BERKELEY LAB
Bringing Science Solutions to the World

BioSciences

PLANT SINGLE-CELL SOLUTIONS FOR ENERGY AND THE ENVIRONMENT

SECOND WORKSHOP REPORT

Lawrence Berkeley National Laboratory | April 29, 2021

Scientific Organizing Committee

Ben Cole, JGI/LBNL
Diane Dickel, LBNL
Chris Mungall, LBNL
Ronan O'Malley, JGI/LBNL
Axel Visel, JGI/LBNL
Henrik Scheller, LBNL

Strategy and Logistics Organizing Committee

Lauren Jabusch, LBNL
Katy Christiansen, LBNL

Contents

Executive Summary	2
About the Workshop	3
Introduction	3
Workshop Insights	4
Experimental Barriers to Entry and Potential Solutions	5
Data Limited to Transcriptomics; Lack of Multi-omics Platforms with Anatomical Resolution	5
No Methods to Capture Plant Interactions with Fungi or Bacteria	6
Computational Barriers to Entry and Potential Solutions	7
Resources for Effective Analysis of Plant Single-Cell Data Sets Are Costly, Difficult to Use, and Hard to Access	7
Lack of Analysis Methods for Investigating Plant Single-Cell Data Sets	7
Need for Data-reporting Standards for Single-Cell Data Sets	8
Integrative and Organizational Barriers to Entry and Potential Solutions	9
Lack of Forums for Idea Exchange and Collaboration Across Communities	9
Need for Training in both Experimental and Computational Approaches	9
Lack of Agreement on a Set of Common Approaches and Species Relevant to Energy and Environmental Research	10
Increased Coordination Between Federal Granting Agencies	11
References	11
Appendix 1: Workshop Participants	12
Appendix 2: Workshop Agenda	14
Speaker Biographies	15
Scientific Organizing Committee	15
Strategy and Logistics Organizing Committee	15
Appendix 3: Word cloud of workshop generated terms	16

Executive Summary

Plants are important sources of energy and materials, and they collectively represent a critical component of Earth's ecosystem. With increasing environmental stresses due to climate change and intensive agricultural practices, the need for resilient plants is greater than ever before. To secure plant resources for bioenergy, biomaterials, food, and ecosystem adaptation, a deeper understanding of the fundamental biology of plants at a cellular level is urgently needed. Plants contain a multitude of specialized cell types that compose tissues and organs. Pathogens often target specific cell types within plants, and the response of one cell to a particular stimulus is likely to be distinct from its neighbor because of underlying molecular and contextual differences. Understanding how these responses are distributed among cells, the main goal of single-cell approaches, will substantially enhance our ability to use targeted engineering for improving plant productivity and resilience. Furthermore, single-cell approaches are necessary to understand the interactions between plants and other ecosystem members such as fungi, bacteria, and archaea. Unlocking these gene-response mechanisms at a cellular level can improve our ability to adapt plants to environmental stresses, increasing their utility as feedstocks for biomaterials and bioenergy.

Recent advances in high-throughput sequencing, mass spectrometry, microfluidics and miniaturization, artificial intelligence and machine learning, and bioinformatics have greatly improved our ability to detect and understand processes at a cellular level. In mammalian systems, single-cell transcriptomics has already led to many advances, such as newly identified cell types and cell-targeted treatment of diseases, and mass spectrometry-based single-cell proteomics has recently been demonstrated as a promising emerging technology. However, plant single-cell omics has lagged behind mammalian approaches due to the high cost of the technologies relative to available resources and to the innate biological features of plants, including the complexity of the cell wall and polyploidy.

To better understand how single-cell methods could enable plant science, Lawrence Berkeley National Laboratory (Berkeley Lab) hosted a workshop on April 29, 2021, that brought together a diverse group of leaders in plant and/or single-cell biology. Attendees represented federal research programs and domestic and international academic institutions. During the workshop, three presenters described the current state of research in both experimental and computational approaches. While the focus of the workshop was on factors preventing plant biology researchers from fully adopting single-cell methodologies, workshop participants agreed that most barriers could be overcome with focused, strategic investment and coordinated efforts among institutions leading to significant scientific discoveries that would be difficult to obtain using more conventional technologies.

About the Workshop

On April 29, 2021, Berkeley Lab hosted a virtual workshop on Plant Single-Cell Solutions for Energy and the Environment to identify the most pressing barriers to wider adoption of single-cell sequencing and omics technologies, and discuss solutions to remedy those barriers in order to drive discovery. The workshop built upon findings from a prior workshop (**Plant Single-Cell Solutions for Energy and the Environment Workshop 1 Report**) and recent progress at Berkeley Lab in single-cell discoveries in model species and bioenergy crops. The workshop also highlighted research at other institutions. Featured talks were on programmable plant genetic computation to leverage insights from single-cell genomics, the recently published single-cell Arabidopsis root atlas, and single-cell trajectory analysis to find developmental regulators of key plant bioenergy traits. This workshop brought together researchers from the Biosciences Area at Berkeley Lab (Environmental Genomics and Systems Biology Division, DOE Joint Genome Institute, Joint BioEnergy Institute); national laboratory colleagues (Oak Ridge and Pacific Northwest National Laboratories, Center for Advanced Bioenergy and Bioproducts Innovation); and research leaders from domestic and international institutions and universities (Appendix 1A). This workshop was part of a continued effort to establish research directions for the application and development of plant single-cell multi-omics focused on bioenergy crops. Berkeley Lab looks forward to continued engagement and alignment with other national laboratories, academic partners, and private industry to advance this critical area in bioenergy and other important plant platforms.

The workshop focused on barriers and solutions for improving access and use of plant single-cell sequencing. Attendees engaged in three breakout groups that asked participants to brainstorm barriers to wider adoption, prioritize and categorize those barriers, and propose solutions to the most pressing barriers. The findings from these breakout sessions were divided into three main themes described in this report: experimental, computational, and integrative and organizational barriers.

In total, 43 attendees took part in the workshop. This report describes the participants' discussions and targets for new research programs and collaborative approaches. This workshop was funded with internal Biosciences Strategic Program Development funds.

Introduction

Plants provide innumerable ecosystem and economic services, including tree wood for construction, grasses for grain and biomass, and algae for energy and materials feedstocks. When production and harvesting are well managed, biomass from plants can be a renewable and sustainable resource for both materials and energy applications. Current biomass production consumes vast amounts of land, water, and nutrient resources. Widely used modern cultivation practices cannot support projected human population growth while also preserving wild resources that are essential to a healthy ecosystem. Thus, efficiency and resiliency improvements must be made to ensure future sustainable production of plant-based biomass.

