Title

Life on the leaf: Seasonal activities of the phyllosphere microbiome of perennial crops

Authors

Adina C. Howe[1,2,3], Nejc Stopnisek[4,5,6], Shane K. Dooley[1], Fan Yang[1], Keara L. Grady[4,5], Ashley Shade[4,5,6,7,8]*

Affiliations

1. Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA 50011
2. Department of Bioinformatics and Computational Biology, Iowa State University, Ames, IA 50011
3. Center for Advanced Bioenergy and Bioproducts Innovation, Ames IA 50011
4. The Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI 48824
5. Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824
6. The Plant Resilience Institute, Michigan State University, East Lansing MI 48824
7. Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing MI 48824
8. Program in Ecology, Evolution, and Behavior, Michigan State University, East Lansing MI 48824

Abstract

Plants and microorganisms form beneficial associations. Understanding plant-microbe interactions will inform microbiome management to enhance crop productivity and resilience to stress. Here, we apply a genome-centric approach to identify key leaf microbiome members on field-grown switchgrass and miscanthus, and quantify their activities for switchgrass over two growing seasons. We integrate metagenome and metatranscriptome sequencing from 192 leaf samples collected over key time points in crop phenology. We curated 40 focal metagenome-assembled-genomes (MAGs) and conservatively focus analysis on transcript recruitment to medium and high-quality MAGs that were <10% contaminated and >50% complete. Classes represented by these MAGs (Actinomycetia, Alpha- and Gamma-Proteobacteria, and Bacteroidota) were active and had seasonal dynamics in key functions, including enrichments in transcripts for of short chain dehydrogenase, molybdopterin oxioreductase, and polyketide cyclase in the late season. The majority of MAGs had activated stress-associated pathways, including trehalose metabolism, indole acetic acid degradation, betaine biosynthesis, and reactive oxygen species degradation, suggesting direct engagement with the host environment. We also detected seasonally activated biosynthetic pathways for terpenes (carotenoid and isoprenoids), and for various non-ribosomal peptide pathways that were poorly annotated. Overall, this study overcame laboratory and bioinformatic challenges associated with field-based leaf metatranscriptome analysis to inform potential key activities of these phyllosphere populations. These activities collectively support that leaf-associated bacterial populations are seasonally dynamic, responsive to host cues and interactively engage in feedbacks with the plant.

47

48 **Keywords**

49 Metagenome, metatranscriptome, leaf, bacteria, archaea, bioenergy, biofuel feedstocks,

50 phytobiome, switchgrass, miscanthus, time-series, ROS, IAA, isoprene

51

52 **Introduction**

53 Perennial plants are a crucial target for the sustainable development of biofuels
54 (Robertson *et al.*, 2017; Hestrin *et al.*, 2021; Ma *et al.*, 2021). In addition to yielding high
55 biomass that can be converted to biofuels and bioproducts, perennial crops offer a broad range
56 of ecosystem services that support efforts to mediate climate change, including greenhouse gas
57 mitigation and promotion of soil nutrient cycling (Heaton *et al.*, 2008; Langholtz *et al.*, 2016;
58 Robertson *et al.*, 2017; Roley *et al.*, 2018). Like all plants, perennials harbor diverse microbiota,
59 and many of these microbes are either known or expected to benefit their hosts. For example,
60 plant-associated microbes can increase productivity and protect against environmental
61 stressors. Because of the intimate engagement of many plant-associated microbiota with the
62 host, management of the plant microbiome is one tool proposed to promote crop vigor and
63 support crop resilience to global climate changes (Busby *et al.*, 2017; Toju *et al.*, 2018; Haskett
64 *et al.*, 2020; Wang and Haney, 2020). Therefore, along with selective breeding and data-
65 informed field management, regulating the plant microbiome is expected to be key for the
66 sustainable production of perennial biofuel feedstocks.

67 Plants have anatomical compartments that each are inhabited by distinctive microbial
68 consortia. Generally, the diversity and composition of the plant microbiome narrows from
69 external compartments to internal, and the plant plays an active role in filtering the
70 microbiome composition inward (Hacquard *et al.*, 2015; Hardoim *et al.*, 2015; Gopal and Gupta,
71 2016). External plant compartments include the root zone, rhizosphere and rhizoplane below
72 ground, and the epiphytic phyllosphere above ground (Andrews and Harris, 2000). External
73 compartments have relatively higher representation of transient or commensal microbial taxa,

74 and these compartments engage with and recruit microbes from the immediate environment.
75 Internal compartments include the endosphere of above- and below-ground tissues, and these
76 have relatively low richness and harbor the most selected microbiota (Bulgarelli *et al.*, 2013;
77 Müller *et al.*, 2016). Of these compartments, the rhizosphere has received the most attention
78 as a key site of microbial-plant interactions that are important for nutrient and water
79 acquisition (e.g., Kuzyakov and Razavi, 2019). However, members of the microbiota that
80 inhabit the phyllosphere also provide important plant functions, such as pathogen exclusion
81 and immune priming (Bell *et al.*, 2019; Chen *et al.*, 2020). Phyllosphere microorganisms have
82 specialized adaptations to their exposed lifestyle (Lindow and Brandl, 2003; Vorholt, 2012;
83 Müller *et al.*, 2016; Koskella, 2020) and contribute to global carbon and other biogeochemical
84 cycling, including transformations relevant for climate change (Bringel and Couée, 2015;
85 Dorokhov *et al.*, 2018; Cavicchioli *et al.*, 2019). Because perennial biofuel feedstocks are
86 selected to maximize foliage surface area, understanding the phyllosphere microbiome is
87 expected to provide insights into key microbial engagements that benefit the plant to support
88 productivity and stress resilience.

89 There are two general challenges in regulating the microbiome to promote crop vigor
90 and resilience to environmental stress. The first challenge is to distinguish the beneficial
91 members of the plant microbiome from transient or commensal members, with recognition
92 that some members that provide host benefit likely change situationally, either over plant
93 development or given environmental stress (Edwards *et al.*, 2015; Xu *et al.*, 2018; Zhalnina *et*
94 *al.*, 2018), while others are stable (Shade and Stopnisek, 2019; Stopnisek and Shade, 2021). The
95 second challenge is that plants and their agroecosystems are temporally dynamic over the

96 growing season, and their associated microbiota are also dynamic. It is currently unclear what
97 functions may be associated with phyllosphere microbial dynamics and their potential
98 interactions with plant hosts.

99 Previously, we used 16S rRNA gene amplicon analysis to identify a "core" cohort of
100 bacterial and archaeal taxa that were persistently associated with the phyllosphere
101 microbiomes' of two perennial biofuel feedstocks, miscanthus and switchgrass. Persistent
102 membership was established by collecting leaf samples over replicated field plots, over a
103 temperate seasonal cycle, and across two annual growing seasons for switchgrass (Grady *et al.*,
104 2019). Other studies in switchgrass have similarly reported that the leaf, and other plant
105 compartments, can be distinguished by microbiome compositions (e.g., Bulgarelli *et al.*, 2012;
106 Lundberg *et al.*, 2012; Bahulikar *et al.*, 2014; Roley *et al.*, 2019; Singer *et al.*, 2019), suggesting
107 selection to the leaf compartment. In the current study, we aimed to understand the functional
108 attributes and activities of persistent phyllosphere taxa, with an interest in specialized
109 adaptations to the leaf and interactions with the host plant that may inform the mechanisms
110 and nature of plant-microbe engagements. Therefore, we performed a seasonal analysis of
111 phyllosphere metagenomes for both miscanthus and switchgrass. We paired our metagenome
112 longitudinal series with metatranscriptome analyses at select time points in switchgrass
113 phenology to determine which functions were active seasonally. We performed genome-centric
114 analyses of the metagenome and focused on understanding the seasonal dynamics and
115 functions of a focal subset of medium- and high-quality metagenome-assembled-genomes
116 (MAGs) that we could bin from these data. Our results reveal functions supportive of a leaf-
117 associated lifestyle and seasonal activities of persistent phyllosphere members. Finally, we

118 provide evidence that these genomes were detected in various sites and years beyond our
119 original study plots, suggesting that they are general, consistent inhabitants of bioenergy
120 grasses.

121 **Materials and Methods**

122 *Site description and sampling scheme*

123 Switchgrass and miscanthus leaves and corresponding contextual data were collected at
124 the Great Lakes Bioenergy Research Center (GLBRC) located at the Kellogg Biological Station in
125 Hickory Corners, MI, USA (42°23'41.6" N, 85°22'23.1" W) (**Figure 1**). We sampled switchgrass
126 (*Panicum virgatum L.* cultivar Cave-in-rock) and miscanthus (*Miscanthus x giganteus*), from the
127 Biofuel Cropping System Experiment (BCSE) sites, plot replicates 1-4, as previously described
128 (Grady *et al.*, 2019). This included collecting leaves from switchgrass and miscanthus at 8 and 9
129 time points, respectively, in 2016 and switchgrass at 7 time points in 2017 seasons (**Table 1,**
130 **Figure 1, Dataset 1, Dataset 2**). We collected leaves for RNA isolation at three phenology-
131 informed switchgrass time points in 2016 (emergence, peak growth and senescence according
132 to GLBRC standard phenology methods https://data.sustainability.glbrc.org/protocols/165) to
133 assess the potential for sufficient mass and quality RNA extraction from the switchgrass leaf
134 surface, and then expanded to include leaves from all switchgrass sampling time points in 2017.
135 Leaves for RNA isolation were flash-frozen in liquid nitrogen immediately and stored at -80°C
136 until processing.

137

138 *Phyllosphere RNA and DNA isolation*

139 Phyllosphere epiphyte DNA was isolated and processed (Grady *et al.*, 2019). DNA
140 concentrations were normalized to 4 ng/ml. Phyllosphere epiphyte RNA was isolated using a
141 benzyl chloride liquid:liquid extraction, based on (Grady *et al.*, 2019) that was newly modified
142 for RNA isolation based on the published methods of (Suzuki *et al.*, 2001). Approximately 5 g of
143 intact, frozen leaf material was added to a 50 ml polypropylene conical tube (Corning #430290)
144 and kept frozen on liquid nitrogen while samples were weighed and transferred. Denaturing
145 Solution (DS) was prepared with 4.2 M guanidine thiocyanate, 25 mM sodium citrate dihydrate
146 pH 7.0, 0.5% (v/v) sodium n-laroyl sarcosine in Milli-q water and was filter sterilized through
147 0.22 mm filters. Immediately prior to extraction, a working stock of DS was prepared fresh by
148 adding 2-mercaptoethanol to a final concentration of 5% (v/v, DS/2-ME). To each leaf tube, 5
149 ml of benzyl chloride, 2.5 ml of 3M sodium acetate (pH 5.2), and 5 ml of the working stock of
150 DS/2-ME was added. The tube was incubated in a 60°C water bath for 20 minutes with
151 vortexing every 1 min. The leaves were removed from the conical tube using ethanol-sterilized
152 forceps and discarded.

153 Five ml of chloroform:isoamyl alcohol (24:1) were added to the remaining solution in
154 each conical tube, which were then shaken by hand for 15 seconds and incubated on ice for 15
155 minutes. The tubes were then centrifuged at 12,000 x g for 15 min at 4°C to separate aqueous
156 and organic phases. Up to 5 ml of the upper, aqueous phase was transferred to a clean 15 ml
157 polypropylene tube, without disrupting the white interface. Two and a half ml of sodium
158 citrate buffer (1.2 M sodium chloride, 0.8 M sodium citrate dihydrate in Milli-q water, filter
159 sterilized at 0.22 mm) and ice-cold isopropanol were added to achieve a final volume of 12.5
160 ml. Next, the tubes were centrifuged at 12,000 x g for 15 minutes to pellet the RNA, and the

161 remaining supernatant aspirated using a pipette. The RNA pellets were resuspended in 0.3 ml
162 of working DS/2-ME solution, and afterwards 0.3 ml of ice-cold isopropanol was added and
163 mixed by pipetting gently. The solution was incubated for 30 minutes at -20°C, and then the full
164 volume was transferred to a clean nuclease-free 1.7 ml tube and centrifuged at 16,000 x g for
165 15 minutes at 4 °C. The supernatant was removed, and the pellet washed in 1 ml of nuclease-
166 free 75% ethanol. Tubes were then centrifuged at 16,000 x g for 15 minutes at 4 °C, and
167 supernatant again removed using a pipette. The remaining pellet was air dried to completely
168 remove residual ethanol, then resuspended in 30 ml of nuclease-free Tris-EDTA buffer, pH 8.0.

169 Genomic DNA (gDNA) was removed using RNase-free DNase I (Thermo Fisher #AM2222)
170 per manufacturer's instructions. The RNA was then purified using the RNeasy MinElute Cleanup
171 Kit (Qiagen Germantown, MD, USA) according to manufacturer's instructions. The absence of
172 contaminating gDNA was confirmed by lack of amplification of the 16S rRNA gene V4 region by
173 PCR (Caporaso *et al.*, 2011) with positive and negative controls. This RNA isolation method was
174 developed to most closely align with our established phyllosphere epiphyte DNA isolation
175 (Grady *et al.*, 2019) in order to minimize potential bias introduced during biofilm disruption or
176 microbial cell lysis, as well as to minimize contamination from host RNA or genomic DNA by
177 leaving the plant tissue intact. Commercial RNA extraction kits are primarily based on grinding
178 or bead beating whole tissue samples, which would result in over-representation of host-
179 derived nucleic acids and potentially introduce bias in microbial cell lysis efficiencies.

180

181 *Metagenome and metatrascriptome library preparation*

182 The Joint Genome Institute (JGI) performed the library preparation and sequencing from
183 submitted RNA and DNA samples. Plate-based DNA library preparation for Illumina sequencing
184 was performed on the PerkinElmer Sciclone NGS robotic liquid handling system using Kapa
185 Biosystems library preparation kit. 1.82 ng of sample DNA was sheared to 436 bp using a
186 Covaris LE220 focused-ultrasonicator. The sheared DNA fragments were size selected by
187 double-SPRI and then the selected fragments were end-repaired, A-tailed, and ligated with
188 Illumina compatible sequencing adaptors from IDT containing a unique molecular index
189 barcode for each sample library. The prepared libraries were quantified using KAPA
190 Biosystems' next-generation sequencing library qPCR kit and run on a Roche LightCycler 480
191 real-time PCR instrument. Sequencing of the flowcell was performed on the Illumina HiSeq
192 sequencer following a 2x151 indexed run recipe.

193 At JGI, plate-based RNA sample prep was performed on the PerkinElmer Sciclone NGS
194 robotic liquid handling system using Illumina Ribo-Zero rRNA Removal Kit (Bacteria) and the
195 TruSeq Stranded Total RNA HT sample prep kit following the protocol outlined by Illumina in
196 their user guide: https://support.illumina.com/sequencing/sequencing_kits/truseqstranded-
197 total-rna.html, and with the following conditions: total RNA starting material of 100 ng per
198 sample and 10 cycles of PCR for library amplification. The prepared libraries were quantified
199 using KAPA Biosystems' next-generation sequencing library qPCR kit and run on a Roche
200 LightCycler 480 real-time PCR instrument. Sequencing of the flowcell was performed on the
201 Illumina NovaSeq sequencer using NovaSeq XP V1 reagent kits, S4 flow cell, following a 2x151
202 indexed run recipe.

