conditions, resulting in many varying datasets for the same plant host. Such datasets create opportunities for comparative studies aimed at understanding the potential generalizability of findings, and the robustness of microbiome compositional and functional dynamics in relation to plant performance across diverse production systems. Capturing the knowledge of large-scale processes and underlying trends from our collective agricultural microbiome data will require such comparative analyses and, where possible, integration of existing datasets and research findings from across many laboratories. This will only be possible with standardized metadata. Finally, it is important to note that agricultural researchers are likely to require metadata types that are not commonly included in microbiome datasets, including metrics related to crop productivity, management practices, weather conditions, and cropping history.

Here we advance the discussion of agriculture-specific metadata for microbiome research by proposing a first draft of metadata standards for agricultural plant and soil microbiomes. This contribution comes out of a summer 2018 workshop (Boston, MA) titled “Agricultural Microbiome Data: Data Platforms, Standards, and Analytical Tools Needed for Advancing our Science.” This workshop was co-organized by the National Science Foundation Agricultural Microbiomes Research Coordination Network (AgMicrobiomes RCN; <https://agmicrobiomercn.umn.edu/>) and leaders of both the Genomic Standards Consortium (GSC) (<https://press3.mcs.anl.gov/gensc/>; Yilmaz et al. 2011b) and the National Microbiome Data Collaborative (<https://microbiomedata.org/>). This workshop brought together researchers with expertise in plant pathology, soil science, agronomy, microbial ecology, bioinformatics, and data science to propose a preliminary list of metadata fields toward building metadata standards specific for the agricultural microbiomes research community, and to propose defined controlled vocabulary associated with specific ontologies. The workshop built upon initial efforts arising from the Phytobiomes Initiative of The American Phytopathological Society (<http://www.phytobiomes.org>) and a subsequent 2016 workshop organized by the International Alliance for Phytobiomes Research (<http://www.phytobiomesalliance.org>) and the

U.S. National Institute of Standards and Technology to explore priorities, standards, protocols, reference materials, and reference data for phytobiomes research.

## BUILDING UPON EXISTING METADATA STANDARDS

Agricultural microbiome researchers are not starting from scratch with regard to establishing standards for metadata. Multiple environmental metadata databases have already been established including for example the Genomes OnLine Database (GOLD) (Mukherjee et al. 2017) and the National Center for Biotechnology Information's BioSample Database (<https://www.ncbi.nlm.nih.gov/biosample>). The GSC pioneered and coordinated the generation of community-driven standards for collecting and managing relevant contextual information associated with genomic data (Field et al. 2008). Starting over a decade ago, the GSC community established minimum information (MIxS: minimum information about any [x] sequence) standards for describing sequencing data and the associated sample environment, including defining specific parameters for sample description, as well as the documentation of analytical approaches (Glass et al. 2014; Yilmaz et al. 2011a). The MIxS standards consist of checklists for describing minimum information about marker genes (MIMARKS), genomes (MIGS), and metagenomes (MIMS), and 15 environmental packages to enable the standardized description of environmental and host-associated contextual data specific to distinct environments (e.g., air, soil, water, and sediment), hosts (e.g., humans and plants), and tissues (e.g., human gut, oral cavity, and skin) (Field and Sansone 2006; Taylor et al. 2008). Similarly, the International Human Microbiome Standards project (IHMS) (<http://www.microbiome-standards.org>) was established to promote development and implementation of standard operating procedures, including sample identification, collection, and processing, needed to optimize data quality and comparability in human microbiome research. Of more particular relevance to agricultural researchers, the Terragenome International Soil Metagenome Sequencing Consortium (Vogel et al. 2009) established soil data types and standards for describing soil features relevant to soil biology research. The consortium established controlled vocabulary and simple definitions to help researchers to query and retrieve soil metagenome data using project contextual data (Cole et al. 2010). Through collaboration with the GSC, this vocabulary and definitions were incorporated into the environmental metadata packages within MIxS, including for the MIxS soil- and plant-associated checklist that were subsequently endorsed and promoted by the Earth Microbiome Project (Thompson et al. 2017). More recently, a group of scientists from the U.S. Department of Agriculture-Natural Resources Conservation Service published a list of recommended soil health indicators and associated methods appropriate for high-throughput soil test laboratories (Technical Note No. 450-03; Stott 2019). This information was also adopted by the Soil Health Institute for developing methods for evaluating soil health indicators at a continental scale (<https://soilhealthinstitute.org/north-american-project-to-evaluate-soil-health-measurements>). While genomics is included in their Tier 2 checklist (described as "additional research is needed before users can have the same level of confidence in its measurement, use, and interpretation than Tier 1"), no other microbiome-related indicators have been incorporated today.