The application of genomics has unlocked critical insights into plant biology and ecosystem interactions. Exploring how genomes interact with environmental factors to improve the productivity and stress resilience of key biomass crops can enable advancements in plant breeding, engineering, and ecosystem management. Now, in the omics era of biology, researchers hope to solve pressing issues in plant science with a deeper, comprehensive understanding of RNA (transcriptomics), proteins (proteomics), and metabolites (metabolomics) in plants. The combination of these tools can identify molecular changes in plants resulting from nutrient shortages, temperature fluctuations, water deficiencies, and other environmental stresses associated with climate change and land management. However, most studies are focused on whole plants or heterogeneous tissues. Plants contain many distinct cell types with specialized roles and responses; the response even within one complex tissue typically contains inputs from a host of different cell types carrying out discrete biological functions. While understanding the phenotypic response of the whole plant can yield broad insights, the important actions of individual cells are lost. Understanding the characteristics and response profiles of cell types will provide a more complete and accurate model of how plants behave.

Advances in single-cell studies in mammals have led to important breakthroughs with respect to how the genome, epigenome, and transcriptome orchestrate development, control normal physiology, and undergo dysregulation in disease. As single-cell approaches are becoming more widely adopted in plant research, we are gaining a more comprehensive picture of the identity

and function of the primary gene products and pathways that determine cell type specificity. Furthermore, we are beginning to understand how individual cells contribute to coordinated tissue responses to biotic and abiotic stresses.[1] As more species and conditions are explored, a better understanding of which of these single-cell properties and genetic pathways are shared across plant species can lead to the development of widely applicable strategies for crop improvement. Conversely, the identification of species-specific pathways and cell types can be valuable for developing novel beneficial traits.

Workshop Insights

The following insights reflect the discussions during the workshop breakout sessions and full group report-backs. The extent of the participants' knowledge of and application of single-cell technologies varied. Attendees studied a diverse array of organisms, including bioenergy crops, model plants, agricultural crops, and aquatic species (Figure 1). Others had experience with mouse, fungi, and other eukaryotes. Their experience with single-cell technologies also varied: 76% of attendees had some (novice or expert) experience in plant single-cell experimentation; 62% had some (novice or expert) experience in plant single-cell computation; and 41% had some (novice or expert) experience in other types of single-cell research. Finally, 19% of participants had no experience with single-cell research but wanted to bring this technology to their research programs. From these wide-ranging points of view, priority barriers and potential solutions were identified, along with their connection and benefit to grand science challenges in plant science, environmental research, and bioenergy production.

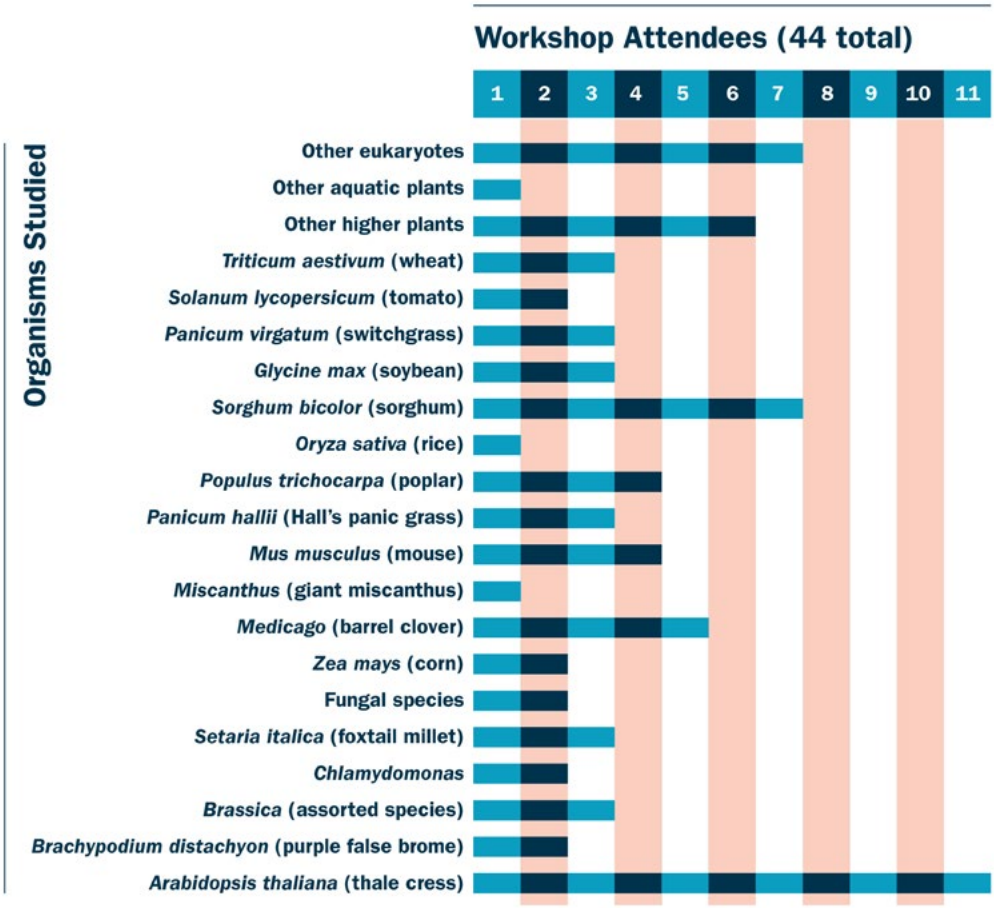


Figure 1. Workshop participants brought expertise in a wide range of eukaryotes, including many terrestrial plant species, mouse, fungi, and aquatic plants.

Experimental Barriers to Entry and Potential Solutions

The experimental approaches from biomedical research have been a starting point for single-cell investigations in other multicellular organisms, but barriers remain for widespread adoption of these technologies in plants. The plant cell wall and the species-specific features of the cell wall remain a challenge in the analysis of individual tissues. Full gene-to-function studies require multi-omic approaches, currently underexplored in plant single-cell research. Finally, understanding the plant as a holobiont — host to an ecosystem of many microorganisms — is currently not possible with the present state of single-cell technologies.