203

204 *Quality filtering of metagenomes and metatranscriptomes*

205 We proceeded with bioinformatic analysis (**Figure 2**) of 192 metagenome (**Dataset 1**)
206 and 78 metranscriptome (**Dataset 2**) observations that met JGI standards for raw data quality
207 based on the Illumina proprietary software. We used Trimmomatic (v0.39) (Bolger *et al.*, 2014)
208 to remove adaptors and filter low-quality reads from fastq files using the following arguments:
209 PE -phred33 ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:8:TRUE LEADING:3 TRAILING:3
210 SLIDINGWINDOW:4:15 MINLEN:36. After assembly, plant host reads were filtered out (for both
211 metagenomes and metatranscriptomes) by removing all reads that mapped to the switchgrass
212 genome (*Panicum virgatum* v1.0, DOE-JGI, http://phytozome.jgi.doe.gov/) or miscanthus
213 genome (*Miscanthus sinensis* V7.1 http://phytozome.jgi.doe.gov/) using bowtie2 (v 2.4.1),
214 samtools (v 1.13), and bedtools (v2.30.0) (Li *et al.*, 2009; Quinlan and Hall, 2010; Langmead and
215 Salzberg, 2012) (**Figure S1)**. To remove the fungal reads and improve the prokaryotic signal in
216 metagenome samples, we also filtered reads against 7 fungal genomes (Gostinčar *et al.*, 2014;
217 Druzhinina *et al.*, 2018; Gill *et al.*, 2019; Haridas *et al.*, 2020) that represent close relatives of
218 the most abundant fungal species in this system that we assessed and reported in our prior
219 work (Bowsher *et al.*, 2020) (**Table S1**). The genomes of these fungal species were retrieved
220 from the JGI Genome Portal.

221 *Metagenome assemblies and metagenome-assembled genome binning, curation, refinement,*
222 *annotation*

223 Two metagenome assemblies, one for switchgrass and one for miscanthus, were created based
224 on metagenomes collected in 2016. These filtered metagenome reads were combined and used
225 for co-assembly with MEGAHIT (v 1.2.9) using--kmin-1pass (low sequencing depth) and--presets

226 meta-large (complex metagenome) (Li *et al.*, 2015). Additionally, we curated metagenome
227 assembled genomes (MAGs) from the 2016 switchgrass and miscanthus metagenome libraries
228 (n = 136 metagenomes) using Metabat (v2.2.15) (Kang *et al.*, 2019). MAG assemblies were
229 performed using filtered reads from switchgrass and miscanthus sampled from 2016,
230 separately, to maximize completeness and reliability, as also done in other studies (Nayfach *et*
231 *al.*, 2019). To assess the MAGs quality and completeness, we used CheckM (v.1.13 with the
232 lineage_wf option) estimates of quality and completeness (Parks *et al.*, 2015). Among 238
233 MAGs assembled from the switchgrass or miscanthus phyllosphere metagenomes (**Dataset 3**),
234 we selected a subset of MAGs based on: completeness greater than 50% and contamination
235 less than 10%. We identified replicated bins associated with MAGs using dRep (v3.2.0,Olm et
236 al. 2017), resulting in the removal of a single MAG. An additional MAG was removed due to
237 insufficient read recruitment in metatranscriptomes (described below), for a final total of 40
238 focal MAGs, including 7 high-quality and 33 medium quality (Bowers *et al.*, 2017) (**Dataset 3).**

239 Read recruitment was performed with MAGs that meet either medium- and/or high-
240 quality standards (Swan *et al.*, 2013; Bowers *et al.*, 2017). The metagenome abundance of
241 contigs in MAGs in each sample was estimated based on the median coverage of filtered
242 metagenome reads associated to each MAG contig. Specifically, Bowtie2 (v2.2.2) was used to
243 align reads to all focal MAG contigs (using default setting and allowing for a single read to map
244 to only one reference). Bedtools (v2.28) was used to estimate the coverage of each basepair
245 within the contig. The estimated abundance of contig was based on the median basepair
246 coverage of all reads mapped to the contig, and the estimated abundance of a MAG was based
247 on the average median coverage of all contig within its bins. The metatranscriptome

248 abundance was estimated based on protein-encoding genes identified in MAGs. For each MAG
249 contig, open reading frames (ORFs) and functional genes were identified using Prodigal (v2.6.3,
250 default parameters). Transcripts were mapped to ORFs associated with each MAG to estimate
251 median base coverage of each ORF (Bowtie2, default parameters, no multiple mappings
252 allowed). To normalize varying sequencing depths, estimated abundances were normalized by
253 the sum of the median base pair coverage of housekeeping genes identified in each sample.
254 Housekeeping genes were identified based on full sequence alignment to the HMM models of
255 71 housekeeping single-copy genes with E-value of less than 1e-5 (Eren *et al.*, 2015; Lee, 2019).
256 If housekeeping genes were not identified in a metagenome or metatranscriptome, samples
257 were removed from further analysis. We also estimated the total reads associated with each
258 MAG for metagenomes and metatranscriptomes (**Dataset 4**). One MAG, M22, had an average
259 of 48 reads map to metatranscriptome and was removed for further analysis. For
260 metatranscriptome analyses, only ORFs with the top 75% of observed abundances and in at
261 least 10% of samples were considered.

262 Functional annotation of ORFs in focal MAGs was done with the DRAM tool (v1.1.1,
263 (Shaffer *et al.*, 2020) using UniRef90, MEROPS, PFAM, dbCAN-HMMdb (v8) databases (all
264 compiled with DRAM on February 12, 2021), and the KEGG database which we manually added
265 to the DRAM pipeline (release January 1, 2018). To obtain functions related to terpene
266 metabolism, ORFs associated with any KEGG annotation that contained the phrase “terpen”
267 were selected.

268 MAGs were assigned taxonomy using GTDB-tk (v1.4.0, (Parks *et al.*, 2018)). Assembled
269 focal MAGs were aligned againts the chloroplast genome of *Panicum virgatum* (NC015990)

270 (BLAST, 2.10.1). We detected partial matches to eight bins and no full alignments, confirming
271 that those bins were not chloroplast genomes. Additionally, we compared the focal MAGs
272 taxonomy to our previously detected 16S rRNA gene core cohort (Grady *et al.*, 2019). This
273 cohort consists of 61 phylogenetically diverse bacterial OTUs (97% clustering). Taxonomy of
274 these 61 OTUs was compared with the 40 MAGs at the genus level only. Due to differences in
275 nomenclature between the GTDB and SILVA databases we removed extensions for MAGs
276 classified as *Pseudomonas_E* or *Aeromonas_A*.

277

278 *Statistical analyses*

279 Pairwise comparisons of functional roles (the cumulative sum of all associated ORFs) between
280 the early (May – June) and late (July – Sep) season were performed with the Kruskal-Wallis test
281 by ranks. Early and late season were delineated by plant phenology (flowering/senescence in
282 late season). The distribution and dynamics of MAGs' transcripts were compared with Ward's
283 method for hierchical clustering using euclidian distances of the estimated metatranscriptome
284 abundances using *pvclust* (version 2.20)(Suzuki and Shimodaira, 2006).

285

286 *Assessment of biosynthetic gene clusters (BGC) on focal MAGs*

287 Biosynthetic gene clusters (BGC) were predicted by antiSMASH (v6.0) (Blin *et al.*, 2021)
288 and further annotated by Big-SCAPE (v1.1.0) (Navarro-Muñoz *et al.*, 2020). While there are only
289 8 BGC classes used in Big-SCAPE (i.e. PKS I, PKS other, NRPS, RiPPs, Saccharides, Terpenes,
290 PKS/NRPS Hybrids, Others), antiSMASH provides more detailed classification. We leveraged

291 both outputs to investigate the diversity and expression of predicted BGCs in focal MAGs. From
292 each predicted gene cluster, we extracted the location of the biosynthetic gene (i.e.,
293 gene_kind="biosynthetic" and core_position). To evaluate the transcription of putative BGCs
294 within a MAG, we searched for any transcripts that are mapped to the same genomic region of
295 predicted biosynthetic genes.

296 *Predicting functional potential of focal MAGs*

297 We used gapseq (Zimmermann *et al.*, 2021) to predict completed metabolic pathways in
298 our focal MAGs. We used the 'find -p all –b 200' option to search for pathways against the
299 MetaCyc database. We filtered out incomplete pathways and remaining pathways were
300 grouped into broader categories using MetaCyc classification and manually curated to focus on
301 understanding the pathways relevant either to the plant environment or for microbial
302 interactions with plants. These categories were defined as potential involvement in: i) plant
303 (using plant metabolites/cell components), ii) phytohormone (known/potential involvement in
304 phytohormone homeostasis), iii) stress (e.g., drought, reactive oxygen species), and iv) general
305 (pathways that utilize potential plant derived products). Furthermore, we also manually
306 searched for genes/pathways that were missed by gapseq but are known to be related to
307 adaptation to plant-associated lifestyle, including secretion systems (Tseng *et al.*, 2009; Lucke *et*
308 *al.*, 2020), oxidation of trace gases (H_2 and CO)(Bay *et al.*, 2021; Palmer *et al.*, 2021), oxalate
309 degradation, and phytohormone production/degradation. Similar to our biosynthetic gene
310 analysis, the activity of predicted pathways was estimated based on mapped transcripts to key
311 identified genes (**Table S2**). Contigs of focal MAGS were aligned (BLAST v2.7.1+) against nine

312 isoprenoid precursor biosynthesis genes previously reported (Julsing *et al.*, 2007). The top hit
313 of each query was considered and accepted as aligned if E-value was less than 1e-5.

314

315 *Detection of MAGS in public metagenomes*

316 To evaluate the presence of MAGs in other switchgrass and miscanthus metagenomes, 55
317 publicly available switchgrass, miscanthus, and corn metagenomes were used (**Dataset 5**).
318 Metagenome reads were mapped to focal MAGs, using the same methods for mapping and
319 abundance estimation described above for the metagenomes generated in this study.

320

321 *Data and code availability*

322 Metagenome and metatranscriptome data are available through the Joint Genome Institute
323 Genome Portal under proposal ID 503249. Code and links to data and metadata are available
324 on GitHub (https://github.com/ShadeLab/PAPER_Howe_2021_switchgrass_MetaT). MAGs are
325 available in NCBI under project IMG. All metadata, including metadata standards for
326 metagenomes, metatranscriptomes, and metagenome-assembled genomes are provided as
327 supporting datasets (**Datasets 1-6**).

328

329 **Results**

330 *Key leaf bacterial populations and general dynamics*

331 Among all assembled MAGs (n=238), we focused analysis on 40 that were high- and
332 medium- quality based on completeness and contamination standards (**Figure 3A**, **Dataset**
333 **3**)(Bowers *et al.*, 2017). The consistent detection of these 40 MAGs in both miscanthus and
334 switchgrass samples (**Figure 4**) suggests that their originating populations were not host-
335 specific but rather distributed among these perennial grasses; this supports our previous results
336 from an 16S rRNA gene amplicon survey that revealed substantial overlap in the major leaf
337 bacterial taxa across the two crops (Grady *et al.*, 2019). Furthermore, there were 38 "core"
338 OTUs from our prior survey that were also found in 25 focal MAGs here per their taxonomic
339 identifications (**Figure 3B**); our previous study prioritized OTUs based on abundance and
340 seasonal and field-replicated occupancy, and so this conservatively shows that at least half of
341 the focal MAGs represent the most abundant and consistently detected taxa in this ecosystem.
342 The 40 focal MAGs most represented the orders Rhizobiales (n=8/40), Actinomycetales
343 (n=6/40), Burkholderiales (n=6/40), and Sphingomonadales (n=6/40) (**Dataset 3**).

344 Focal MAGs exhibited varied seasonal patterns in their overall metagenome (**Figure 4A-**
345 **C**) and metatranscriptome (**Figure 4D-E**) read recruitment, with early, late, and consistent
346 seasonal detection observed. However, in the metatranscriptomes of both years for
347 switchgrass, there was a strong seasonal enrichment over time, with increases in the later
348 months of sampling. Phenologically, August samples corresponded with (late) peak biomass
349 and fruiting for switchgrass, and with a closed canopy for both switchgrass and miscanthus.
350 Senescence occurred as early as mid-September for switchgrass and as late as November for
351 miscanthus. Notably, transcript recruitment to the MAGs was robust even when there was
352 inconsistent detection in metagenome read recruitment, suggesting that these MAGs were

353 consistently active and had relatively high activity even when their genome detection was
354 obscured by the flanking community or host signal.

355 MAGs were clustered based on coherence in seasonal transcript dynamics (**Figure 4D-E,**
356 **Figure S2**). Five coherent and statistically supported groups (clusters) of 40 MAGs were
357 identified (**Figure S2**). Cluster 1 contained a single MAG S28, assigned to genus *Pseudomonas*
358 (S28, 98% complete, 1.9% contamination) and its genes were enriched early in the season
359 (2017) but the MAG was highly active throughout the season. The other four clusters contained
360 numerous MAGs and were relatively more dynamic, with trends towards late-season
361 enrichment in transcripts. Several clusters contained MAGs annotated to the same class,
362 suggestive of taxonomic coherence in seasonal activities. For example, cluster 3 included all
363 but one (of eight) Actinomycetia MAGs.

364

365 *MAG genes and activities support a leaf-associated lifestyle*

366 Not surprisingly, the most abundant subsystems identified among the 40 focal MAGs were
367 associated with bacterial growth, such as carbohydrates, energy, amino acid, and nucleotide
368 metabolisms (**Figure 5**). Leaf metatranscriptomes also showed that most subsystem transcripts
369 were steadily enriched over the season. However, subsystem associated with colonization and
370 adaptation, such as cell motility, signal transduction, community-associated pathways, and
371 environmental adaptation trended down in their normalized transcripts over the season. A
372 closer investigation of these down-trending functions revealed that they could be attributed
373 almost exclusively to MAG S28 (*Pseudomodales* from cluster 1 in **Figure 4DE** and **Figure S3**).

374 Finally, a few KEGG classifications were relatively stable and lacked overall seasonal trends,
375 including transcripts for translation, transcription, and antimicrobial resistance.

376 We selected the most prominent phyllosphere transcripts based on their high detection
377 among the 40 focal MAGs, hypothesizing that some of these functions shared between
378 bacterial populations may represent functions for survival on the leaf. A total of 124 distinct
379 functional roles were identified in at least 39 of the 40 MAGs (**Dataset 6, Figure S4**). Of these,
380 most functions were associated to translation (n=48/124). Other broadly present functions
381 were associated with metabolism of energy (n=29/124) and carbohydrates (n=28/124). There
382 were also several functional roles associated with cofactors and vitamins (n=18/124). While
383 many functions were enriched in transcripts in the late season (July-Sept) relative to the early
384 season (May-June), three functional roles in particular stood out, with enrichments greater
385 than 20-fold in the late season transcripts and included proteins classified as short chain
386 dehydrogenase, molybdopterin oxidoreductase, and polyketide cyclase (**Table 2**, **Dataset 6,**
387 **Figure S4**).