To make microbiome data across research studies and data repositories more tractable and easier to interpret, it is paramount to define a controlled vocabulary, or ontology, encompassing the semantics, properties, and relationships of a particular domain (Huss 2014). Standard terminology and clear definitions are of critical importance to reduce the siloed nature of microbiome research, and to

prevent confusion regarding the use and interpretation of critical terms. For example, Huss (2014) noted the potential conflict in the dual usage of the term microbiome, when used with the intention to refer to the collective microbiota genome of a host organism, versus the ecological interpretation of the term as a biome of microbes, as in ecology. In other cases, the use of synonyms or imprecise terms for metadata annotation can obscure the intended description or hinder data retrieval from existing repositories (Huttenhower et al. 2014). In agricultural microbiome research, such challenges are raised when terms are used interchangeably or without providing a specific working definition to refer to the origin of the microbiome of interest. Some common examples are as follows: rhizosphere, rhizoplane, or root-associated; root-endophytes versus root-compartmentalized; and soil, bulk soil, or potting soil.

The Environmental Ontology (EnvO) (Buttigieg et al. 2013) and the EMP Ontology (EMPO) (Thompson et al. 2017) are ontologies used to describe microbiome data based on description of environments linked to biological specimens, samples and observations. Similarly, the Open Biological and Biomedical Ontology (OBO Foundry) (Smith et al. 2007) archives various plant-related ontologies used to describe plant attributes that are relevant to plant science experiments, including plant traits and phenology, as well as agronomic practices. However, despite the existence of multiple sequencing platforms and databases types, the agricultural microbiome research community lacks consensus metadata standards and defined ontologies integrating agricultural and microbiome-specific research principles.

Following the example of the MIxS soil- and MIxS plant-associated checklists, participants in the AgMicrobiomes RCN workshop sought to identify the minimum information and categories to be included in a standard metadata checklist for agricultural microbiome studies (Table 1). Discussion focused on identifying metadata that should be required (essential for the dataset) versus desirable (important, but either more difficult or expensive to measure or irrelevant in some contexts). Participants of the workshop recognized the tension between the costs of collecting, reporting, and storing metadata versus the value of additional metadata for research synthesis, and were challenged to limit required metadata.

Although our efforts are extensions of metadata standards developed for and by other microbiome research communities, our proposed checklist of required and desirable (but optional) metadata standards addresses the unique needs of the agricultural microbiome research community. Our proposed checklist deviates from currently available resources by integrating plant, soil, field, and climate metadata, and incorporating metrics specific to cropping systems, productivity, and management practices. Moreover, our checklist highlights the importance of including categories for sampling protocols, processing, and storage for both environmental samples and sequenced materials.

## TOWARD THE ESTABLISHMENT OF AN AGRICULTURALLY FOCUSED METADATA STANDARD

This manuscript is intended to stimulate discussion and move our community toward standardized reporting of metadata, sampling, processing, and analytical pipelines in agricultural microbiome research. The next step is to develop, along with members of the GSC, a MIxS-Ag metadata standard and ontology that will be incorporated into the GSC MIxS collection and released to other commonly used data management platforms and repositories. Further efforts should also focus on developing mechanisms and incentives that promote the use of these standards within our field.

**TABLE 1**  
**Proposed required and desirable metadata categories to be included in a future minimum information metadata standards checklist for describing agricultural microbiome studies (MixS-Ag)**