Lack of Streamlined, Accessible, and Universal Protocols for Bioenergy Plant Species

Single-cell genomics approaches have been transformative in the biomedical field, providing cell-level resolution of which genes are expressed and how they contribute to cell-type specificity. The plant biology field has been slower to adopt single-cell approaches, but the first wave of publications in the field provides evidence that these technologies will be equivalently transformative. The delay in adoption is due in part to technical challenges specific to plants. Successful single-cell approaches require the enzymatic removal of the cell wall to release individual cells, or physical isolation of nuclei from cellular debris within complex tissues. Additionally, while animal studies are typically restricted to few organisms (e.g., human, mouse, and *Drosophila*), the plant field has a much wider number of species of interest. These include: experimental models (e.g., *Arabidopsis*, *Brachypodium*, and *Chlamydomonas*); food crops (e.g., maize, rice, and tomato); bioenergy crops (e.g., *Sorghum*, *Poplar*, and *Camelina*); and ecologically-relevant species (e.g. aquatic plants). Because each species has distinct properties that can impact protoplast and nuclei isolation, including cell wall composition and the presence of tissue-associated wax and other coatings, bespoke protocols are more likely to be required than in the animal field. Single-cell approaches have only been reported in a few studies of plants, and only a few protocols are available for researchers who want to enter the field. The lack of specialized protocols for non-model species and diverse tissues means that the advances made in biomedical single-cell omics are currently unattainable for most plant researchers.

Solutions

Due to the high cost of single-cell omics approaches, it is important that reliable methods for cell dissociation and upfront quality control steps are established to avoid issues associated with lysis, degradation of nucleic acids, clumping, and other potential pitfalls associated with cell or nuclei isolation. These barriers will be lowered when more researchers enter the field. As cell quality assessment primarily relies on optimizing extraction protocols (i.e., lysis approaches, extraction buffer conditions, etc.) and fluorescence microscopy, workshops and availability of video-based protocols could provide hands-on or virtual training for those new to the methods. In addition to the practical training, such workshops would also be highly useful in generating a scientific network of researchers in the field.

Relevance to Scientific Grand Challenges

Improved methods to facilitate plant single-cell technologies will unlock deeper understanding of the Department of Energy's (DOE's) flagship plants. Improved methods in single-cell preparation will allow for interrogations of cell type–specific gene functions in a wider array of plants and plant organs.

Data Limited to Transcriptomics; Lack of Multi-omics Platforms with Anatomical Resolution

While most recent plant single-cell omics studies have focused on gene expression (scRNA-seq) and gene expression regulation (scATAC-seq), a host of new approaches are broadening the types of information that can be captured to provide a richer view of cell type-specific molecular properties. These newly developed single-cell molecular modalities include additional epigenomic features such as DNA methylation and chromatin modifications as well as lipidomic, metabolomic, and proteomic profiling. The global characterization of epigenomic, protein, and metabolite properties in plants has been primarily limited to whole-tissue analyses, mainly due to the lack of sensitive and reliable technology platforms coupled with efficient cell-isolation protocols. Furthermore, while understanding

cell-type transcriptomic, epigenomic, and biomolecule properties is valuable, it will be critical to assess all of these properties within the three-dimensional anatomical context of plant organs at a temporal scale during development and night/day cycles. Commercial solutions are already available for profiling transcriptomes in an anatomical context, a technology commonly referred to as spatial transcriptomics. Developing plant-specific spatial transcriptomics experimental protocols, and subsequent development of spatial omics and biomolecule characterization as they become available, will be important for a deep understanding of plant biology. While RNA transcripts are used as a surrogate for proteins, it is widely recognized that RNA abundance cannot precisely predict its corresponding protein abundance. The direct measurement of proteins is limited by inefficient sample preparation and the sensitivity of mass spectrometry.

Solutions

Recent advancements in metabolomic and proteomic profiling of individual or small numbers of cells in plant tissues using mass spectrometry indicate the power of such technologies in identifying cell type-specific and spatially localized proteins and metabolites with important biological functions. These should be further developed to increase sensitivity and throughput, and should be paired with novel spatial transcriptomics capabilities. Adoption of spatial and other anatomically resolved molecular analysis in plants will require plant- and tissue-specific methods. Workshops and video-based protocols would be invaluable to researchers as they learn methods in this new field, as well as helping to develop a scientific network of collaborators.

Relevance to Scientific Grand Challenges

The research enabled by anatomically resolved multi-omics and platform development will lead to advances in characterizing the genotype and phenotype of individual cells. Improved throughput will lead to improved translation from the plant molecular to cellular realm using tools in genomics, transcriptomics, proteomics, and metabolomics.

No Methods to Capture Plant Interactions with Fungi or Bacteria

Understanding the interactions between plants and microbes, including beneficial bacteria and fungi in the rhizosphere, is one of the scientific community's and the DOE's grand challenges. Single-cell technologies are particularly relevant for such studies, as microbes may only interact directly with a limited number of cells within a plant. Hence, neighboring cells are likely to have widely different transcriptomic and physiological profiles that, in turn, influence microbial physiology. However, current single-cell methods are generally not able to capture the transcriptome of both eukaryotic and prokaryotic cells. Spatially resolved transcriptomics can potentially address some of these questions, but the methods are in their infancy, have poor spatial resolution, are very expensive, and have not been tested using relevant plant tissues.

Solutions

To overcome this barrier it will be necessary to directly support research and development of omics methods that can be used simultaneously on eukaryotic and prokaryotic cells, and to help in building networks of researchers in the field. Specifically, single-cell profiling of cross-kingdom interactions will require close collaboration among research groups with diverse technical and domain expertise, and could benefit from strategic partnerships with industry groups that currently support biomedical applications exclusively.

Relevance to Scientific Grand Challenges

With the development of methods such as spatially resolved single-cell transcriptomics, a deeper understanding of the biological complexity of plant and microbial metabolism and interfaces across scales spanning molecules to ecosystems could be understood.

Computational Barriers to Entry and Potential Solutions

Data sets generated using single-cell approaches tend to be very large and have unique properties that require innovation in computational biology. These computational advancements have already made mammalian single-cell studies able to link genes to functions, but plant-specific needs remain. In this section, we identify three barriers and paired solutions to address the computational needs of plant studies.

Resources for Effective Analysis of Plant Single-Cell Data Sets Are Costly, Difficult to Use, and Hard to Access

Single-cell technologies produce a dizzying amount of data with unique properties (i.e., high dimensionality and sparsity), with some applications producing several terabytes-worth of data after a single experiment. In order to be useful, these data need to be securely stored, publicly accessible, and usable by the research community. While there are numerous storage solutions for raw single-cell data (e.g., the National Center for Biotechnology Information Sequence Read Archive), these platforms often do not support, or do not enforce storage of processed data in standard formats. In addition, because of the high dimensionality of single-cell data sets, analysis pipelines to process and interpret single-cell data are memory-intensive and usually necessitate the use of high-performance computing (HPC). The nature of HPC also complicates tasks such as data normalization, filtering, clustering, annotation, and trajectory inference, which require extensive iteration. The ever-increasing array of analysis tools often requires expert computational knowledge, precluding a large number of biologists from directly benefiting from these data sets. These tools also need to be maintained as they gain new features or become more efficient. Lastly, the expense associated with the type of computing required for analysis of single-cell data prevents the large-scale accessibility of the technology.