388 We then performed metabolic, biosynthetic, and plant-associated gene pathway
389 analyses to understand the functions detected among focal MAGs in the phyllosphere and their
390 activities inferred by transcript recruitment to the pathways (**Figure 6**). Pathways for terpene
391 metabolism (34/40), betaine biosynthesis (30/40), trehalose metabolism (25/40), cyanide
392 degradation (21/40), ROS degradation (23/40), and indole acetic acid (IAA) degradation (40/40)
393 were the most common pathways shared and active across MAGs, and were found in lineages
394 from all four classes represented by the focal MAGs. Thus, these pathways likely are common
395 among phyllosphere members and highly supportive of a leaf-associated lifestyle. In addition,

396 there were some functional pathways and activities that may be more specialized functions
397 because they were detected in only a few MAGs (e.g., xylitol metabolism). However, there was
398 no clear phylogenetic pattern to the distributions of these putative specialized functions.

399 Because of the near-ubiquitous detection of terpene-related biosynthetic genes on the
400 focal MAGs, we examined their annotations and recruitment more closely using gene function
401 predictions (i.e., ORFs), which was expected to improve sensitivity of functional annotation
402 relative to metabolic pathway tools (e.g., BGC). We detected 278 ORFs associated with
403 terpenes annotations in the metatranscripts. Collectively, these terpene-related genes
404 followed the trend of gradual enrichment over the season (**Figure 7A**) and were relatively
405 abundant in both years. The highest enrichment subcategory was terpenoid backbone
406 biosynthesis, which included 29 ORFs. Among several isoprene biosynthesis genes detected, of
407 interest was relatively high, seasonal enrichment of the two terminal enzymes in the non-
408 mevalonate isoprene biosynthesis pathway that is employed by bacteria, the *gcpE* and *lytB*
409 (**Figure 7B, Figure S5**). *GcpE* and *lytB* gene transcripts were detected in more than a third of the
410 MAGs (13/40), and these included MAGs representing all three phyla detected. Six of the eight
411 Proteobacteria that recruited isoprene biosynthesis transcripts were annotated to
412 *Methylobacterium* (alpha). The four Actinomycetia were more distributed phylogenetically and
413 annotated as genera *Frigoribacterium* M1, *Microbacterium* M105, *Amnibacterium* M67, and
414 Pseudokineococcus M86. The single Bacteroidota that had isoprene biosynthesis transcripts
415 was a *Hymenobacter* M9. We then investigated the 40 MAGs for detection of any of the nine
416 genes previously reported to be involved in bacterial biosynthesis of isopentenyl diphosphate, a
417 precursor terpenoids like isoprene (Julsing *et al.*, 2007) and found that 29/40 MAGs had six or

418 more of the genes detected, and that all MAGs had at minimum 2 of the genes (**Table S2**). The
419 consistent detection of genes associated with isoprene pathways in these MAGs, which are
420 >50% complete, suggests that biosynthesis of isoprene-related molecules may be a prominent
421 leaf strategy by phyllosphere bacteria. *Psuedomonas* MAG S28, noted previously to be the
422 dominant population that colonized and activated early in the season (**Figure 4, Figure S2**), had
423 high isoprene biosynthesis transcript enrichment early in the season that then declined.
424 However, the other eleven MAGs harboring genes from isoprene biosynthesis pathways
425 compensated with increased activity in the late season.

426 Returning to the BGC annotations (by antiSMASH , **Figure 6**), most of the terpene
427 transcripts were associated to pigment biosynthesis (e.g., carotenoids). The second most
428 consistent class of BGC transcripts on the focal MAGs were non-ribosomal peptide synthase
429 (NRPK) genes, but the majority of these did not have additional annotations beyond the general
430 category. Therefore, the ORF analysis and BGC detection were complementary in results,
431 especially for terpene-related functions. Another notable BGC finding was that all but two (of
432 eight) Actinomycetia MAGs had transcripts for type III polyketide synthases, while this BGC was
433 less commonly detected among Proteobacteria and Bacteroidota.

434

435 *MAGs were detected in metagenomes from different field sites, crops, and years*

436 We next asked if these focal MAGs were detected more broadly in other crop
437 metagenomes. Because leaf metagenomes were not findable, we curated related soil
438 metagenomes from local cropping systems in Michigan and also from the publicly available

439 data within the Integrated Microbial Genomes (IMG) database, which serves as a repository for
440 the Joint Genome Institute's sequencing efforts through the US Department of Energy. We
441 were surprised to find that, for each metagenome investigated, reads could be mapped to this
442 set of MAGs. This included soil metagenomes from both switchgrass and miscanthus fields,
443 three locations (Iowa, Michigan, and Wisconsin) and sampled across five different years (**Figure**
444 **8)**. These results suggest that the detected and analyzed MAGs can be widely distributed in
445 midwestern agroecosystems, and potentially of general importance for perennial crop
446 environments.

447

448 **Discussion**

449 Here, we report the first multi-year seasonal metagenome and metatranscriptome
450 assessment of plant phyllosphere in agricultural field conditions, focusing on the bacterial
451 functions associated with two promising biofuel feedstocks. We expect these findings to have
452 relevance for other grasses or systems with substantial aerial biomass, including native prairie.
453 Furthermore, our collection of MAGs included phyllosphere members previously reported as
454 abundant, persistent, or key for microbiome assembly in other plants, including the model
455 *Arabidopsis* (e.g., (Carlström *et al.*, 2019), specifically: Sphingomonadales, Pseudomonadales,
456 Actinomycetales, Burkholderiales, and Rhizobiales. Therefore, the patterns and consistently
457 detected functions among this curated collection offer general insights as well as seasonal
458 phyllosphere functions.

459 There are multiple lines of evidence that the focal MAGs discussed here represent key
460 lineages in the switchgrass and miscanthus phyllosphere. First, there is ample overlap among
461 these focal MAGs and the core taxa of high abundance and occupancy from our previous 16S
462 rRNA gene amplicon analysis of the switchgrass and miscanthus phyllosphere diversity (Grady
463 *et al.*, 2019), including a MAG associated to *Hymenobacter* M9, which is a genus that had high
464 occupancy across fields and over time, as well as several OTUs assigned. This same time series
465 was investigated in our previous amplicon analysis and the core taxa were prioritized using
466 consistency across replicated fields at the same time point, persistence over time, and relative
467 abundance. This suggests that the populations represented by the MAGs are not rare taxa that
468 are transient to the system. Second, we were able to generate quality assemblies from the
469 complex metagenome data, which is a process that is generally biased towards abundant
470 members. These MAGs are highly represented in read abundance in metagenomes and
471 recruited relatively more metatranscriptome reads as well, suggesting that they are both
472 abundant and active in the phyllosphere. Though it is possible that there are key lineages
473 missing among our collection, we are confident that those discussed here are among the major
474 host- or environment-selected populations inhabiting the switchgrass and miscanthus
475 phyllosphere.

476

477 *Stress responses: trehalose, betaine, reactive oxygen, IAA*

478 Trehalose is a disaccharide that protects cells against salt, water, and osmolyte stress by
479 serving as a stabilizing chemical chaperone, either by displacing water from protein surfaces or

480 by vitrifying around protein structures to shield them (Laskowska *et al.*, 2020). Similarly,
481 betaine is another commonly biosynthesized osmoprotectant used by microorganisms to
482 contend with water, salt, and temperature stress (Zou *et al.*, 2016). Both have been
483 hypothesized to be important survival strategies of microorganisms in the phyllosphere, with
484 supporting evidence from isolate genome analyses (Rastogi *et al.*, 2013). Here, we show both
485 trehalose and betaine to be prominent among MAG populations and activated consistently on
486 the leaf surface, suggesting that they are not seasonally activated but rather necessary for the
487 leaf-associated lifestyle.

488 In bacteria, trehalose metabolism prevents cellular overflow metabolism and carbon
489 stress by redirecting glucose-6-phosphate from conversion to pyruvate (Laskowska *et al.*, 2020).
490 Trehalose biosynthesis is common among bacteria and archaea that live in arid, saline, thermal,
491 or seasonally dry environments (e.g., (Urrejola *et al.*, 2019), (Dai *et al.*, 2015; Lacerda-Júnior *et*
492 *al.*, 2019), and it has also been reported to be induced in a pseudomonad by ethanol (Harty *et*
493 *al.*, 2019), which can originate inside plant cells (Kimmerer and MacDonald, 1987) or more
494 generally in the roots (e.g., Ferner *et al.*, 2012), especially during stress, fruit ripening, or
495 senescence (Kimmerer and Kozlowski, 1982). In switchgrass and plants in general, trehalose
496 concentration is increased in response to drought conditions (Liu *et al.*, 2015), and its
497 precursor, trehalose-6-phosphate, induces senescence when carbon is readily available
498 (Wingler *et al.*, 2012). Furthermore, the *A. thaliana* phyllosphere member *Sphingomonas*
499 *melonis* was reported to regulate trehalose biosynthesis during growth conditions that
500 promoted mild stress (Gottschlich *et al.*, 2019). Given this, it makes sense that the majority of

501 MAGs had enrichment of transcripts related to trehalose metabolism, which would support a
502 plant-associated lifestyle during drought and host senescence (Gottschlich *et al.*, 2019).

503 Betaine biosynthesis in bacteria often begins with oxidation of choline, which is a part of
504 plant tissues and can be transported into the cell (Chen *et al.*, 2013). Indeed, choline
505 degradation was detected in 9/10 Gammaproteobacteria MAGs here (**Figure 6**). Microbial-
506 derived osmolytes such as betaine and trehalose have been suggested as targets for
507 biotechnology development to support crop stress tolerance. However, plants can
508 biosynthesize betaine and can be divided into groups of those that do and do not accumulate it
509 in concentrations that are supportive for stress tolerance (Valenzuela-Soto and Figueroa-Soto,
510 2019). For example, in switchgrass, the concentration of glycine betaine was not predictive of
511 differences in drought tolerance among different genotypes, while trehalose was, along with
512 abscisic acid, spermine, and fructose (Liu *et al.*, 2015). Furthermore, our data suggest no
513 notable microbial limitations in the genetic potential or activation of betaine biosynthesis in
514 phyllosphere.

515 Reactive oxygen species (ROS) serve as signals for various developmental and cellular
516 processes in plants, and here we detected active pathways for ROS degradation in the majority
517 of focal MAGs. Though the precise mechanisms are unclear, homeostasis of ROS are expected
518 to be involved in senescence (Kerchev *et al.*, 2015), which is relevant to our study given that at
519 least some to a majority of senescing plants were observed per plot in August and September
520 sampling dates, respectively. Additionally, ROS accumulate in plants that are exposed to abiotic
521 stress, to negative effects. ROS degradation is one of many functions phyllosphere

522 microorganisms employ to contend with expected fluctuations in ROS on the leaf surface
523 (though, these fluxes are difficult to measure (Kerchev *et al.*, 2015)). Previously, several genes
524 relevant for oxidative stress response were differentially regulated in a wild-type phyllosphere
525 bacteria *Sphingomonas melonis* strain Fr1 as compared to a knock-out mutant for regulation of
526 general stress response, when both were grown in a medium expected to induce low-levels of
527 stress. (Gottschlich *et al.*, 2019). Given that managing plant ROS is a target for reducing crop
528 stress and regulating plant development (Considine and Foyer, 2014; Kerchev *et al.*, 2015), it is
529 possible that manipulating microbial ROS degradation could be applied as one tool to achieve
530 such efforts, but much more research is needed to understand the microbial-host interaction
531 given ROS exposure or accumulation, and any possible ROS signaling between them.

532 IAA is a phytohormone produced by plants to regulate many processes in growth and
533 stress response (e.g., Spaepen *et al.*, 2007; Egamberdieva *et al.*, 2017). It is also made by many
534 microorganisms, including those shown to support plant grown promotion (e.g., Egamberdieva
535 *et al.*, 2017). Therefore, the activity of IAA degradation pathways by focal MAGs is expected
536 given the redundancies between plants and microorganisms in synthesizing and responding to
537 IAA and demonstrates microbiome responsiveness to feedbacks in the host environment.

538

539 *Biosynthesis of isoprene-related molecules*

540 Most of the functions identified in our MAGs suggest general requirements for growth
541 and maintenance given a leaf-associated lifestyle (e.g., carbohydrate and amino acid
542 metabolism, pigment production to protect from radiation, similar to previous reports, e.g.,

543 Lajoie *et al.*, 2020). However, BGC analysis revealed surprising consistency in terpene
544 metabolism pathways, leading us to look more closely at transcript ORFs associated to
545 terpenes. This analysis revealed particular enrichment in pathways and key genes associated
546 with isoprenoid biosynthesis. Isoprenoids are a class of volatile terpenes that are generally
547 abundant and reactive, and they engage in indirect and complex feedbacks with methane and
548 nitrous oxide greenhouse gases (McGenity *et al.*, 2018). Isoprene is one of the simplest
549 isoprenoids. It is released by many plant species, and much of it is synthesized within the
550 methylerythritol phosphate pathway of the chloroplast (MEP, aka:non-mevalonate) (Sharkey *et*
551 *al.*, 2008). Isoprene is thought to act as a signaling molecule in stress response (Zuo *et al.*,
552 2019). Studies have also found that isoprene emission protects leaf photosynthesis against
553 short episodes of high-temperature (Sharkey and Yeh, 2001). Plants emit isoprene from
554 matured, photosynthetically active leaves, and emissions are light responsive (Sharkey *et al.*,
555 2008). However, senescing leaves have been reported to decrease in their isoprene emissions
556 relative to leaves at peak growth (Sharkey *et al.*, 1991). Both switchgrass and miscanthus have
557 been reported to emit relatively low basal levels of isoprene (Eller *et al.*, 2011; Morrison *et al.*,
558 2016).

559 We hypothesize that that members of the biofuel feedstock phyllosphere bacterial
560 community may be either compensating for the loss of plant-derived isoprene, engaging in
561 interspecies isoprenoid signaling with the host, protecting plant photosynthesis from thermal
562 damage, quenching reactive oxygen species, or possibly producing isoprenoids as overflow
563 metabolites (as hypothesized for *Bacillus subtilus* (Sivy *et al.*, 2002)). Bacterial isoprene
564 degraders and synthesizers are widespread in nature (McGenity *et al.*, 2018) and have been

565 previously investigated in phyllosphere communities of the relatively high isoprene emitter
566 *Populus* spp (Crombie *et al.*, 2018), as well as in soils (El Khawand *et al.*, 2016), which can serve
567 as an isoprene sink. Stable isotope assays have been used to determine that a subset of
568 bacteria community members degrade isoprene, including several Actinobacteria (*Rhodococcus*
569 spp.) and *Variovorax* (Proteobacteria) (El Khawand *et al.*, 2016; Crombie *et al.*, 2018). Our MAG
570 collection contains several Actinomycetia, a *Hymenobacter*, several *Methylobacterium*, and
571 *Pseudomonas* MAG S28 that show activation of genes involved in isoprenoid biosynthesis, and
572 adding support for their involvement in related molecular feedbacks in the phyllosphere.
573 Furthermore, given that a third of our modest collection had activity of key isoprenoid
574 biosynthesis genes, there may be many more leaf bacterial contributors that we did not detect
575 here. In addition, these activities were commonly to three Bacteria phyla, it suggests that
576 biosynthesis of isoprene-related molecules may be a very common phyllosphere microbiome
577 function. As isoprene is a precursor to sidechains needed for several quinones (Nowicka and
578 Kruk, 2010), it could be speculated that leaf bacteria scavange isoprene emitted by the host
579 plant to supplement bacterial synthesis of these sidechains, but then compensate with de novo
580 biosynthesis if host decreases production. We observe that isoprenoid synthesis increases
581 seasonally in the majority of MAGs containing these pathways, and concurrently with when
582 plant isoprene emissions also is expected to decrease, directs future work to understand these
583 dynamics and potential isoprenoid-mediated bacterial-host engagement.