Required or Desired	Category	
<b>Site or location context</b>		
Required	Environment/habitat	GPS coordinates or location description
Required	Agricultural field	GPS coordinates or location description
	Urban	GPS coordinates or location description
	Natural	GPS coordinates or location description
	Other	GPS coordinates or location description
	Controlled environment (i.e., laboratory/greenhouse)	GPS coordinates or location description
		Description of conditions (i.e., T, % humidity, day length, substrate)
	Growth substrate (i.e., potting media, rockwool, water)	
Postfield/harvest	GPS coordinates or location description	
Desirable	Other information	
<b>Sample type and description</b>		
Required	Sample description	Sample volume/mass
Required	Sample type	
	( ) Soil- or root-associated	
	( ) Aboveground plant-associated	
<b>Soil-associated sample source</b>		
Required	Soil sample info	Soil type (textural class)
Desirable		Classification (national classification category or World reference)
Desirable		Sample depth
Desirable		Soil horizon
Desirable		Soil physical properties (density/structure/porosity/aggregate size)
Desirable		Soil moisture and temperature at the time of sampling
Required	Sample type	Brief sample description
	( ) Rhizosphere	Sample collection methods
	( ) Rhizoplane	Sample processing (i.e., disinfection/sieving/air- or freeze-drying)
	( ) Bulk soil	
<b>Plant-associated sample source</b>		
Required	Sample type	Brief sample description
	( ) Epiphytic	Sample collection methods
Desirable	( ) Endophytic	Sample processing (i.e., disinfection)
Required	Plant tissue	Plant host species/variety/genome sequencing/disease/health status
		Plant tissue type (i.e., roots/leaf/stems or shoots/fruits/seed or vegetative propagule/flowers, pollen/other)
		Sample description
		Plant growth stage/developmental scale
Desirable	Plant tissue	Tissue chemistry analysis
Desirable	Insect pest/vector	Sample description (genus, species, biotype)
Desirable	Other (water, air, etc.)	Sample description
Desirable	Other information	
<b>Sample collection</b>		
Required	Sampling	Protocol (number/design), collection date and time (brief description)
Required	Collector/project/associated publication	Brief description

*(Continued on next page)*



**TABLE 1** (Continued from previous page)

Required or Desired		Category	
Required	Storage conditions	Transportation (duration and conditions)	
		Preprocessing storage (duration and condition)	
		Postprocessing storage (duration and condition)	
Required	Composite or pooling	Brief description	
Desirable	Other information		
<b>Sample processing</b>			
Required	Environmental sample	Brief sample description (i.e., enrichment protocol)	
Required	Sequenced material	Brief sample description	
Required		DNA extraction and storage	
Required		PCR protocol (i.e., primers used)	
Required		Sequencing methods	
Desirable		Sequencing facility	
Desirable	Data	Cleanup	
Desirable		Brief description of the analysis pipelines (bioinformatics programs and software)	
Desirable	Other information		
<b>Climate</b>			
Desirable	Field average temperature (air/ambient)	Growing season (maximum and minimum), measurement source GPS coordinate	
Required		At the time of sampling	
Desirable	Field average relative humidity (air/ambient)	Growing season (maximum and minimum), measurement source GPS coordinate	
Required		At the time of sampling	
Desirable	Field average rainfall	Growing season (average), measurement source GPS coordinate	
Desirable	Other information		
<b>Management practices history and abiotic characteristics</b>			
Desirable	Crop yield	Current (if available)	
Desirable	Current and previous land use history	Description (agriculture/fallow/forest/pasture/woodland/wetland)	
Desirable	Pesticide applications	Preplanting treatment	Type and dose
Desirable		Seed treatment	Type and dose
Desirable		Soil treatment	Type and dose
Desirable		Foliar application treatment	Type and dose
Desirable		Postharvest treatment	Type and dose
Desirable		Nutrient inputs	Foliar
Desirable		Soil	Type and dose
Desirable	Soil legacy	Tillage	Type/timing/residue
Desirable		Cropping systems	Crop rotation/sequence
Desirable	Soil chemistry	N, P, K	
Desirable		pH	
Desirable		Organic matter	
Desirable		Micronutrients	
Desirable		Dry matter	
Desirable			Other (electrical conductivity / cation exchange capacity)
Desirable	Water inputs	Watering state (irrigated or flooded field), frequency of irrigation	
Desirable	Photosynthetic activity	Measurement (i.e., leaf gas exchange/chlorophyll fluorescence emissions/reflectance/transpiration)	
Desirable	Other information		

With the goal of stimulating discussion around the topic and its content, the MiXS-Ag will be presented at various national and international conferences, offering multiple open Q&A and annotation sessions. During the development of the MiXS-Ag, we will seek feedback and endorsement from agricultural microbiome researchers representing diverse public and private sectors (e.g., the International Alliance for Phytobiomes Research, the Sustainable Innovation of Microbiome Applications in the Food System [SIMBA] project, the Soil Health Partnership, among others). Moreover, the AgMicrobiomes RCN will work with core researchers in our community to produce and deposit MiXS-Ag biosample datasets. Journals publishing agricultural microbiomes research and scientific funding agencies can also support these efforts by requiring metadata compliance with community standards. We have already established connections with editorial boards from relevant scientific journals regarding the establishment of a policy that promotes the use of the MiXS-Ag metadata.