Solutions

A publically accessible, full-featured computing platform that can handle single-cell data from many species (e.g., DOE flagship plants) is needed to surpass this barrier. This platform would be integrated with other existing plant informatics tools. For example, connections to both the DOE Systems Biology Knowledgebase (KBase) and Joint Genome Institute's (JGI's) Phytozome portal would enable reproducibility of their analysis and access to comparative genomics tools with single-cell data. This infrastructure would support some of the unique features of plants and other environmental organisms (e.g., high ploidy and complex orthology relationships between organisms), with a backend having enough memory/storage to run standard analysis tools in a user-friendly interface. In addition, the infrastructure would continuously maintain and update existing analysis tools and onboard new tools as they become available. The cyberinfrastructure would also allow users to prototype cutting-edge analytical methods through a sandbox-type environment. The computing platform would be co-located with the raw data sets, avoiding the need for costly downloads to run analyses locally.

Relevance to Scientific Grand Challenges

The development of this resource would improve researchers' ability to assemble capabilities for processing large, complex, and heterogeneous systems biology data into open-access analysis platforms addressing grand challenges. This resource would also represent a next-generation data system that could host algorithms needed for large-scale systems biology data science connecting observations across scales and modalities using omics data relevant to cellular and multicellular processes.

Lack of Analysis Methods for Investigating Plant Single-Cell Data Sets

The plant lineage has accumulated many fascinating features while adapting to diverse habitats over millions of years of evolution. However, most tools for computational analysis of single-cell data sets—including cross-species integration and comparative transcriptomics—were developed with mammals in mind, which have much closer evolutionary distances. As such, these tools fall short when confronted with the complex genomic features

of plants and fungi (e.g., polyploid genomes and the presence of both plastid and mitochondrial organelles). Multimodal and multi-species integration is an active area of research even for the more well-studied mammalian model systems. These efforts are made substantially more challenging in plants due to the lack of high-quality data sets from multiple species and across different modalities (e.g., spatial transcriptomics, chromatin accessibility, and single-cell protein abundance data sets), as well as the absence of conserved, cell type-specific marker genes that span the plant lineage. There is a real need to both develop these foundational data sets, as well as implement cutting-edge algorithms that can integrate them, while taking into consideration the unique features of plants.

Solution

Analysis tools to examine and integrate plant single-cell data sets need to be developed or adapted. Specifically, tools are needed that can generalize homology relationships and are ploidy-aware, challenges not yet encountered for studies in animal systems. For cross-species integration, many well-established tools require one-to-one mapping between homologs, which does not work well in plants; recent tools are more flexible but have not been tested using plant data.[1] These methods should also be integrated with genomics resources within Phytozome and KBase. Reads mapping to the mitochondrial genome is a standard quality-control metric for most single-cell transcriptomics analysis methods in animals. Unfortunately, assemblies of chloroplast and mitochondrial genomes are not currently standard practice for plant and fungal species. These organelle genomes should be immediately sequenced or assembled from existing data for all plant species with available nuclear genomes, especially DOE flagship plants. Standards need to be established when determining overall single-cell data quality from plants and fungi, as the commonly used metrics developed for mammals (e.g., mitochondrial transcript abundance for scRNA-seq data) may not be as applicable to other organisms. This will necessitate large-scale experimentation on diverse plant species with meticulous recording of a variety of QC metrics, in addition to an unbiased investigation into transcriptomic markers associated with poor data quality. As more data from different species becomes available, establishing a standard set of cell type-specific markers across plants will substantially simplify the annotation and analysis of new single-cell data sets, especially by non-specialist researchers.

Relevance to Scientific Grand Challenges

New suites of tools specifically engineered for single-cell analysis of plants and fungi would address grand challenges, including the creation of algorithms needed for large-scale systems biology data science that connects observations across scales. In addition, these new tools would implement artificial intelligence algorithms to identify relationships among different parts of genomes and build integrated biological models that capture higher-order complexity of the interactions among cellular components that lead to phenotypic differences.

Need for Data-reporting Standards for Single-Cell Data Sets

There are currently no established standards for reporting single-cell data, outside of minimal raw data deposition requirements by most academic journals. Similarly, there are no standards regarding how data processing (e.g., data normalization and filtering) is reported. Even this basic standard is not well-tailored for single-cell transcriptomics and chromatin accessibility data, which has specific features that must be captured for data reusability (e.g., a cell barcode read and a genomic read). These special features are not always deposited, or they are deposited in unconventional ways. This impedes scientific reproducibility, reanalysis of single-cell data, and data synthesis across studies. In particular, single-cell metadata (including, but not limited to: cell type or cell cluster identifier for all cells in a scRNA-seq dataset; developmental age/state of cells; growth and treatment conditions; and other sample preparation information) is typically not collected or made available. Lastly, high-quality genome assemblies and annotations are often lacking for non-model organisms, especially for cultivars that could diverge significantly from the sequenced genotype.

Solution

Community standards need to be implemented in a way that ensures FAIR (findable, accessible, interoperable, and reproducible) data principles for all new single-cell data sets, and they should be applied retroactively to all currently published data sets. This can be accomplished by establishing a central curation team for plant/fungal/algal single-cell data (e.g., a resource suggested by solutions to the first computational barrier), supporting a community-based curation effort and cross-collaboration

with similar groups (e.g., the Plant Cell Atlas), and/or through building consensus among journals for the type and format of raw and metadata reporting. In addition, ongoing efforts by the JGI and others to sequence and annotate genomes of important environmental and agricultural plant and fungal species needs continued support. Experience from other standardization efforts, such as the National Microbiome Data Collaborative (NMDC), can be leveraged to drive effective standards for reporting metadata about single-cell data and inferred new cell types. Although the domain is different, NMDC has experience in building the kinds of standards required here. The opportunity to establish these standards is now, before waves of new, dissimilar data are generated and have to be retroactively amended.

Relevance to Scientific Grand Challenges

Established standards and high-quality genome resources would further research goals to assemble capabilities for processing large, complex, and heterogeneous systems biology data into open-access analysis platforms addressing DOE missions. They also may be foundational to the assembly of an integrated systems biology virtual laboratory to accelerate *in silico* ideation and collaboration within the research community. These community standards would also provide a platform that enables integrated analysis of data generated across DOE Biological and Environmental Research (BER) program user facilities and capabilities.