584

585 *MAGs of interest*

586 We highlight three MAG populations that were of interest because of their taxonomy,
587 functions, dynamics, or detection. All were shared with taxa in our prior 16S rRNA survey,
588 supporting their inclusion as part of the "core" set that was selected by abundance and
589 occupancy . First, high-quality MAG S28 (>97% complete, <2% contamination) was a prominent
590 pioneer and active colonizer of the leaf (Figure 4 Group 1). MAG S28 is related to *Pseudomonas*
591 *cerasi*, a species reported to have phytopathogen relatives (Kałużna *et al.*, 2016), but we did not
592 note any disease symptoms on the leaves analyzed. This population had expected traits of a
593 strong surface colonizer, including colonization, adaptation and motility subsystems. It also had
594 six pathways related to phytohormone responses (out of 7 total phytohormone pathways
595 observed in these data), including activated ethene biosynthesis, ACC deaminase, and
596 degradation of ethylene glycol, putrescine, salicylate and IAA. These data suggest that S28 has
597 several mechanisms to engage or respond to the host via phytohormones.

598 Next, MAG M9, identified as *Hymenobacter*, was of interest because it was associated to
599 the most numerous taxonomic group detected in our prior 16S rRNA gene survey (Grady *et al.*,
600 2019) and not among the most typically investigated phyllosphere lineages in the literature.
601 While M9 populations were first detected early in the season, its transcripts were enriched in
602 the late season (**Figure 4** Group 5). MAG M9 had detected and activated galactonate, N-
603 acetylglucosamime, and lactose metabolisms, which were not common among the focal MAGs.
604 It also had activated benzoate, curcumin, and putrescine degradation, as well as cyanide
605 detoxification, type VI secretion, and dihydrogen oxidation. While M9 also had some pathways
606 that were common among these MAGs (e.g., ROS and IAA degredation, terpene biosynthesis),
607 its suite of more sparsely detected pathways and functions suggest a specialized role in the

608 phyllosphere community. Notably, M9 has 65% completion and 0% detected contamination,
609 suggesting more functional potential remains to be discovered for this and similar
610 *Hymenobacter* lineages inhabiting the phyllosphere.

611 Finally, we selected a representative Actinomycetia MAG M60, a *Quadrisphaera* lineage
612 that had activated isoprenoid biosynthesis and had increased activity late in the season along
613 with the majority of focal MAGs (Figure 4, Group 3). Studies have found that members of
614 Actinomycetia are important part of phyllosphere that contribute to disease prevention and
615 plant growth [El-Tarabily 2009; Javed 2021; Anwar 2016]. MAG M60 had several
616 oligo/polysaccharide metabolisms that were infrequently detected in these data, including
617 glycogen, melibiose, and trehalose. Despite its high completeness and low contamination
618 (>95% and < 5%, respectively), M60 was sparsely annotated by the methods we applied.
619 However, *Quadrisphaera* have been reported to be highly abundant in the phyllosphere or
620 endosphere of various plants (Bao *et al.*, 2020).

621

622 *Conclusions*

623 Many recent review, perspective, and opinion pieces have urged integration of multi-
624 omics approaches to improve understanding of the microbiome and its relationship to the host
625 plant (Rastogi *et al.*, 2013; Levy *et al.*, 2018; Remus-Emsermann and Schlechter, 2018; Beilsmith
626 *et al.*, 2019; Trivedi *et al.*, 2020). However, most integrative studies have focused almost
627 exclusively on the rhizosphere as the compartment of soil-plant feedbacks and nutrient and
628 water acquisition for the host. Though leaves are readily accessible for sampling, the

629 phyllosphere microbiome is challenging to investigate using throughput, cultivation-
630 independent approaches like metagenomics and metatranscriptomics. There are high levels of
631 host and chloroplast contamination in leaf samples, and relatively low microbial biomass per
632 leaf that that must be first dislodged from tightly-adhered biofilms. Signal from messenger RNA
633 in metatranscriptome analysis is masked by abundant ribosomal RNA signal, leading to further
634 challenge. Because of the combination of all of these challenges, much of our understanding of
635 the phyllosphere, as the largest surface area of microbial habitation on Earth (Peñuelas and
636 Terradas, 2014), has been learned from studies that employ model hosts and synthetic or
637 model microbial communities in controlled settings, or from description of the community
638 structure by sequencing of marker genes, amplified and bioinformatically depleted of
639 chloroplast genes to overcome the challenges of low signal and host contamination.

640 Here, we report optimized laboratory protocols (to minimize host and chloroplast
641 signals) combined with a genome-centric bioinformatic approach to perform focused functional
642 gene and transcript analysis of seasonally dynamic yet persistent phyllosphere microbiome
643 members. To our knowledge, this is the first untargeted bacterial metatranscriptomic work
644 performed on the leaf phyllosphere of field-grown crops. Other recent leaf metatranscriptome
645 studies have focused the viral communities of tomato and pepper (Choi *et al.*, 2020), soybean
646 (Marzano and Domier, 2016), and rice (Chao *et al.*, 2020 Preprint). A key strength of this work
647 is the challenging integration of phyllosphere metagenome and metatranscriptome data,
648 leveraging the higher coverage of the metagenomes with the activity information available
649 from the metatranscriptomes. Despite the relatively limited coverage of the MAGs (due to
650 substantial host and ribosomal DNA contamination), the analysis proved successful by

651 integrating both datasets and focusing on genome-centric interpretation. Thus, there are likely
652 many more prevalent and functionally active populations of the phyllosphere that were not
653 captured in this study, including those players previously known to be key in the phyllosphere
654 (e.g. Delmotte *et al.*, 2009; Vorholt, 2012). Substantial additional sequencing effort or an
655 enrichment strategy would improve signal for a cultivation-independent approach to target
656 those players. While the use of genome-centric approaches has the obvious shortcoming that
657 we have obviously not captured every microbiome member, our approach does allow us to link
658 actively transcribed functions to specific microbial membership. Furthermore, the functional
659 genes and activities documented here are logical given current understanding of microbial
660 adaptation to the host and phyllosphere environment.

661 Overall, this work provides evidence of a thriving, dynamic, functionally diverse, leaf-
662 specialized, and host-responsive microbiome on the phyllosphere of perennial grasses. It
663 provides evidence of specific phyllosphere functions that are seasonally activated in a
664 temperate agroecosystem and suggests several hypotheses of important host-microbe
665 interactions in the phyllosphere, for example via central metabolism, isoprenoid biosynthesis,
666 and stress response engagements. This research contributes to our broad understanding of the
667 dynamics and activities of phyllosphere microbial communities, and points to specific microbial
668 functions to target that could prove useful for plant-microbiome management.

669

670 **Acknowledgements**

Support for this research was provided by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (Awards DE-SC0018409 and DE-FC02-07ER64494), by the National Science Foundation Long-term Ecological Research Program (DEB 1637653 and 1832042) at the Kellogg Biological Station, Michigan State University AgBioResearch, and by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (US Department of Energy, Office of Science, Office of Biological and Environmental Research under award number DE-SC0018420). This work was also supported in part by Michigan State University through computational resources provided by the Institute for Cyber-Enabled Research and in part by the University of Wisconsin-Madison Wisconsin Energy Institute as supported by GLBRC Information Services. The work (proposal:10.46936/10.25585/60000818) conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. AS acknowledges support by the USDA National Institute of Food and Agriculture and Michigan State University AgBioResearch. NS acknowledges support from the Michigan State University Plant Resilience Institute. We thank three anonymous reviewers for their thoughtful comments on a previous version of this work.

Author contributions

KLG and AS conceived and designed experiments; NS, KLG and AS performed the experiments; ACH, NS, SKD, FY and AS analyzed the data; ACH, NS, SKD, FY, KLG and AS contributed materials/analysis tools; and ACH, NS, SKD, FY, KLG and AS wrote the paper.

694 **Competing Interests statement**

695 The authors declare no competing interests.

696 **References**

697 Andrews, J.H. and Harris, R.F. (2000) The ecology and biogeography of microorganisms on plant
698 surfaces. *Annu Rev Phytopathol* **38**: 145–180.
699 Bahulikar, R.A., Torres-Jerez, I., Worley, E., Craven, K., and Udvardi, M.K. (2014) Diversity of
700 nitrogen-fixing bacteria associated with switchgrass in the native tallgrass prairie of
701 Northern Oklahoma. *Appl Environ Microbiol* **80**: 5636–5643.
702 Bao, L., Cai, W., Cao, J., Zhang, X., Liu, J., Chen, H., et al. (2020) Microbial community overlap
703 between the phyllosphere and rhizosphere of three plants from Yongxing Island, South
704 China Sea. *Microbiologyopen* **9**: 1–18.
705 Bay, S.K., Dong, X., Bradley, J.A., Leung, P.M., Grinter, R., Jirapanjawat, T., et al. (2021) Trace gas
706 oxidizers are widespread and active members of soil microbial communities. *Nat Microbiol*
707 **6**: 246–256.
708 Beilsmith, K., Thoen, M.P.M., Brachi, B., Gloss, A.D., Khan, M.H., and Bergelson, J. (2019)
709 Genome-wide association studies on the phyllosphere microbiome: Embracing complexity
710 in host–microbe interactions. *Plant J* **97**: 164–181.
711 Bell, T.H., Hockett, K.L., Alcalá-Briseño, R.I., Barbercheck, M., Beattie, G.A., Bruns, M.A., et al.
712 (2019) Manipulating wild and tamed phytobiomes: Challenges and opportunities.
713 *Phytobiomes J* **3**: 3–21.
714 Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., Van Wezel, G.P., Medema, M.H., and
715 Weber, T. (2021) AntiSMASH 6.0: Improving cluster detection and comparison capabilities.
716 *Nucleic Acids Res* **49**: W29–W35.
717 Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: A flexible trimmer for Illumina
718 sequence data. *Bioinformatics* **30**: 2114–2120.
719 Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., et
720 al. (2017) Minimum information about a single amplified genome (MISAG) and a
721 metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**:
722 725–731.
723 Bowsher, A.W., Benucci, G.M.N., Bonito, G., and Shade, A. (2020) Seasonal Dynamics of Core
724 Fungi in the Switchgrass Phyllosphere, and Co-Occurrence with Leaf Bacteria. *Phytobiomes*
725 *J* PBIOMES-07-20-0.
726 Bringel, F. and Couée, I. (2015) Pivotal roles of phyllosphere microorganisms at the interface
727 between plant functioning and atmospheric trace gas dynamics. *Front Microbiol* **6**:.
728 Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F.,
729 et al. (2012) Revealing structure and assembly cues for *Arabidopsis* root-inhabiting
730 bacterial microbiota. *Nature* **488**: 91–95.
731 Bulgarelli, D., Schlaeppi, K., Spaepen, S., van Themaat, E.V.L., and Schulze-Lefert, P. (2013)
732 Structure and functions of the bacterial microbiota of plants. *Annu Rev Plant Biol* **64**: 807–
733 838.
734 Busby, P.E., Soman, C., Wagner, M.R., Friesen, M.L., Kremer, J., Bennett, A., et al. (2017)
735 Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS Biol*
736 **15**: 1–14.
737 Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., et
738 al. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per
739 sample. *Proc Natl Acad Sci* **108**: 4516–4522.

740 Carlström, C.I., Field, C.M., Bortfeld-Miller, M., Müller, B., Sunagawa, S., and Vorholt, J.A. (2019)
741 Synthetic microbiota reveal priority effects and keystone strains in the *Arabidopsis*
742 phyllosphere. *Nat Ecol Evol* **3**: 1445–1454.
743 Cavicchioli, R., Ripple, W.J., Timmis, K.N., Azam, F., Bakken, L.R., Baylis, M., et al. (2019)
744 Scientists' warning to humanity: microorganisms and climate change. *Nat Rev Microbiol*
745 **17**: 569–586.
746 Chao, S., Wang, H., Yan, Q., Chen, L., Chen, G., Wu, L., et al. (2020) Metatranscriptomics and
747 small RNA analysis revealed that viral covert coinfection resulted in disease symptoms
748 reminiscent of rice sterility. *Prepr available Res Sq*.
749 Chen, C., Li, S., McKeever, D.R., and Beattie, G.A. (2013) The widespread plant-colonizing
750 bacterial species Pseudomonas syringae detects and exploits an extracellular pool of
751 choline in hosts. *Plant J* **75**: 891–902.
752 Chen, T., Nomura, K., Wang, X., Sohrabi, R., Xu, J., Yao, L., et al. (2020) A plant genetic network
753 for preventing dysbiosis in the phyllosphere. *Nature* **580**: 653–657.
754 Choi, H., Jo, Y., Cho, W.K., Yu, J., Tran, P.T., Salaipeth, L., et al. (2020) Identification of viruses
755 and viroids infecting tomato and pepper plants in vietnam by metatranscriptomics. *Int J*
756 *Mol Sci* **21**: 1–16.
757 Considine, M.J. and Foyer, C.H. (2014) Redox regulation of plant development. *Antioxid Redox*
758 *Signal* **21**: 1305–1326.
759 Crombie, A.T., Larke-Mejia, N.L., Emery, H., Dawson, R., Pratscher, J., Murphy, G.P., et al. (2018)
760 Poplar phyllosphere harbors disparate isoprene-degrading bacteria. *Proc Natl Acad Sci U S*
761 *A* **115**: 13081–13086.
762 Dai, J., Dai, W., Qiu, C., Yang, Z., Zhang, Y., Zhou, M., et al. (2015) Unraveling adaptation of
763 *Pontibacter korlensis* to radiation and infertility in desert through complete genome and
764 comparative transcriptomic analysis. *Sci Rep* **5**: 1–9.
765 Delmotte, N., Knief, C., Chaffron, S., Innerebner, G., Roschitzki, B., Schlapbach, R., et al. (2009)
766 Community proteogenomics reveals insights into the physiology of phyllosphere bacteria.
767 *Proc Natl Acad Sci* **106**: 16428–16433.
768 Dorokhov, Y.L., Sheshukova, E. V., and Komarova, T. V. (2018) Methanol in plant life. *Front Plant*
769 *Sci* **871**: 1–6.
770 Druzhinina, I.S., Chenthamara, K., Zhang, J., Atanasova, L., Yang, D., Miao, Y., et al. (2018)
771 Massive lateral transfer of genes encoding plant cell wall-degrading enzymes to the
772 mycoparasitic fungus Trichoderma from its plant-associated hosts.
773 Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., Podishetty, N.K., Bhatnagar, S., et al.
774 (2015) Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc*
775 *Natl Acad Sci U S A* **112**: E911-20.
776 Egamberdieva, D., Wirth, S.J., Alqarawi, A.A., Abd-Allah, E.F., and Hashem, A. (2017)
777 Phytohormones and beneficial microbes: Essential components for plants to balance stress
778 and fitness. *Front Microbiol* **8**: 1–14.
779 Eller, A.S.D., Sekimoto, K., Gilman, J.B., Kuster, W.C., de Gouw, J.A., Monson, R.K., et al. (2011)
780 Volatile organic compound emissions from switchgrass cultivars used as biofuel crops.
781 *Atmos Environ* **45**: 3333–3337.
782 Eren, A.M., Esen, O.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., and Delmont, T.O.
783 (2015) Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ* **2015**:

1–29.
Ferner, E., Rennenberg, H., and Kreuzwieser, J. (2012) Effect of flooding on C metabolism of flood-tolerant (*Quercus robur*) and non-tolerant (*Fagus sylvatica*) tree species. *Tree Physiol* **32**: 135–145.
Gill, U.S., Nandety, R.S., Krom, N., Dai, X., Zhuang, Z., Tang, Y., et al. (2019) Draft genome sequence resource of switchgrass rust pathogen, puccinia novopanici isolate ard-01. *Phytopathology* **109**: 1513–1515.
Gopal, M. and Gupta, A. (2016) Microbiome selection could spur next-generation plant breeding strategies. *Front Microbiol* **7**: 1971.
Gostinčar, C., Ohm, R.A., Kogej, T., Sonjak, S., Turk, M., Zajc, J., et al. (2014) Genome sequencing of four Aureobasidium pullulans varieties: Biotechnological potential, stress tolerance, and description of new species. *BMC Genomics* **15**:.
Gottschlich, L., Geiser, P., Bortfeld-Miller, M., Field, C.M., and Vorholt, J.A. (2019) Complex general stress response regulation in *Sphingomonas melonis* Fr1 revealed by transcriptional analyses. *Sci Rep* **9**: 1–13.
Grady, K.L., Sorensen, J.W., Stopnisek, N., Guittar, J., and Shade, A. (2019) Assembly and seasonality of core phyllosphere microbiota on perennial biofuel crops. *Nat Commun* **10**:.
Hacquard, S., Garrido-Oter, R., González, A., Spaepen, S., Ackermann, G., Lebeis, S., et al. (2015) Microbiota and host nutrition across plant and animal kingdoms. *Cell Host Microbe* **17**: 603–616.
Hardoim, P.R., van Overbeek, L.S., Berg, G., Pirttilä, A.M., Compant, S., Campisano, A., et al. (2015) The hidden world within plants: Ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiol Mol Biol Rev* **79**: 293–320.
Haridas, S., Albert, R., Binder, M., Bloem, J., LaButti, K., Salamov, A., et al. (2020) 101 Dothideomycetes genomes: A test case for predicting lifestyles and emergence of pathogens. *Stud Mycol* **96**: 141–153.
Harty, C.E., Martins, D., Doing, G., Mould, D.L., Clay, M.E., Occhipinti, P., et al. (2019) Ethanol stimulates trehalose production through a SpoT-DksA-AlgU-dependent pathway in *Pseudomonas aeruginosa*. 1–21.
Haskett, T.L., Tkacz, A., and Poole, P.S. (2020) Engineering rhizobacteria for sustainable agriculture. *ISME J* **15**: 949–964.
Heaton, E.A., Dohleman, F.G., and Long, S.P. (2008) Meeting US biofuel goals with less land: The potential of Miscanthus. *Glob Chang Biol* **14**: 2000–2014.
Hestrin, R., Lee, M.R., Whitaker, B.K., and Pett-Ridge, J. (2021) The switchgrass microbiome: A review of structure, function, and taxonomic distribution. *Phytobiomes J* **5**: 14–28.
Julsing, M.K., Rijpkema, M., Woerdenbag, H.J., Quax, W.J., and Kayser, O. (2007) Functional analysis of genes involved in the biosynthesis of isoprene in *Bacillus subtilis*. *Appl Microbiol Biotechnol* **75**: 1377–1384.
Kałużna, M., Willems, A., Pothier, J.F., Ruinelli, M., Sobiczewski, P., and Puławska, J. (2016) Pseudomonas cerasi sp. nov. (non Griffin, 1911) isolated from diseased tissue of cherry. *Syst Appl Microbiol* **39**: 370–377.
Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019) MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **2019**: 1–13.

828 Kerchev, P., De Smet, B., Waszczak, C., Messens, J., and Van Breusegem, F. (2015) Redox
829 strategies for crop improvement. *Antioxidants Redox Signal* **23**: 1186–1205.
830 El Khawand, M., Crombie, A.T., Johnston, A., Vavlline, D. V., McAuliffe, J.C., Latone, J.A., et al.
831 (2016) Isolation of isoprene degrading bacteria from soils, development of isoA gene
832 probes and identification of the active isoprene-degrading soil community using DNA-
833 stable isotope probing. *Environ Microbiol* **18**: 2743–2753.
834 Kimmerer, T.W. and Kozlowski, T.T. (1982) Ethylene, ethane, acetaldehyde, and ethanol
835 production by plants under stress. *Plant Physiol* **69**: 840–847.
836 Kimmerer, T.W. and MacDonald, R.C. (1987) Acetaldehyde and ethanol biosynthesis in leaves of
837 plants. *Plant Physiol* **84**: 1204–1209.
838 Koskella, B. (2020) The phyllosphere. *Curr Biol* **30**: R1143–R1146.
839 Kuzyakov, Y. and Razavi, B.S. (2019) Rhizosphere size and shape: Temporal dynamics and spatial
840 stationarity. *Soil Biol Biochem* **135**: 343–360.
841 Lacerda-Júnior, G. V., Noronha, M.F., Cabral, L., Delforno, T.P., De Sousa, S.T.P., Fernandes-
842 Júnior, P.I., et al. (2019) Land use and seasonal effects on the soil microbiome of a Brazilian
843 dry forest. *Front Microbiol* **10**: 1–14.
844 Lajoie, G., Maglione, R., and Kembel, S.W. (2020) Adaptive matching between phyllosphere
845 bacteria and their tree hosts in a neotropical forest. *Microbiome* **8**: 1–10.
846 Langholtz, M., Stokes, B., and Eaton, L. (2016) 2016 billion-ton report: Advancing domestic
847 resources for a thriving bioeconomy (Executive Summary). *Ind Biotechnol* **12**: 282–289.
848 Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*
849 **9**: 357–359.
850 Laskowska, E., Dorota Kuczyńska-Wiśnik, ·, and Kuczyńska-Wiśnik, D. (2020) New insight into
851 the mechanisms protecting bacteria during desiccation. *Curr Genet* **66**: 313–318.
852 Lee, M.D. (2019) GToTree: A user-friendly workflow for phylogenomics. *Bioinformatics* **35**:
853 4162–4164.
854 Levy, A., Conway, J.M., Dangl, J.L., and Woyke, T. (2018) Elucidating bacterial gene functions in
855 the plant microbiome. *Cell Host Microbe* **24**: 475–485.
856 Li, D., Liu, C.M., Luo, R., Sadakane, K., and Lam, T.W. (2015) MEGAHIT: An ultra-fast single-node
857 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
858 *Bioinformatics* **31**: 1674–1676.
859 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009) The Sequence
860 Alignment / Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
861 Lindow, S.E. and Brandl, M.T. (2003) Microbiology of the phyllosphere. *Appl Environ Microbiol*
862 **69**: 1875–1883.
863 Liu, Y., Zhang, X., Tran, H., Shan, L., Kim, J., Childs, K., et al. (2015) Assessment of drought
864 tolerance of 49 switchgrass (Panicum virgatum) genotypes using physiological and
865 morphological parameters. *Biotechnol Biofuels* **8**: 1–18.
866 Lucke, M., Correa, M.G., and Levy, A. (2020) The role of secretion systems, effectors, and
867 secondary metabolites of beneficial rhizobia in interactions with plants and microbes.
868 *Front Plant Sci* **11**:.
869 Lundberg, D.S., Lebeis, S.L., Paredes, S.H., Yourstone, S., Gehring, J., Malfatti, S., et al. (2012)
870 Defining the core *Arabidopsis thaliana* root microbiome. *Nature* **488**: 86–90.
871 Ma, L., Rocha, F.I., Lee, J., Choi, J., Tejera, M., Sooksa-Nguan, T., et al. (2021) The impact of

872 stand age and fertilization on the soil microbiome of *Miscanthus × giganteus*. *Phytobiomes*
873 *J* **5**: 51–59.
874 Marzano, S.Y.L. and Domier, L.L. (2016) Novel mycoviruses discovered from
875 metatranscriptomics survey of soybean phyllosphere phytobiomes. *Virus Res* **213**: 332–
876 342.
877 McGenity, T.J., Crombie, A.T., and Murrell, J.C. (2018) Microbial cycling of isoprene, the most
878 abundantly produced biological volatile organic compound on Earth. *ISME J* **12**: 931–941.
879 Morrison, E.C., Drewer, J., and Heal, M.R. (2016) A comparison of isoprene and monoterpene
880 emission rates from the perennial bioenergy crops short-rotation coppice willow and
881 Miscanthus and the annual arable crops wheat and oilseed rape. *GCB Bioenergy* **8**: 211–
882 225.
883 Müller, D.B., Vogel, C., Bai, Y., and Vorholt, J.A. (2016) The Plant Microbiota: Systems-Level
884 Insights and Perspectives. *Annu Rev Genet* **50**: 120215–034952.
885 Navarro-Muñoz, J.C., Selem-Mojica, N., Mullowney, M.W., Kautsar, S.A., Tryon, J.H., Parkinson,
886 E.I., et al. (2020) A computational framework to explore large-scale biosynthetic diversity.
887 *Nat Chem Biol* **16**: 60–68.
888 Nayfach, S., Shi, Z.J., Seshadri, R., Pollard, K.S., and Kyrpides, N.C. (2019) New insights from
889 uncultivated genomes of the global human gut microbiome. *Nature* **568**: 505–510.
890 Nowicka, B. and Kruk, J. (2010) Occurrence, biosynthesis and function of isoprenoid quinones.
891 *Biochim Biophys Acta - Bioenerg* **1797**: 1587–1605.
892 Palmer, J.L., Hilton, S., Picot, E., Bending, G.D., and Schäfer, H. (2021) Tree phyllospheres are a
893 habitat for diverse populations of CO-oxidizing bacteria. *Environ Microbiol* **23**: 6309–6327.
894 Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A., and
895 Hugenholtz, P. (2018) A standardized bacterial taxonomy based on genome phylogeny
896 substantially revises the tree of life. *Nat Biotechnol* **36**: 996.
897 Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM:
898 Assessing the quality of microbial genomes recovered from isolates, single cells, and
899 metagenomes. *Genome Res* **25**: 1043–1055.
900 Peñuelas, J. and Terradas, J. (2014) The foliar microbiome. *Trends Plant Sci* **19**: 278–280.
901 Quinlan, A.R. and Hall, I.M. (2010) BEDTools : a flexible suite of utilities for comparing genomic
902 features. **26**: 841–842.
903 Rastogi, G., Coaker, G.L., and Leveau, J.H.J. (2013) New insights into the structure and function
904 of phyllosphere microbiota through high-throughput molecular approaches. *FEMS*
905 *Microbiol Lett* **348**: 1–10.
906 Remus-Emsermann, M.N.P. and Schlechter, R.O. (2018) Phyllosphere microbiology: at the
907 interface between microbial individuals and the plant host. *New Phytol* **218**: 1327–1333.
908 Robertson, G.P., Hamilton, S.K., Barham, B.L., Dale, B.E., Izaurralde, R.C., Jackson, R.D., et al.
909 (2017) Cellulosic biofuel contributions to a sustainable energy future: Choices and
910 outcomes. *Science (80-)* **356**: 1–9.
911 Roley, S.S., Duncan, D.S., Liang, D., Garoutte, A., Jackson, R.D., Tiedje, J.M., and Robertson, G.P.
912 (2018) Associative nitrogen fixation (ANF) across a nitrogen input gradient. *PLoS One* **13**:
913 1–19.
914 Roley, S.S., Xue, C., Hamilton, S.K., Tiedje, J.M., and Robertson, G.P. (2019) Isotopic evidence for
915 episodic nitrogen fixation in switchgrass (*Panicum virgatum* L.). *Soil Biol Biochem* **129**: 90–

98.
Shade, A. and Stopnisek, N. (2019) Abundance-occupancy distributions to prioritize plant core microbiome membership. *Curr Opin Microbiol* **49**: 50–58.
Shaffer, M., Borton, M.A., McGivern, B.B., Zayed, A.A., La Rosa, S.L. 0003 3527 8101, Solden, L.M., et al. (2020) DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* **48**: 8883–8900.
Sharkey, T.D., Loreto, F., and Delwiche, C. (1991) High carbon dioxide and sun/shade effects on isoprene emission from oak and aspen tree leaves. *Plant, Cell Environ* **14**: 333–338.
Sharkey, T.D., Wiberley, A.E., and Donohue, A.R. (2008) Isoprene emission from plants: Why and how. *Ann Bot* **101**: 5–18.
Sharkey, T.D. and Yeh, S. (2001) Isoprene emission from plants. *Plant Mol Biol* **52**: 407–436.
Singer, E., Bonnette, J., Kenaley, S.C., Woyke, T., and Juenger, T.E. (2019) Plant compartment and genetic variation drive microbiome composition in switchgrass roots. *Environ Microbiol Rep* **11**: 185–195.
Sivy, T.L., Shirk, M.C., and Fall, R. (2002) Isoprene synthase activity parallels fluctuations of isoprene release during growth of *Bacillus subtilis*. *Biochem Biophys Res Commun* **294**: 71–75.
Spaepen, S., Vanderleyden, J., and Remans, R. (2007) Indole-3-acetic acid in microbial and microorganism-plant signaling. *FEMS Microbiol Rev* **31**: 425–448.
Stopnisek, N. and Shade, A. (2021) Persistent microbiome members in the common bean rhizosphere: an integrated analysis of space, time, and plant genotype. *ISME J*.
Suzuki, R. and Shimodaira, H. (2006) Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**: 1540–1542.
Suzuki, Y., Makino, A., and Mae, T. (2001) An efficient method for extraction of RNA from rice leaves at different ages using benzyl chloride. *J Exp Bot* **52**: 1575–1579.
Swan, B.K., Tupper, B., Sczyrba, A., Lauro, F.M., Martinez-Garcia, M., Gonźalez, J.M., et al. (2013) Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci U S A* **110**: 11463–11468.
Toju, H., Peay, K.G., Yamamichi, M., Narisawa, K., Hiruma, K., Naito, K., et al. (2018) Core microbiomes for sustainable agroecosystems. *Nat Plants* **4**: 247–257.
Trivedi, P., Leach, J.E., Tringe, S.G., Sa, T., and Singh, B.K. (2020) Plant–microbiome interactions: from community assembly to plant health. *Nat Rev Microbiol* **18**: 607–621.
Tseng, T.T., Tyler, B.M., and Setubal, J.C. (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol* **9**: 1–9.
Urrejola, C., Alcorta, J., Salas, L., Vásquez, M., Polz, M.F., Vicuña, R., and Díez, B. (2019) Genomic features for desiccation tolerance and sugar biosynthesis in the extremophile *gloeocapsopsis* sp. UTEX B3054. *Front Microbiol* **10**: 1–11.
Valenzuela-Soto, E.M. and Figueroa-Soto, C.G. (2019) Biosynthesis and Degradation of Glycine Betaine and Its Potential to Control Plant Growth and Development. In *Osmoprotectant-Mediated Abiotic Stress Tolerance in Plants*. Anwar Hossain, M., Kumar, Vi., Burritt, D.J., Fujita, M., and Makela, P.S.A. (eds). Springer, pp. 241–256.
Vorholt, J.A. (2012) Microbial life in the phyllosphere. *Nat Rev Microbiol* **10**: 828–840.
Wang, N.R. and Haney, C.H. (2020) Harnessing the genetic potential of the plant microbiome. *Biochem (Lond)* **42**: 20–25.