Many factors will influence the further development and evolution of metadata standards for agricultural microbiome research. For example, the rapid rate at which the spectrum of omics technologies and computing and processing infrastructures are advancing suggests the potential need for more thorough documentation of analytical methods and procedures. Similarly, new technologies for collecting environmental and plant phenotypic data in the field are changing the scale, types, and magnitude of data collected, with significant repercussions for metadata collection. Although detailed, highly resolved metadata are desirable, their production and management can be laborious, especially in the absence of automated procedures for incorporation into standardized data formats. Increased granularity of metadata must be coupled with increased abilities to document and export information in machine-readable formats that allow automated data parsing and integration into analytical pipelines for downstream analyses (Cole et al. 2010; Huttenhower et al. 2014). Ultimately, metadata selection principles must focus not only on what we can measure, but also on the biologically relevant scales of space and time for the target microbiome.

An additional challenge to promoting data sharing, transparency, and comparative analyses is in assuring privacy and confidentiality of sensitive grower or landowner information, as well as proprietary industry data. Data protection measures similar to those used to prevent misuse of human subjects' data are needed to guarantee secure and efficient sharing of sensitive data in agricultural research studies, particularly those incorporating data from private grower fields. In addition, documentation of data provenance and intellectual property issues are factors that are likely to be faced by our field in the coming years. Finally, we recognize that a flexible rather than a fixed approach to agricultural microbiomes metadata will be needed over the long term. While standardization of protocols and metadata collection and sharing are vital for ensuring that microbiome data are findable, accessible, interoperable, and reusable (FAIR) (Wilkinson et al. 2016), it is important to ensure that standardized procedures do not constrain the development of innovative or optimized approaches that can advance our field. As new sequencing, analytical, and environmental monitoring technologies evolve, our standards must evolve in parallel.

## CONCLUSION

Here we propose metadata standards for the agricultural microbiome research community. Our goal is not to delineate a strict or fixed list of metadata requirements or defined format for data deposition, but to stimulate conversation to drive the field toward standardization of metadata capture and sharing. The proposed list of metadata standards should serve as a guide and

citable tool to assist researchers in evaluating whether their own data, or that of others, contains key information needed to facilitate comparative and synthetic analyses. We anticipate that the proposed metadata standards will evolve in response to community feedback and advances in our field. Moreover, we expect that these standards will incentivize future efforts for consistent metadata collection, sharing, archiving, and retrieval within our community. Finally, it is our hope that by promoting the use of standard terminology and agricultural system metadata, we will accelerate the potential for synthesis and integration of data across research studies, encourage collaboration among agricultural microbiome researchers, facilitate the identification of generalizable principles of crop–microbiome interactions, and advance our capacities to exploit the functional potential of agricultural microbiomes to improve crop production, sustainability, and nutritional quality.

## ACKNOWLEDGMENTS

We thank The American Phytopathological Society for the support provided in organizing and hosting our workshop during the 2018 International Conference of Plant Pathology. Special thanks go to workshop participants and invited speakers. We also thank Leland Pierson, Ewa Lojkowska, Deborah Samac, Thais Egreja, Paul Schulze-Lefert, Tijana Glavina del Rio, and James Tiedje for their participation in the “Agricultural microbiome data: Data platforms, standards, and analytical tools needed for advancing our science” at the ICCP meeting in Boston, MA. For more information about the AgMicrobiomes RCN, including a complete summary of workshop discussion and outcomes, visit our website <https://agmicrobiomercn.umn.edu>.