Integrative and Organizational Barriers to Entry and Potential Solutions

The interdisciplinary nature of plant single-cell research is an opportunity to remove silos from information, expertise, and coordination. We see potential for cross-discipline, cross-institutional idea sharing, both at the research level and with federal leadership.

Lack of Forums for Idea Exchange and Collaboration Across Communities

Workshop participants identified the lack of dedicated forums for the exchange of scientific and technological advances as a major barrier for progress of single-cell applications for plants. In particular, both experimental and computational investigators currently find it difficult to identify and approach suitable collaborators with complementary skills.

Solutions

The organization of meetings, symposia, and other forums for idea exchange and opportunities for interactions will be critical to facilitate the emergence of a more coherent community around plant single-cell omics. These might be dedicated meetings or dedicated sessions occurring within existing meetings relevant to energy and environmental genomics, such as the annual Plant and Animal Genomes meeting, the annual JGI-organized "Genomics of Energy & Environment" meeting, or the BER-organized "Genomic Science Program Annual Principal Investigator Meeting." An important goal of such community gatherings will be the collective identification of grand challenges that require extensive collaboration across experts in relevant plant biology and single-cell technologies. This type of community building is also a current major focus area of the emerging Plant Cell Atlas initiative, which will offer additional avenues for community engagement.

Relevance to Scientific Grand Challenges

These activities will establish scientific connections for multidisciplinary researchers and create a framework for progress towards solving the most pressing and complex questions in energy and environmental research. It will also guide the development of joint collaborative efforts among DOE user facilities whose capabilities complement current research efforts.

Need for Training in both Experimental and Computational Approaches

Single-cell approaches allow unprecedented resolution in measuring biological activities. Because the approaches are relatively new, the experimental procedures are complicated and constantly evolving. Similarly, while there is rapid growth in the number and complexity of computational tools for analyzing single-cell data, practical knowledge of computational workflows and best practices is available only in a small number of

laboratories. As a result, the general research community, although interested in utilizing these approaches, find it exceedingly difficult to kick-start single-cell studies.

Solutions

To remedy these barriers, leading labs providing single-cell solutions in the experimental and/or computational realms, particularly those with existing DOE BER research efforts, should be identified and invited to serve as a core expert group for the community. These core experts should receive investment to support their community engagements. The core expert group can spearhead the following three activities. First, the core experts will establish experimental and/or computational workshops on single-cell technologies. Such workshops on experimental procedures should feature an overview of the approaches (including discussion of their limitations), theoretical considerations of major steps, and hands-on components for users to handle and process biological materials. The computational workshops should also provide an overview of standard workflows, the computational and quantitative underpinnings of key algorithms/tools, and a hands-on session on analyzing existing single-cell data sets from initial preprocessing steps to downstream analyses, providing solutions to most common problems. Second, step-by-step tutorials and protocols for both experimental and computational approaches should be developed for a wider audience. Users with a background in basic molecular and cellular biology and/or basic bioinformatic analysis should be able to follow the protocols. The workshops will serve as a testbed for the tutorials and protocols to evaluate their effectiveness and troubleshoot any issues. Beyond written tutorials and protocols, it will be particularly useful to share them as videos where experts demonstrate the steps visually. These videos should be distributed through venues such as the *Journal of Visualized Experiments*, major video-sharing sites such as YouTube, or web portals of relevant DOE user facilities (discussed below). The entire computational workflow can be provided as a virtual machine or a Docker image so users can run the entire environment without software installation and version issues, facilitating reproducibility. The data used in the computational tutorial should also be readily available from public repositories. Third, we advocate for the establishment of resources to match experimental and computational experts in single-cell research, initially the core expert group, to share expertise or collaborate with interested users with little or no experience. One potential way to establish such resources is through a user facility that is geared toward or already generating single-cell data. Such user facilities can provide support on both experimental and computational approaches, as well as hosting workshops, training resources, and a portal for collaboration.

Relevance to Scientific Grand Challenges

These proposed solutions can increase access to cutting-edge single-cell technologies that are beyond the expertise and scope of most individual investigators and create new capabilities at existing BER user facilities and among existing collaborative research efforts.

Lack of Agreement on a Set of Common Approaches and Species Relevant to Energy and Environmental Research

Within the nascent field of plant single-cell science, consensus has not yet emerged regarding which species or assay types to focus on. The resulting set of largely unconnected and incompatible data sets currently available in the literature represents a major hurdle for data integration and comparative interpretation. Furthermore, the species investigated by single-cell technologies so far have not focused on species that are relevant for energy and environmental purposes.

Solutions

In recognition of these barriers, there is a major opportunity for DOE science programs and user facilities to guide the community towards generating unified data sets, focusing on a common set of relevant species and strains (such as DOE flagship plants), using standardized approaches, and reporting data and metadata in standardized formats to enable downstream analyses. The identification of suitable target species will require careful consideration of input from the community and stakeholders. Mechanisms for guiding the community may include: relevant focus areas in calls by user facilities; development and sharing of protocols optimized for these species offered by user facilities; creation of data resources and analysis systems specifically for species of interest (e.g., dedicated portals in the JGI-operated Phytozome system for plant comparative genomics); and investment opportunities focused on plants of high importance for bioenergy.

Relevance to Scientific Grand Challenges

Overcoming this barrier will enable the creation of deep and multifaceted layers of data for a limited set of plant species to enable understanding the molecular and genetic underpinnings of plant traits relevant for energy and environment (the "flagship" approach). It will also expand our understanding of the functions of genes, genomic regulatory networks, and metabolites in plants and microbes—particularly those that enable crop adaptation to extreme conditions, changing environments, and episodic environmental events.

References

[1] Cole, B., Bergmann, D., Blaby-Haas, C.E. et al. Plant single-cell solutions for energy and the environment. *Commun Biol* **4**, 962 (2021). **https://doi.org/10.1038/s42003-021-02477-4**

[2] U.S. DOE. 2021. *Biological Systems Science Division Strategic Plan*, DOE/SC-0205, U.S. Department of Energy Office of Science (**genomicscience.energy.gov/2021bssdstrategicplan/**).

Increased Coordination Between Federal Granting Agencies

The development of integrated resources for plant single-cell omics can be accelerated through coordination across funding agencies to generate field-wide data repositories and comparative and integrative analysis platforms. When efforts are coordinated, the resources and data systems generated will be less likely to result in siloed, scattered, and non-integratable data sets. Coordination can foster opportunities to maximize the conversion of data into knowledge that advances plant biology research.