960 Wingler, A., Delatte, T.L., O'Hara, L.E., Primavesi, L.F., Jhurreea, D., Paul, M.J., and
961 Schluepmann, H. (2012) Trehalose 6-phosphate is required for the onset of leaf
962 senescence associated with high carbon availability. *Plant Physiol* **158**: 1241–1251.
963 Xu, L., Naylor, D., Dong, Z., Simmons, T., Pierroz, G., Hixson, K.K., et al. (2018) Drought delays
964 development of the sorghum root microbiome and enriches for monoderm bacteria. *Proc*
965 *Natl Acad Sci* **115**: E4284–E4293.
966 Zhalnina, K., Louie, K.B., Hao, Z., Mansoori, N., da Rocha, U.N., Shi, S., et al. (2018) Dynamic root
967 exudate chemistry and microbial substrate preferences drive patterns in rhizosphere
968 microbial community assembly. *Nat Microbiol* **3**: 470.
969 Zimmermann, J., Kaleta, C., and Waschina, S. (2021) Gapseq: informed prediction of bacterial
970 metabolic pathways and reconstruction of accurate metabolic models. *Genome Biol* **22**: 1–
971 35.
972 Zou, H., Chen, N., Shi, M., Xian, M., Song, Y., and Liu, J. (2016) The metabolism and
973 biotechnological application of betaine in microorganism. *Appl Microbiol Biotechnol* **100**:
974 3865–3876.
975 Zuo, Z., Weraduwage, S.M., Lantz, A.T., Sanchez, L.M., Weise, S.E., Wang, J., et al. (2019)
976 Isoprene acts as a signaling molecule in gene networks important for stress responses and
977 plant growth. *Plant Physiol* **180**: 124–152.
978
979

980 **Figures and Tables**
981
982 Figures
983
984 **Figure 1.** Phyllosphere microbiome field sampling strategy at the Great Lakes Bioenergy
985 Research Center Bioenergy Cropping System Experiment (BCSE) in 2016 and 2017. (A) The
986 study site is at Kellogg Biological Station, a Long-Term Ecological Research site focused on
987 agroecosystems located in southwest Michigan. (B) Four replicate randomized cropping system
988 blocks from the BCSE were sampled at each time point for switchgrass and/or miscanthus, and
989 within each plot there was a fertilized main plot and unfertilized subplot sampled. (C) In 2016,
990 both switchgrass and miscanthus were sampled, and in 2017 only switchgrass was sampled.
991 Switchgrass leaves were flash frozen in liquid nitrogen for RNA extraction and
992 metatranscriptome analysis at a subset of time points in 2016 and at all points in 2017.
993
994 **Figure 2.** Overview of bioinformatic processing of the metagenome (green solid arrows) and
995 metatranscriptome (red solid arrows) datasets. Switchgrass reads are shown in light green and
996 miscanthus is dark green. The solid grey arrow represents the common analysis step for both
997 datasets. Figure was made with Lucidchart.
998
999 **Figure 3**. Summary of leaf-associated MAGs. **(A)** Abundance and occupancy of genomes
1000 assembled and binned from switchgrass and miscanthus phyllosphere metagenomes. Quality
1001 and contamination assessment were determined using checkM. 40 focal MAGs were selected
1002 **(B)** Taxonomy of focal MAGs, as annotated with GTDB-tk, and taxonomic overlap with the
1003 persistent taxa detected in our previous 16S rRNA gene amplicon survey.
1004
1005 **Figure 4**. Seasonal patterns of the 40 focal MAG metagenome and metatranscriptome read
1006 recruitment. MAG abundances for: (**A**) Miscanthus 2016 metagenomes (metaG); (**B**)
1007 Switchgrass 2016 metagenomes; (**C**) Switchgrass 2017 metagenomes; (**D**) Switchgrass 2016
1008 metatranscriptomes (metaT) and (E) Switchgrass 2017 metatranscriptomes. Abundances of
1009 metagenome contigs were estimated with the median basepair of read recruitment divided by
1010 the average median basepair coverage of housekeeping genes. Abundances of
1011 metatranscriptome ORFs were estimated based on the median basepair coverage of all reads
1012 mapped to ORFs and divided by median basepair coverage of housekeeping genes. The same
1013 dendrogram is applied to all panel, and it is the result of hierarchical clustering (see Figure S2)
1014 of metatranscriptome diversity and abundances in switchgrass.
1015
1016 **Figure 5**. 2016 (circle) and 2017 (triangle) switchgrass leaf transcript dynamics of KEGG
1017 metabolism classifications to the 40 focal MAGs. The y-axis is scaled for each classification.
1018
1019 **Figure 6**. Key functional gene pathways detected in the 40 focal MAGs and their activities
1020 (mapped transcripts) during the 2016-2017 switchgrass growing season. Functional gene
1021 pathways were curated using antiSMASH for biosynthetic gene clusters, gapseq for general
1022 metabolic pathways, and manual selection for plant-associative functions reported in the
1023 literature. Pathways that were discovered in the MAGs but not detected in the transcripts are

1024 represented by open circles, and pathways detected in the MAGs and mapped by transcripts
1025 are represented by filled circles. Colors categorize different functional groups of pathways.
1026
1027 **Figure 7**. **(A).** 2016 (circle) and 2017 (triangle) switchgrass leaf transcript dynamics of KEGG
1028 metabolism classifications associated with terpene metabolism. **(B).** Transcripts in MAGs
1029 associated with terminal enzymes in the non-mevalonate isoprene biosynthesis, *gcpE* and *lytB*.
1030 MAG IDs include predicted taxonomy at the genus level: Methyl = *Methylobacterium*, Frigor =
1031 *Frigoribacterium*, Pseudokineo = *Pseudokineococcus*, Microbac = *Microbacterium*, Amnibac =
1032 *Amnibacterium*, Hymeno = *Hymenobacter,* Sphingo = *Sphingomonas,* and Pseudo =
1033 *Pseudomonas*.
1034
1035 **Figure 8**. Detection of 36 focal MAGs (out of 40) in publicly available metagenomes and
1036 metatranscriptomes of bioenergy grasses.
1037
1038

1039 Tables
1040
1041 **Table 1**. Summary of RNA and DNA samples that returned reads and passed Illumina standard
1042 quality control at the Joint Genome Institute. Total collected samples submitted for sequencing
1043 is provided first, and the number of quality sequencing datasets returned for analysis is given in
1044 parentheses.

	2016		2017	
	Metagenome	Metatranscriptome	Metagenome	Metatranscriptome
Switchgrass	8 timepoints (64/64 successful)	3 (22/24 successful)	7 (56/56 successful)	7 (56/56 successful)
Miscanthus	9 (72/72 successful)	Not assessed	Not assessed	Not assessed

1045
1046
1047

1048 **Table 2.** Functional roles that exhibited strong seasonality with transcript enrichment by more
1049 than 20-fold in the late season than early. These roles were consistently detected among
1050 phyllosphere focal MAGs (39/40 detections).
1051

Functional role	*Major classification(s)*	*No. features*	*Late:early ratio*
short chain dehydrogenase [PF00106.26]; Enoyl-(Acyl carrier protein) reductase [PF13561.7]; KR domain [PF08659.11]	Carbohydrate metabolism, Lipid metabolism, nucleotide metabolism, metabolism of cofactors and vitamins	4	53.2
Molybdopterin oxidoreductase [PF00384.23], Molydopterin dinucleotide binding domain [PF01568.22]	Carbohydrate metabolism, metabolism of cofactors and vitamins, Energy metabolism, Signal transduction	4	29.7
Polyketide cyclase / dehydrase and lipid transport [PF10604.10]	Energy metabolism	1	23.6

1052

1053 **Datasets**
1054 **Dataset 1.** Metadata for metagenome samples according to MIMs standards.
1055 **Dataset 2.** Metadata for metatranscriptome samples according to MIMs standards.
1056 **Dataset 3.** Excel file. *Sheet 1.* Metadata for selected focal 40 metagenome-assembled genomes
1057 (>50% complete, <10% contamination) according to MiMAG standards *Sheet 2.* Other MAGs
1058 (non-focal) assembled from switchgrass or miscanthus metagenomes.
1059 **Dataset 4**. Excel file. *Sheet 1.* Summary of metagenome reads mapped to the focal MAGs.
1060 *Sheet 2.* Summary of metatranscriptome reads mapped to the focal MAGs. *Sheet 3*. Total reads
1061 mapped from each metagenome. *Sheet 4.* Total reads mapped from each metatranscriptome.
1062 **Dataset 5.** Public metagenomes used for comparison to the switchgrass and miscanthus focal
1063 MAGs.
1064 **Dataset 6.** Core KEGG functions identified in at least 39 focal MAGs and their estimated
1065 abundances in early and late season.
1066
1067

1068 **Supplementary Figures**
1069 **Figure S1**. Contaminating plant or fungal sequences. Proportion of sequencing in metagenomes
1070 originating from miscanthus and switchgrass phyllospheres associated with miscanthus and
1071 switchgrass host genomes (A) and prevalent fungal genomes (B, Table S1). Proportion of
1072 sequencing in metatreanscriptomes originating from and switchgrass phyllospheres associated
1073 with switchgrass host genomes (C) and fungal genomes (D, Table S1).
1074
1075 **Figure S2.** Hierarchical clustering identified five clusters of MAGS to identify MAG populations
1076 with coherent seasonal activity dynamics. Clustering was based on metatranscriptome diversity
1077 and abundances.
1078
1079 **Figure S3.** Mean transcript dynamics by month and year for each MAG cluster. Filled circles
1080 indicate that identification of transcripts, and empty circles indicate no detection.
1081
1082 **Figure S4**. Summary of transcript seasonality of phyllosphere open reading frames (ORFs) that
1083 could be annotated as KEGG functional roles and were consistently detected among focal MAGs
1084 (at least 39/40 detections). Ratios are late-to-early normalized transcript abundances on
1085 MAGs.
1086
1087 **Figure S5**. 2016 (circle) and 2017 (triangle) switchgrass leaf transcript dynamics of KEGG
1088 metabolism classifications associated with terpenoid backbone biosynthesis.
1089
1090
1091
1092

1093
1094 **Supplementary Tables**
1095
1096 **Table S1**. Fungal genomes used for filtering metagenome reads to remove eukaryotic
1097 contamination. Genomes were selected to use for filtering based on the taxonomic identities
1098 of prevalent fungal taxa detected in our previous ITS2 amplicon survey that was conducted at
1099 the same location (Bowsher *et al.*, 2020).
1100

Fungal Genome	**JGI Project ID (*) or GenBank accession (**)**	**Reference**
Trichoderma harzianum	403727*	(Druzhinina et al., 2018)
Pleomassaria siparia	1011309*	(Haridas et al., 2020)
Aureobasidium pullulans	403628*	(Gostinčar et al., 2014)
Alternaria alternata	1103683*	NA
Didymella zeae-maydis	https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=Didma1	NA
Sporobolomyces roseus	16892*	NA
Puccinia novopanici	GCA_004348175.1**	(Gill et al., 2019)

1101
1102
1103
1104

1105 **Table S2.**
1106
1107 Summary of the nine genes involved the *Bacillus subtilus* isoprene biosynthesis pathway that
1108 were detected among focal MAG contigs. Except for *ypgA*, these genes were directly linked to
1109 isoprene accumulation by (Julsing *et al.*, 2007).
1110

Gene	*dxr*	*dxs*	*gcpE*	*idi*	*lytB*	*yacM*	*ychB*	*ygbB*	*yqiD*
NCBI Acc. No.	NP_389537.2	NP_390307.1	NP_390386.1	BAB32625.1	NP_390395.2	NP_387971.1	NP_387927.1	NP_387972.1	NP_390308.2
M94	+	+	+	+	+		+	+	+
S30	+	+	+	+	+		+	+	+
M105	+	+	+		+		+	+	+
M12	+	+	+	+	+			+	+
M32	+	+	+		+		+	+	+
M52	+	+	+	+	+			+	+
M67	+	+	+		+	+	+		+
M77	+	+	+		+		+	+	+
S117		+	+	+	+		+	+	+
S28	+	+	+		+		+	+	+
S56	+	+	+	+			+	+	+
M100	+	+	+		+	+			+
M109	+	+	+		+			+	+
M1	+	+			+		+	+	+
M21	+	+	+		+		+		+
M44	+	+		+	+		+		+
M47		+	+		+	+		+	+
M60		+	+		+		+	+	+
M66	+	+	+				+	+	+
M86	+	+	+		+		+		+
M8		+	+		+		+	+	+
M99	+	+			+	+		+	+
M9	+	+	+		+			+	+
S27	+	+	+		+			+	+
S61	+	+	+		+			+	+
S71	+	+			+		+	+	+
S74	+	+	+		+	+			+
M102	+	+	+					+	+
M55	+				+		+	+	+
S80		+	+		+	+			+
M111			+	+	+				+
M87	+	+						+	+
S120		+	+		+		+		
S29		+					+	+	+
S50	+	+	+				+		
S36		+	+		+				
M17		+							+
M35		+							+
S8		+							+
S9		+							+