## LITERATURE CITED

- Bodelier, P. L. E. 2011. Toward understanding, managing, and protecting microbial ecosystems. *Front. Microbiol.* 2:80.
- Busby, P. E., Soman, C., Wagner, M. R., Friesen, M. L., Kremer, J., Bennett, A., Morsy, M., Eisen, J. A., Leach, J. E., and Dangl, J. L. 2017. Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS Biol.* 15:e2001793.
- Buttigieg, P. L., Morrison, N., Smith, B., Mungall, C. J., and Lewis, S. E., and the ENVO Consortium. 2013. The environment ontology: Contextualizing biological and biomedical entities. *J. Biomed. Semantics* 4:43.
- Cole, J. R., Myrold, D. D., Nakatsu, C. H., Owens, P. R., Kowalchuk, G., Tebbe, C., and Tiedje, J. M. 2010. Development of soil metadata standards for international DNA sequence databases. 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia. <https://press.igsb.anl.gov/terragenome/files/2011/03/Symp232.pdf>
- Erlach, Y., and Narayanan, A. 2014. Routes for breaching and protecting genetic privacy. *Nat. Rev. Genet.* 15:409-421.
- Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., et al. 2008. The minimum information about a genome sequence (MIGS) specification. *Nat. Biotechnol.* 26:541-547.
- Field, D., and Sansone, S.-A. 2006. A special issue on data standards. *OMICS* 10:84-93.
- Finkel, O. M., Castrillo, G., Herrera Paredes, S., González, I. S., and Dangl, J. L. 2017. Understanding and exploiting plant beneficial microbes. *Curr. Opin. Plant Biol.* 38:155-163.
- Glass, E. M., Dribinsky, Y., Yilmaz, P., Levin, H., Van Pelt, R., Wendel, D., Wilke, A., Eisen, J. A., Huse, S., Shipanova, A., Sogin, M., Stajich, J., Knight, R., Meyer, F., and Schriml, L. M. 2014. MiXS-BE: A MiXS extension defining a minimum information standard for sequence data from the built environment. *ISME J.* 8:1-3.
- Hooper, L. V., Littman, D. R., and Macpherson, A. J. 2012. Interactions between the microbiota and the immune system. *Science* 336:1268-1273.
- Huss, J. 2014. Methodology and ontology in microbiome research. *Biol. Theory* 9:392-400.
- Huttenhower, C., Knight, R., Brown, C. T., Caporaso, J. G., Clemente, J. C., Gevers, D., Franzosa, E. A., Kelley, S. T., Knights, D., Ley, R. E., Mahurkar,



- A., Ravel, J., The Scientists for Advancement of Microbiome Research, and White, O. 2014. Advancing the microbiome research community. *Cell* 159: 227-230.
- Jacoby, R., Peukert, M., Succurro, A., Koprivova, A., and Kopriva, S. 2017. The role of soil microorganisms in plant mineral nutrition—current knowledge and future directions. *Front. Plant Sci.* 8:1617.
- Jones, M. B., Schildhauer, M. P., Reichman, O. J., and Bowers, S. 2006. The new bioinformatics: Integrating ecological data from the gene to the biosphere. *Annu. Rev. Ecol. Evol. Syst.* 37:519-544.
- Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Verezemskaja, O., Isbandi, M., Thomas, A., D., Ali, R., Sharma, K., Kyrpides, N. C., and Reddy, T. B. K. 2017. Genomes OnLine Database (GOLD) v.6: Data updates and feature enhancements. *Nucleic Acids Res.* 45:D446-D456.
- Sarkar, A., Saha, M., and Meena, V. S. 2017. Plant beneficial rhizospheric microbes (PBRMs): Prospects for increasing productivity and sustaining the resilience of soil fertility. Pages 3-29 in: *Agriculturally Important Microbes for Sustainable Agriculture*. V. Meena, P. Mishra, J. Bisht, and A. Pattanayak, eds. Springer, Singapore.
- Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., et al. 2007. The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* 25:1251-1255.
- Stott, D. E. 2019. Recommended soil health indicators and associated laboratory procedures. Soil Health Technical Note No. 450-03. U.S. Department of Agriculture, Natural Resources Conservation Service.
- Taylor, C. F., Field, D., Sansone, S.-A., Aerts, J., Apweiler, R., Ashburner, M., et al. 2008. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: The MIBBI project. *Nat. Biotechnol.* 26: 889-896.
- Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., et al. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551:457-463.
- Vogel, T. M., Simonet, P., Jansson, J. K., Hirsch, P. R., Tiedje, J. M., Van Elsas, J. D., Bailey, M. J., Nalin, R., and Philippot, L. 2009. TerraGenome: A consortium for the sequencing of a soil metagenome. *Nat. Rev. Microbiol.* 7:252.
- Vonaesch, P., Anderson, M., and Sansonetti, P. J. 2018. Pathogens, microbiome and the host: Emergence of the ecological Koch's postulates. *FEMS Microbiol. Rev.* 42:273-292.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3:160018.
- Yilmaz, P., Gilbert, J. A., Knight, R., Amaral-Zettler, L., Karsch-Mizrachi, I., Cochrane, G., Nakamura, Y., Sansone, S.-A., Glöckner, F. O., and Field, D. 2011a. The genomic standards consortium: Bringing standards to life for microbial ecology. *ISME J.* 5:1565-1567.
- Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., et al. 2011b. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIXS) specifications. *Nat. Biotechnol.* 29:415-420.