Solutions

Coordinated efforts and funding mechanisms jointly supported by relevant agencies (e.g., DOE, USDA, NSF) would enable the creation of resources that will be broadly used, while avoiding duplicative efforts. There is successful precedent for such joint-funding initiatives: for example, the jointly supported DOE-USDA Biomass Genomics Research Program. The synergies emerging from such programs would maximize the impact of federal investment in plant genomics and accelerate progress of the field, with impactful advances benefiting all participating agencies.

Relevance to Scientific Grand Challenges

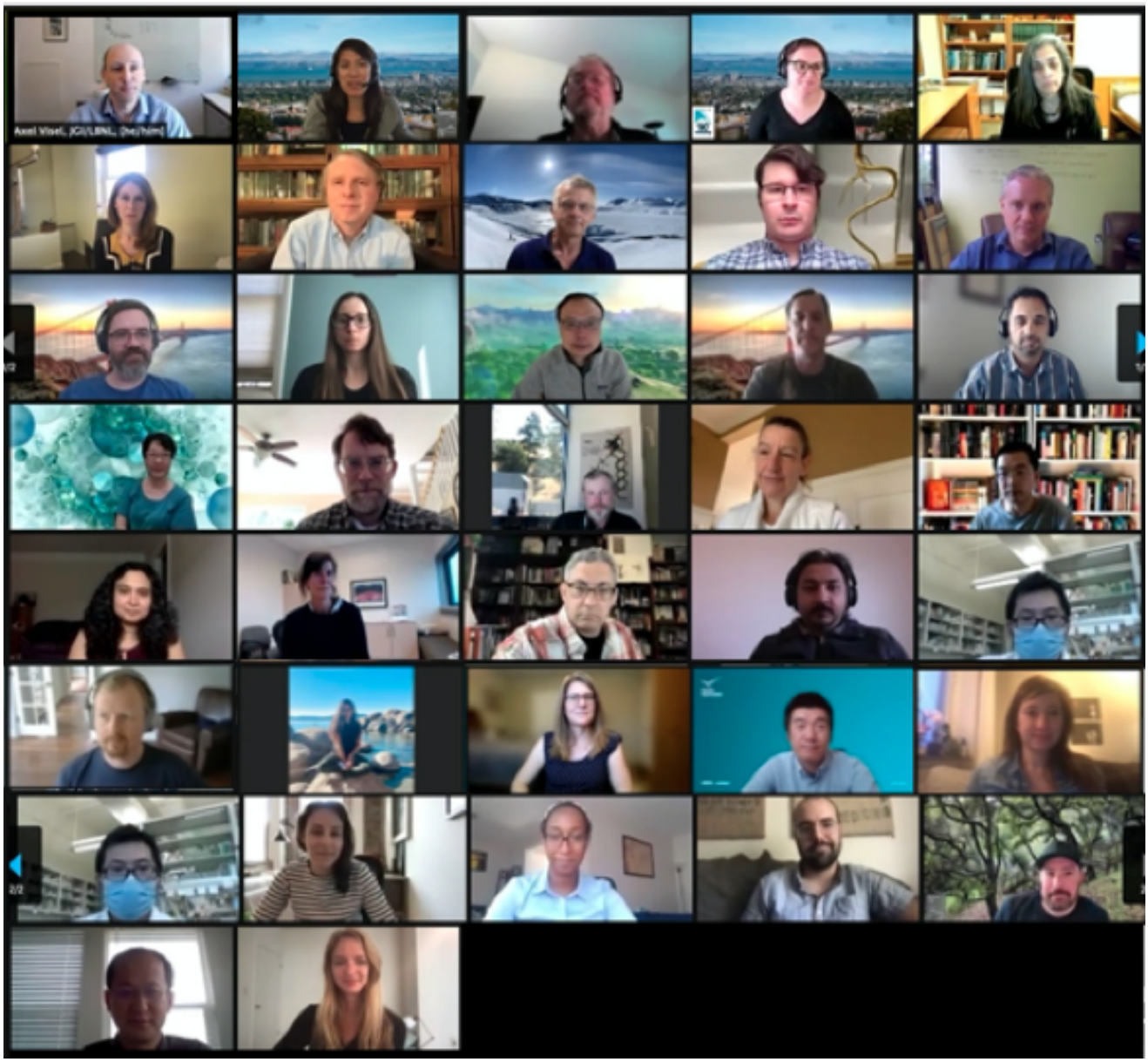
Successful implementation of joint funding mechanisms has strong potential for participating agencies to guide community leaders in the development of data production and analysis standards and to enable seamless exchange and analysis of data and creation of knowledge.

Appendix 1: Workshop Participants

A) Table of Workshop Participants

Institution	Workshop Participants
Center for Advanced Bioenergy and Bioproducts Innovation (CABBI), University of Illinois	Andrew Leakey
Cornell, Boyce Thompson Institute	Maria Harrison
Duke University	Rachel Shahan
Hudson Alpha, Center for Advanced Bioenergy and Bioproducts Innovation (CABBI)	Kankshita Swaminathan
Hudson Alpha, Joint Genome Institute	Jeremy Schmutz
Joint Genome Institute	David Goodstein Jonelle Basso Li Lei Margot Bezrutczyk Rex Malmstrom Ronan O'Malley
Joint Genome Institute, Lawrence Berkeley National Lab	Axel Visel Ben Cole Len Pennacchio
Joint Genome Institute/University of California, Berkeley	Dan Rokhsar Karen Serrano
Lawrence Berkeley National Lab	Chris Mungall Diane Dickel Henrik Scheller John Vogel Katy Christiansen Lauren Jabusch
Michigan State University	Shin-han Shiu Federica Brandizzi
Pacific Northwest National Lab	Amir Ahkami
	Ying Zhu
SciLifeLab, KTH Royal Institute of Technology (Sweden)	Stefania Giacomello
Stanford University	Jose Dinneny
Carnegie Institution for Science	
	Seung Yon (Sue) Rhee
University of California, Berkeley	Sabeeha Merchant Sunnyjoy Dupuis
University of California, Davis	Alex Canto-Pastor Gitta Coaker Jie Zhu
University of Florida	Matias Kirst
University of Massachusetts, Amherst	Sam Hazen
University of Minnesota	Kathleen Greenham

Institution	Workshop Participants
University of Missouri	Gary Stacey
University of Tennessee	Scott Lenaghan
University of Texas, Austin	Renhou Wang
University of Washington	Hardik P. Gala
University of Western Australia	Ryan Lister
University of Wisconsin-Madison	Jean-Michael Ane
Virginia Tech	Song Li