1111
1112
1113
1114

Figure 1

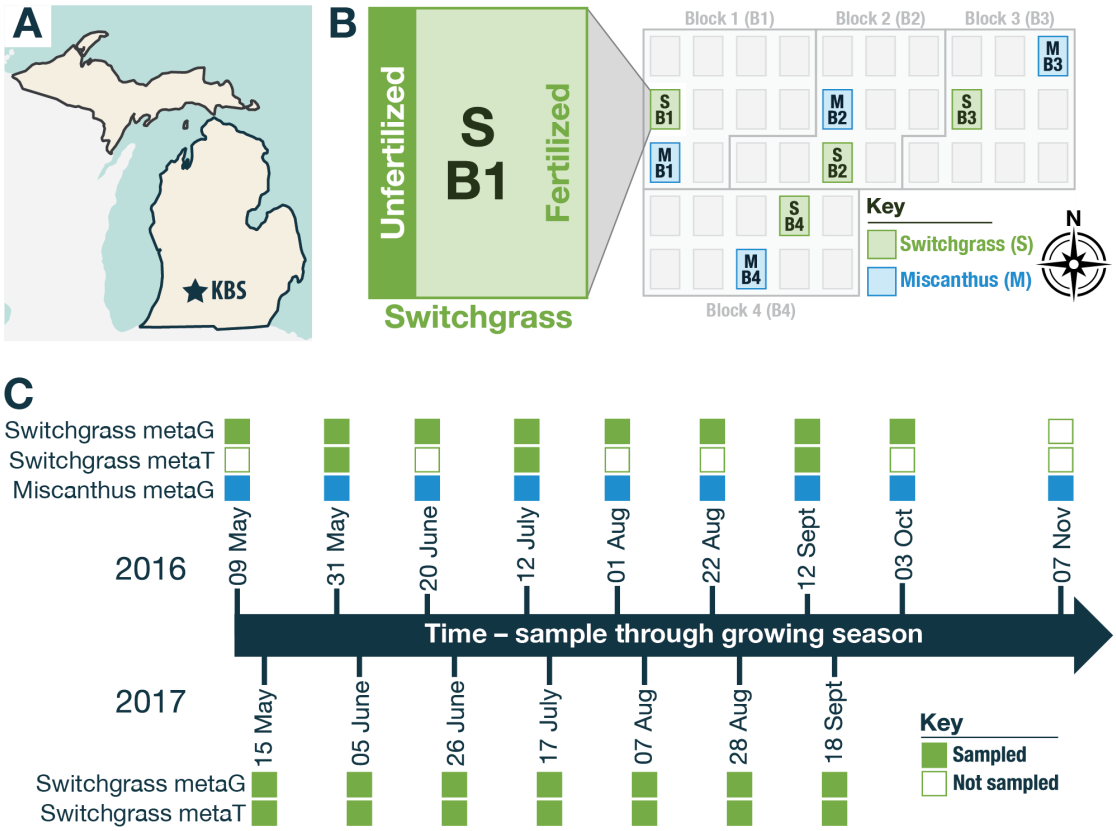

Figure 1. Phyllosphere microbiome field sampling strategy at the Great Lakes Bioenergy Research Center Bioenergy Cropping System Experiment (BCSE) in 2016 and 2017. (A) The study site is at Kellogg Biological Station, a Long-Term Ecological Research site focused on agroecosystems located in southwest Michigan. (B) Four replicate randomized cropping system blocks from the BCSE were sampled at each time point for switchgrass and/or miscanthus, and within each plot there was a fertilized main plot and unfertilized subplot sampled. (C) In 2016, both switchgrass and miscanthus were sampled, and in 2017 only switchgrass was sampled. Switchgrass leaves were flash frozen in liquid nitrogen for RNA extraction and metatranscriptome analysis at a subset of time points in 2016 and at all points in 2017.

Figure 2

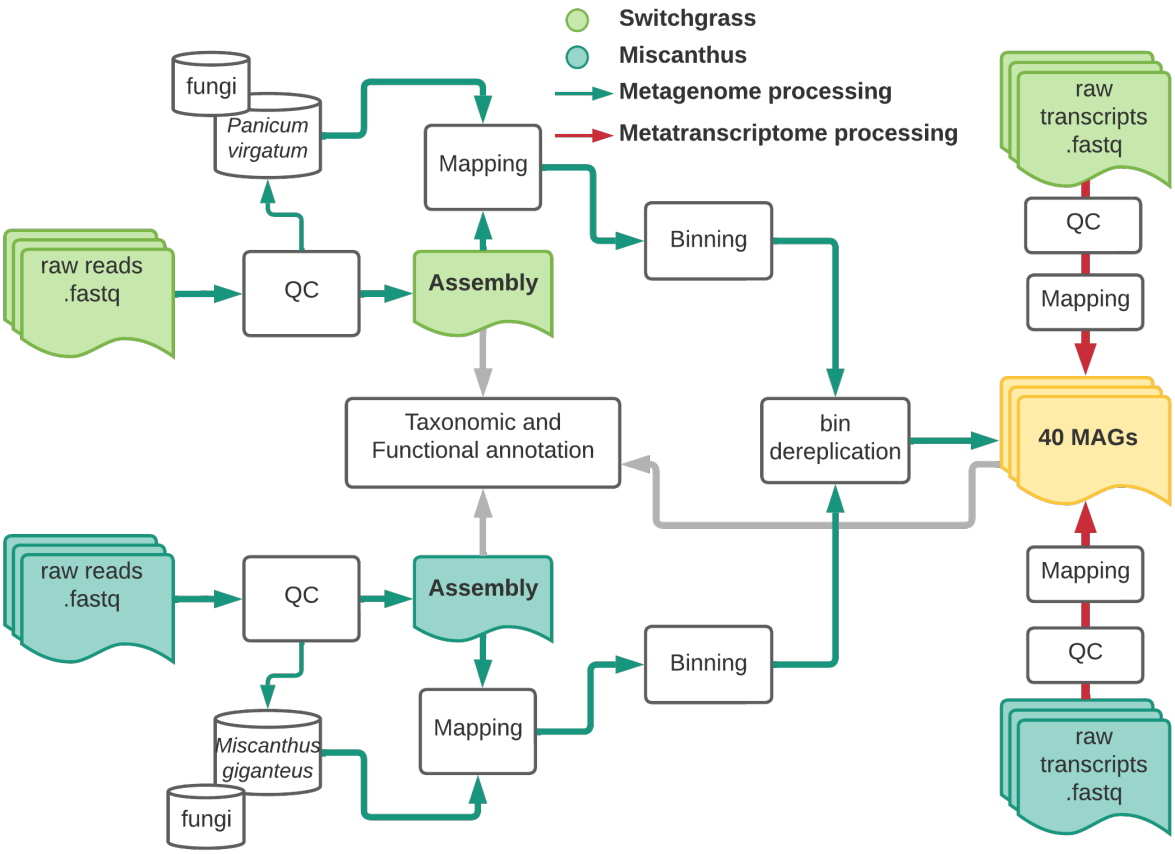

Figure 2. Overview of bioinformatic processing of the metagenome (green solid arrows) and metatranscriptome (red solid arrows) datasets. Switchgrass reads are shown in light green and miscanthus is dark green. The solid grey arrow represents the common analysis step for both data types. Figure was made with Lucidchart.

Figure 3

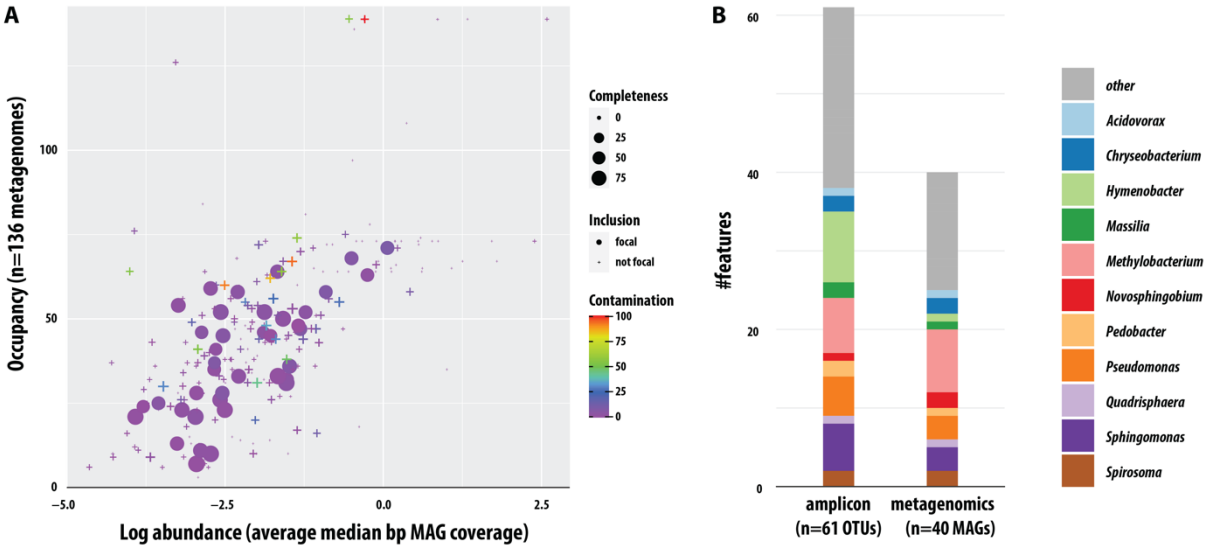

Figure 3. Summary of leaf-associated MAGs. **(A)** Abundance and occupancy of genomes assembled and binned from switchgrass and miscanthus phyllosphere metagenomes. Quality and contamination assessment were determined using checkM. 40 focal MAGs were selected **(B)** Taxonomy of focal MAGs, as annotated with GTDB-tk, and taxonomic overlap with the persistent taxa detected in our previous 16S rRNA gene amplicon survey.

Figure 4

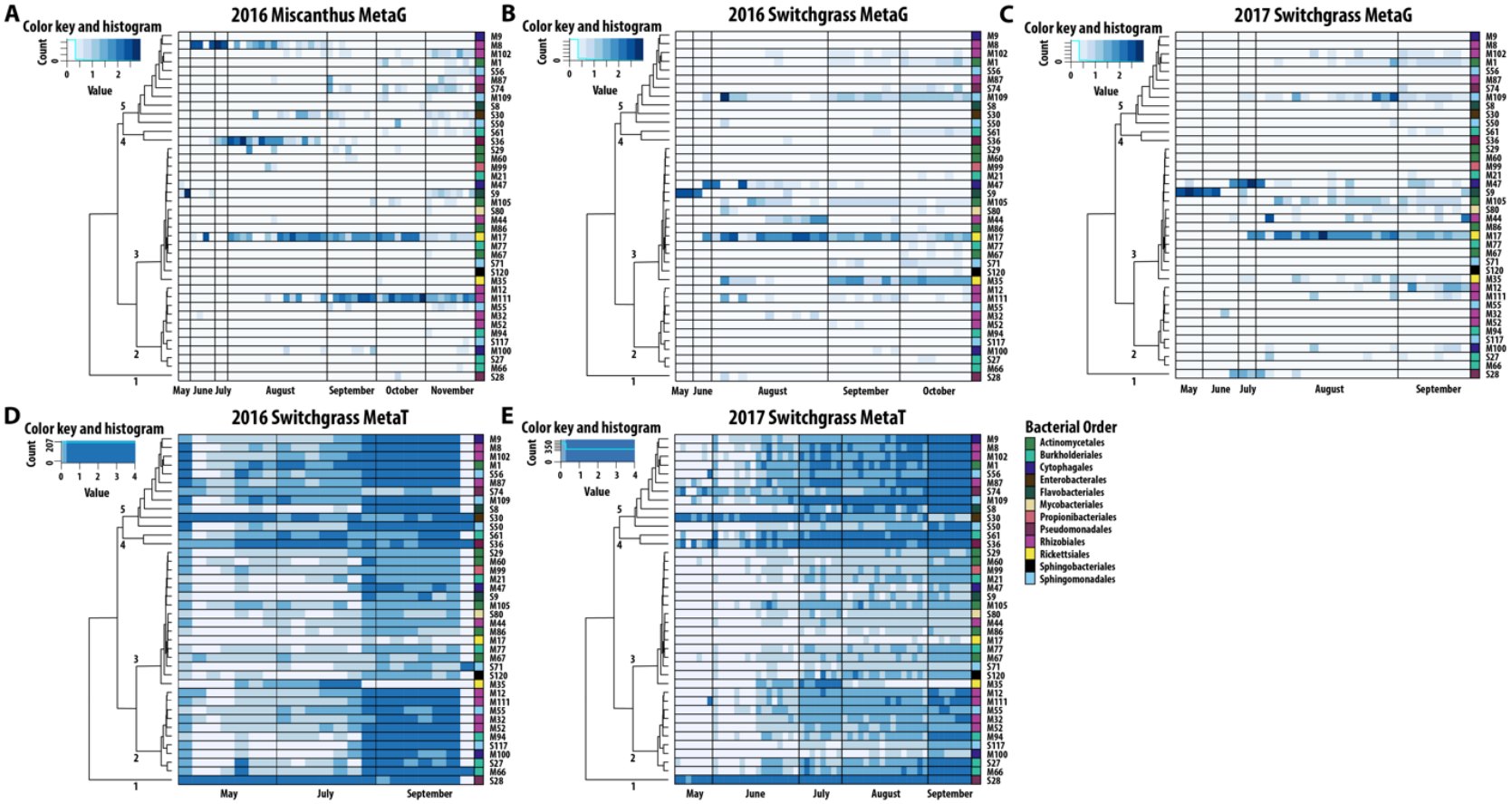

Figure 4. Seasonal patterns of the 40 focal MAG metagenome and metatranscriptome read recruitment. MAG abundances for: (**A**) Miscanthus 2016 metagenomes; (**B**) Switchgrass 2016 metagenomes (metaG); (**C**) Switchgrass 2017 metagenomes; (**D**) Switchgrass 2016 metatranscriptomes (metaT) and (E) Switchgrass 2017 metatranscriptomes. Abundances of metagenome contigs were estimated with the median basepair of read recruitment divided by the average median basepair coverage of housekeeping genes. Abundances of metatranscriptome ORFs were estimated based on the median basepair coverage of all reads mapped to ORFs and divided by median basepair coverage of housekeeping genes. The same dendrogram is applied to all panel, and it is the result of hierarchical clustering (see Figure S2) of metatranscriptome diversity and abundances in switchgrass.

Figure 5

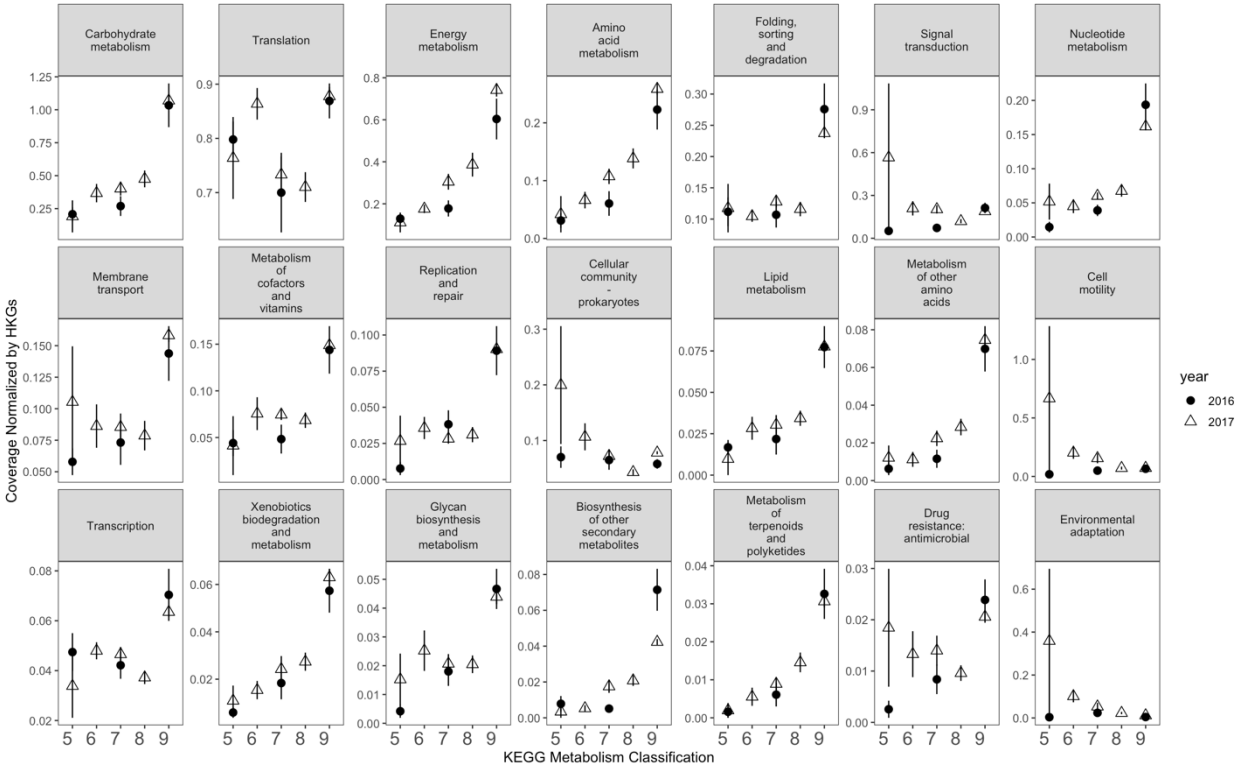

Figure 5. 2016 (circle) and 2017 (triangle) switchgrass leaf transcript dynamics of KEGG metabolism classifications to the 40 focal MAGs. The y-axis is scaled for each classification. Coverage was estimated as the average median base pair coverage for each gene. All coverage estimates were normalized by estimated coverage of housekeeping genes in each sample.