Photography Credit

Composite of screen captures of the video call roster by Lauren Jabusch

Appendix 2: Workshop Agenda

Plant Single-Cell Solutions for Energy and the Environment Workshop 2

April 29, 2021
8:00 am–12:15 pm

8:00–8:10 am	**Welcome: Diane Dickel**
8:10–8:20 am	**Workshop context: Ben Cole**
8:20–8:35 am	**Science Talk: Ryan Lister** "Development of programmable genetic computation in plants to leverage insights from single–cell genomics"
8:35–8:50 am	**Science Talk: Rachel Shahan** "A single–cell Arabidopsis root atlas reveals developmental trajectories in wild type and cell identity mutants"
8:50–9:05 am	**Science Talk: Matias Kirst** "Single–cell trajectory analysis uncovers developmental regulators of key plant bioenergy traits"
9:05–9:15 am	Break
9:15–9:30 am	Introduction to Miro and practice exercise
9:30–10:00 am	Discussion 1: Brainstorming of barriers to wider adoption
10:00–10:15 am	Discussion 2: Prioritization and categorization of the barriers to wider adoption
10:15–10:20 am	Stretch Break
10:20–10:30 am	Synthesis and report back
10:30–11:15 am	Discussion 3: Solutions to top barriers to wider adoption
11:15–11:25 am	Break
11:25–11:55 am	Synthesis and report back
11:55 am–12:05 pm	Community Engagement: future involvement
12:05–12:15 pm	Closing

Speaker Biographies

Ryan Lister

Ryan Lister is a Professor in the School of Molecular Sciences, ARC Centre for Plant Energy Biology at The University of Western Australia. He received his PhD also from The University of Western Australia and a postdoctoral fellowship at the Salk Institute for Biological Studies. He has a wide-range of projects in plant and mammalian genomics, epigenomics and synthetic biology which heavily leverage single-cell approaches to drive these research directions.

Rachel Shahan

Rachel Shahan is a Ruth L. Kirschstein postdoctoral fellow in the lab of Dr. Philip Benfey at Duke University. Using the Arabidopsis root as a model, she is employing single-cell genomics approaches to study gene regulatory networks underlying cell differentiation and organogenesis. Prior to joining the Benfey lab in 2018, Rachel completed her thesis work in the lab of Dr. Zhongchi Liu at the University of Maryland. Her dissertation focused on the molecular regulation of fruit development in Fragaria vesca, a diploid strawberry.

Matias Kirst

Matias Kirst is a Professor in the School of Forest Resources and Conservation and Co-Director of the Cooperative Forest Genetics Research Program at the University of Florida. Prior to starting his faculty position, Dr. Kirst received his PhD from NC State and carried out postdoc work in maize genomics at Cornell. His lab carries out a variety of research in plant genomics and breeding, including using genomics and single-cell transcriptomics to characterize and improve bioenergy traits.

Technology and logistics questions can be directed to Lauren Jabusch (**lkjabusch@lbl.gov**)