Figure 6

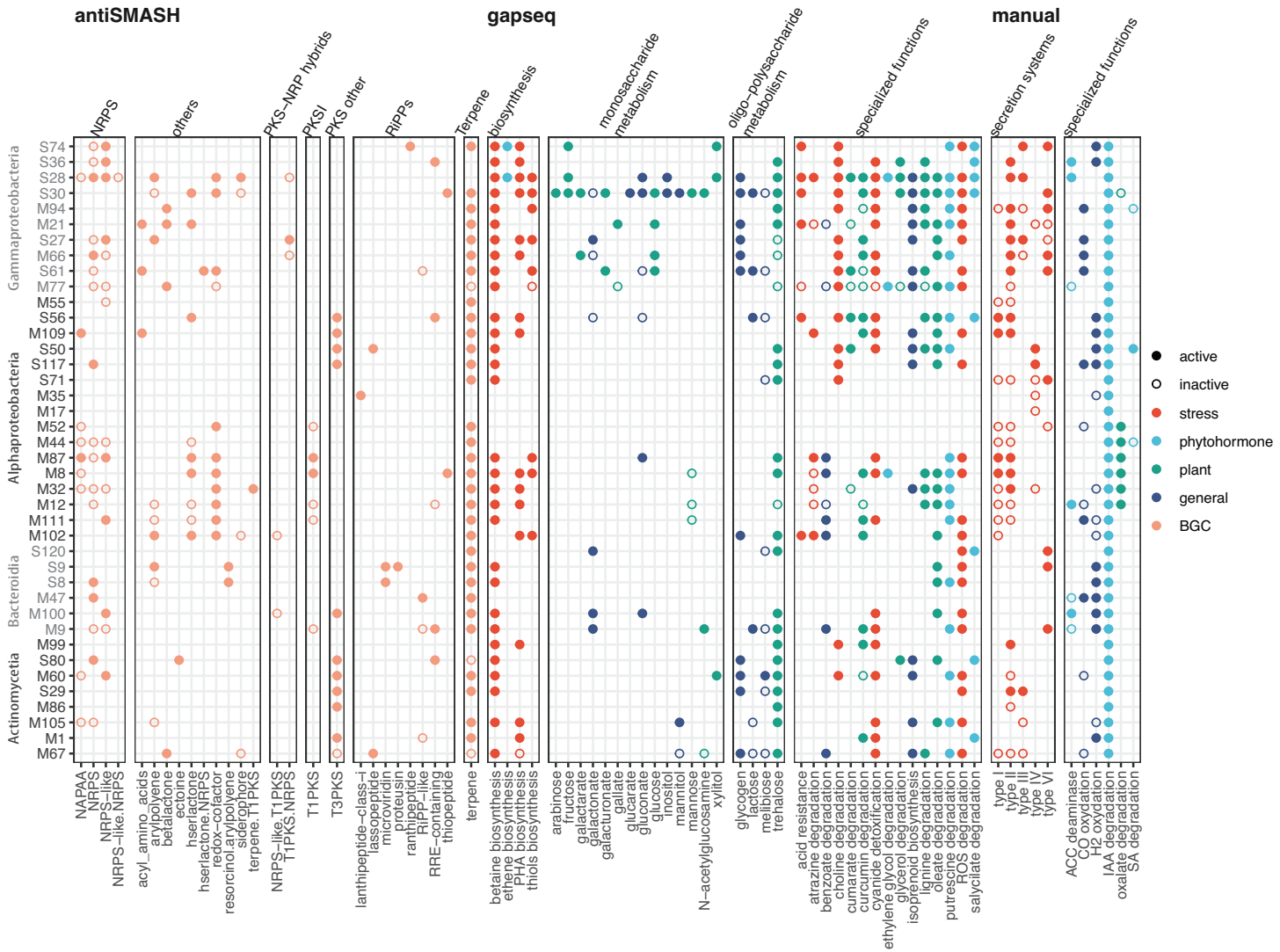

Figure 6. Key functional gene pathways detected in the 40 focal MAGs and their activities (mapped transcripts) during the 2016-2017 switchgrass growing season. Functional gene pathways were curated using antiSMASH for biosynthetic gene clusters, gapseq

for general metabolic pathways, and manual selection for plant-associative functions reported in the literature. Pathways that were discovered in the MAGs but not detected in the transcripts are represented by open circles, and pathways detected in the MAGs and mapped by transcripts are represented by filled circles. Colors categorize different functional groups of pathways.

Figure 7

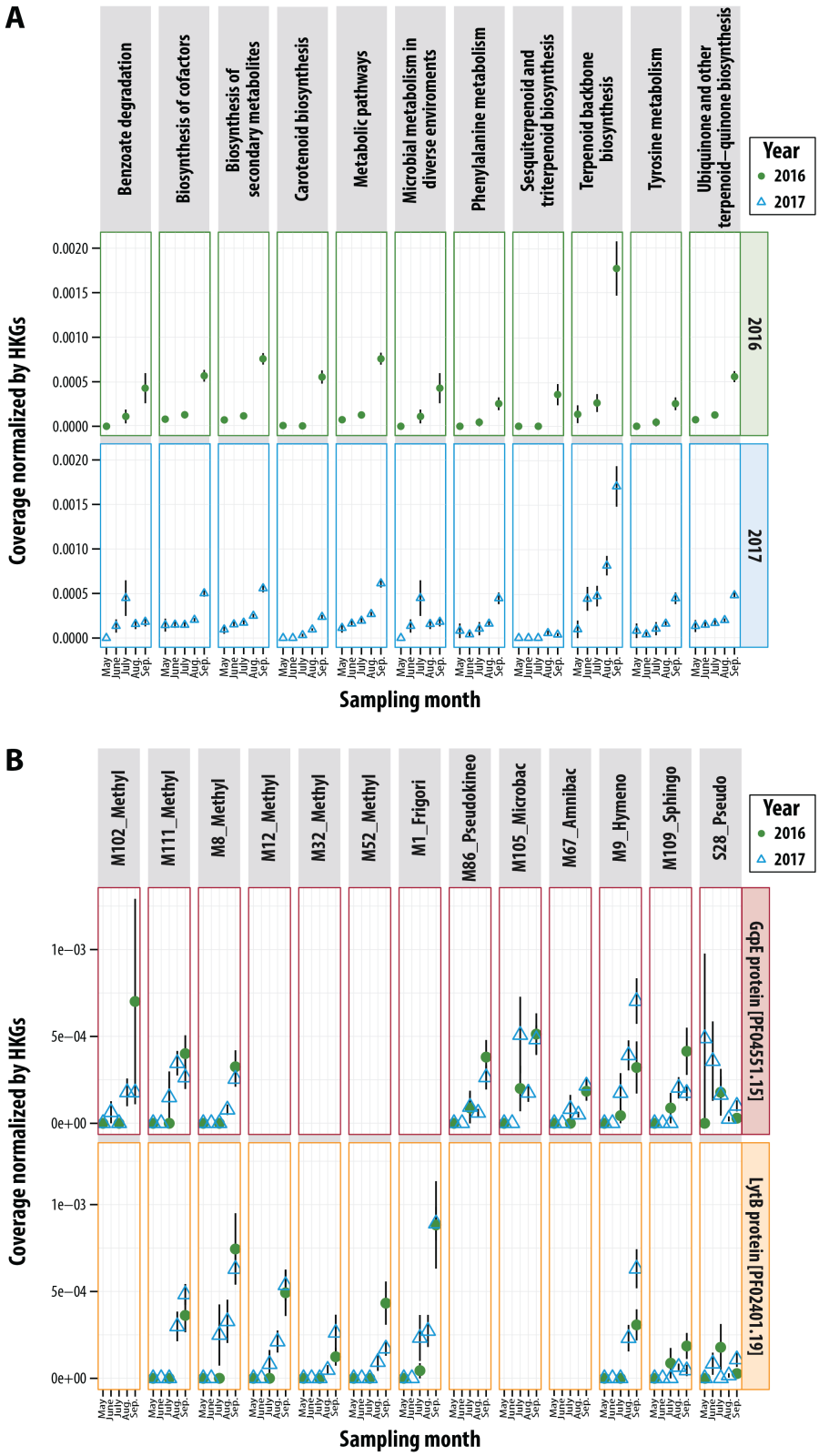
A
Benzoate degradation
Biosynthesis of cofactors
Biosynthesis of secondary metabolites
Carotenoid biosynthesis
Metabolic pathways
Microbial metabolism in diverse enviroments
Phenylalanine metabolism
Sesquiterpenoid and triterpenoid biosynthesis
Terpenoid backbone biosynthesis
Tyrosine metabolism
Ubiquinone and other terpenoid−quinone biosynthesis
Year
2016
2017
Coverage normalized by HKGs
0.0000
0.0005
0.0010
0.0015
0.0020
May
June
July
Aug.
Sep.
Sampling month
B
M102_Methyl
M111_Methyl
M8_Methyl
M12_Methyl
M32_Methyl
M52_Methyl
M1_Frigori
M86_Pseudokineo
M105_Microbac
M67_Amnibac
M9_Hymeno
M109_Sphingo
S28_Pseudo
GcpE protein [PF04551.15]
LytB protein [PF02401.19]
0e+00
5e−04
1e−03

Figure 7. **(A).** 2016 (circle) and 2017 (triangle) switchgrass leaf transcript dynamics of KEGG metabolism classifications associated with terpene metabolism. **(B).** Transcripts in MAGs associated with terminal enzymes in the non-mevalonate isoprene biosynthesis, *gcpE* and *lytB*. MAG IDs include predicted taxonomy at the genus level: Methyl = *Methylobacterium*, Frigor = *Frigoribacterium*, Pseudokineo = *Pseudokineococcus*, Microbac = *Microbacterium*, Amnibac = *Amnibacterium*, Hymeno = *Hymenobacter,* Sphingo = *Sphingomonas,* and Pseudo = *Pseudomonas*.

Figure 8

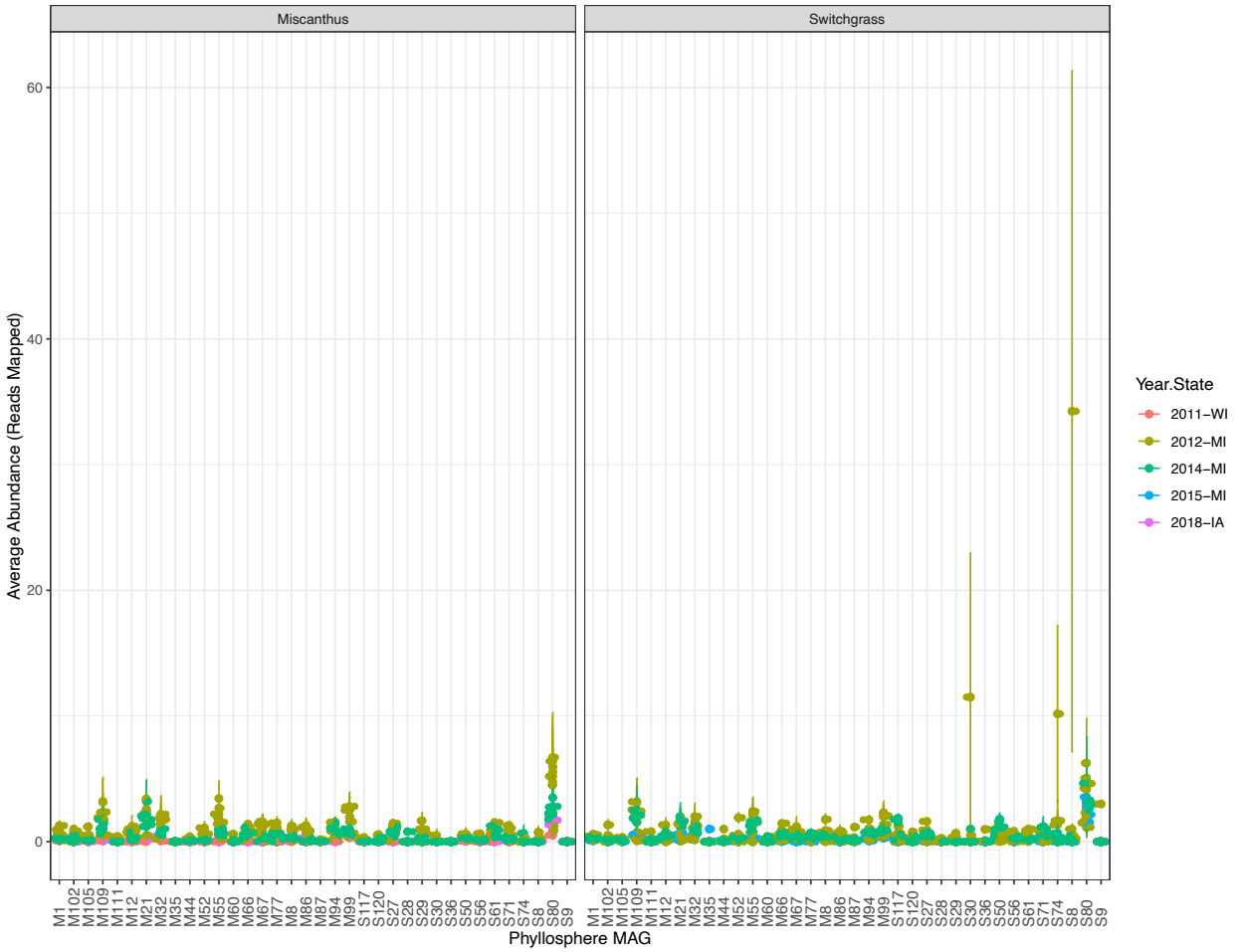

Figure 8. Detection of MAGs in publicly available metagenomes and metatranscriptomes of bioenergy grasses.