Prior workshop report

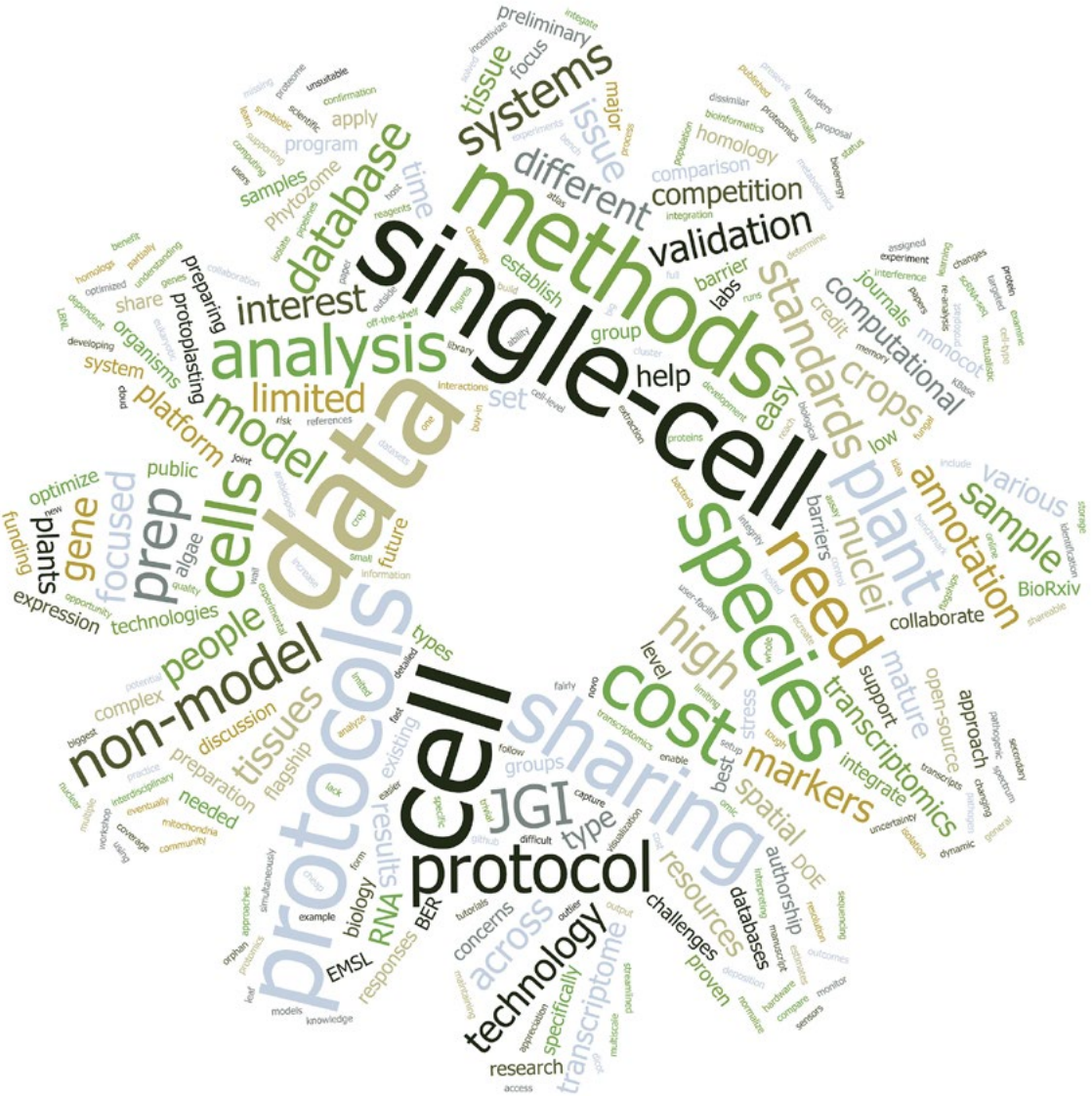
Scientific Organizing Committee

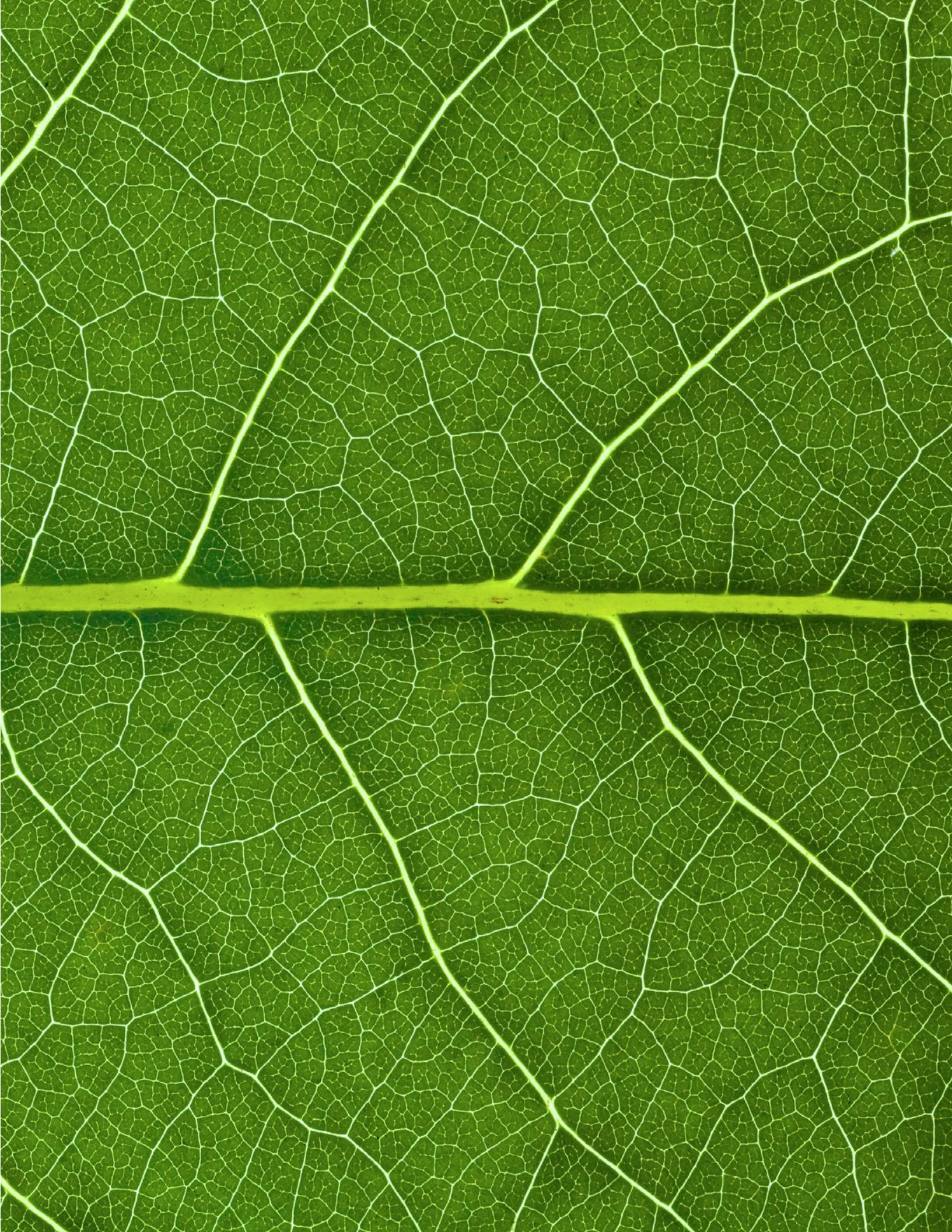
Ben Cole, JGI/LBNL
Diane Dickel, LBNL
Chris Mungall, LBNL
Ronan O'Malley, JGI/LBNL
Axel Visel, JGI/LBNL
Henrik Scheller, LBNL

Strategy and Logistics Organizing Committee

Lauren Jabusch, LBNL
Katy Christiansen, LBNL

Appendix 3: Word cloud of workshop generated terms


single-cell
methods
data
cell
protocol
protocols
sharing
cost
species
need
plant
standards
systems
analysis
cells
non-model
technology
JGI
markers
transcriptomics
transcriptome
resources
tissues
prep
model
database
interest
validation
competition
different
tissue
issue
preliminary
crops
computational
annotation
sample
various
BioRxiv
collaborate
high
mature
open-source
approach
spatial
DOE
authorship
databases
challenges
proven
concerns
across
specifically
research
RNA
EMSL
BER
responses
biology
results
existing
types
level
stress
best
integrate
support
people
technologies
expression
plants
gene
focused
optimize
public
platform
limited
organisms
system
share
preparing
protoplasting
Phytozome
samples
program
apply
time
establish
barrier
labs
group
set
help
easy
low
nuclei
barriers
future
algae
funding
flagship
needed
preparation
discussion
complex
groups
mitochondria
community
workshop
interdisciplinary
multiscale
coverage
using
knowledge
models
orphan
approaches
simultaneously
example
tutorials
output
gather
difficult
fairly
novo
transcriptomics
enable
setup
trough
omic
capture
specific
fast
detailed
limited
analyze
easier
lack
bigger
potential
practice
eventually
nuclear
multiple
mapping
proteome
unsuitable
confirmation
scientific
symbiotic
learn
supporting
computing
users
genes
genome
isolate
reagents
host
atlas
challenge
build
off-the-shelf
library
stability
cluster
cell-level
interactions
risk
references
cloud
joint
crop
small
information
experimental
quality
opportunity
new
benefit
partially
understanding
optimized
homologs
dependent
developing
LBNL
collaboration
papers
disciplinary
multiscale
approaches
orphan
proteomics
leaf
models
knowledge



BERKELEY LAB
Bringing Science Solutions to the World

JGI
JOINT GENOME INSTITUTE

EGSB
Environmental Genomics & Systems Biology

BioSciences

A U.S. Department of Energy National Laboratory. Managed by the University of California.

21-EGSB-3055