MOLECULAR ECOLOGY
RESOURCES **WILEY**

Obtaining deeper insights into microbiome diversity using a simple method to block host and nontargets in amplicon sequencing

Teresa Mayer[1] | Alfredo Mari[1,2] | Juliana Almario[2] | Mariana Murillo-Roos[1] | Hafiz Syed M. Abdullah[1] | Nina Dombrowski[3] | Stephane Hacquard[3] | Eric M. Kemen[2] | Matthew T. Agler[1]

[1]Plant Microbiosis Laboratory, Department of Microbiology, Friedrich Schiller University Jena, Jena, Germany

[2]Department of Microbial Interactions, IMIT/ZMBP, Eberhardt Karl University of Tübingen, Tübingen, Germany

[3]Department of Plant-Microbe Interactions, Max Planck Institute for Plant Breeding Research, Cologne, Germany

Correspondence
Matthew T. Agler, Plant Microbiosis Laboratory, Department of Microbiology, Friedrich Schiller University Jena, Neugasse 23, 07743 Jena, Germany.
Email: matthew.agler@uni-jena.de

Present address
Alfredo Mari, Department Biomedizin, University of Basel, Basel, Switzerland
Juliana Almario, CNRS, INRA,, UMR5557 Ecologie Microbienne, University of Lyon, Université Claude Bernard Lyon 1, Villeurbanne, France
Nina Dombrowski, Department of Marine Microbiology and Biogeochemistry, Royal Netherlands Institute for Sea Research (NIOZ, Den Burg, The Netherlands

Abstract

Profiling diverse microbiomes is revolutionizing our understanding of biological mechanisms and ecologically relevant problems, including metaorganism (host + microbiome) assembly, functions and adaptation. Amplicon sequencing of multiple conserved, phylogenetically informative loci has therefore become an instrumental tool for many researchers. Investigations in many systems are hindered, however, since essential sequencing depth can be lost by amplification of nontarget DNA from hosts or overabundant microorganisms. Here, we introduce "blocking oligos", a low-cost and flexible method using standard oligonucleotides to block amplification of diverse nontargets and software to aid their design. We apply them primarily in leaves, where exceptional challenges with host amplification prevail. *A. thaliana*-specific blocking oligos applied in eight different target loci reduce undesirable host amplification by up to 90%. To expand applicability, we designed universal 16S and 18S rRNA gene plant blocking oligos for targets that are conserved in diverse plant species and demonstrate that they efficiently block five plant species from five orders spanning monocots and dicots (*Bromus erectus*, *Plantago lanceolata*, *Lotus corniculatus*, *Amaranth* sp., *Arabidopsis thaliana*). These can increase alpha diversity discovery without biasing beta diversity patterns and do not compromise microbial load information inherent to plant-derived 16S rRNA gene amplicon sequencing data. Finally, we designed and tested blocking oligos to avoid amplification of 18S rRNA genes of a sporulating oomycete pathogen, demonstrating their effectiveness in applications well beyond plants. Using these tools, we generated a survey of the *A. thaliana* leaf microbiome based on eight loci targeting bacterial, fungal, oomycete and other eukaryotic microorganisms and discuss complementarity of commonly used amplicon sequencing regions for describing leaf microbiota. This approach has potential to make questions in a variety of study systems more tractable by making amplicon sequencing more targeted, leading to deeper, systems-based insights into microbial discovery. For fast

and easy design for blocking oligos for any nontarget DNA in other study systems, we developed a publicly available R package.

KEYWORDS

amplicon sequencing, bacteria, fungi, holobiont, microbial diversity, microbiome, nontarget amplification, oomycete, protist

1 | INTRODUCTION

A revolution in biology is currently underway as our understanding of various systems is brought into the context of the structures and roles of symbiotic microbial consortia. This transformation is the result of increasing research to characterize the microbiota associated with various abiotic or biotic systems. For example, important roles of microbial communities have been revealed in systems as diverse as biotechnological transformations (Werner et al., 2011) and plant and animal health and fitness (Johnson et al., 2020; Mera & Bourne, 2018; Panke-Buisse et al., 2015; Xu et al., 2018). To do so, many studies rely on microbiota profiles generated by amplicon sequencing of phylogenetically informative genomic loci. These profiles are then linked to specific experimental parameters, host phenotypes or performance measurements (Gould et al., 2018).

Microbiomes often include species from all kingdoms of life. These cohabiting members interact with the environment and influence one another via direct associations (Henriques et al., 2020) or indirectly via a host (Prince et al., 2017). To resolve these interactions, to model microbial community dynamics and to pinpoint important microbes, robust systems approaches with a broad and deep coverage of diversity in a high-throughput manner are needed (Lima-Mendez et al., 2015).

Researchers use many technologies and pipelines to generate and sequence amplicon libraries. A major problem affecting broad-diversity amplicon sequencing pipelines is that universal amplification primers amplify DNA from nontarget or overabundant organisms. This applies above all to host-associated microbiome samples, such as plants (Lundberg et al., 2013) and humans (Vestheim & Jarman, 2008) but also to systems with resident sporulating microorganisms (Agler et al., 2016) or endosymbionts (Simhadri et al., 2017). For example, in extracts of healthy *Arabidopsis thaliana* leaves, the host fraction of DNA is typically far more than 95% (Regalado et al., 2020). This leads to a high fraction of host plastid DNA being amplified and sequenced which reduces the effective sequencing depth that can be obtained from one sequencing run tremendously and obscures microbial diversity.

Methods commonly used to address this problem include peptide nucleic acid (PNA) "clamps" (Lundberg et al., 2013) or oligonucleotides modified with a C3 spacer (Vestheim & Jarman, 2008), which both arrest amplification of nontarget amplicons. These, however, can be costly to design and implement, especially when the needs of researchers are constantly changing. Additionally, the ratio of microbial to host plastid reads in amplicon sequencing data is commonly used as an intrinsic measure of the microbial load of the samples (Humphrey & Whiteman, 2020). Methods such as C3 spacers and PNA clamps are not necessarily designed to retain this quality and may therefore lead to a loss of quantitative insights into microbial loads.

To address this major barrier, here, we introduce a cost-efficient and flexible technique termed "blocking oligos". In contrast to other methods used to block off-target amplification, blocking oligos are standard DNA oligos that can be designed and ordered quickly and easily (synthesis of PNAs can take several weeks of time and is significantly more expensive). Additionally, whereas C3 spacers are modified to prevent amplification, blocking oligos do not have any unusual properties. Blocking oligos have their origins when we previously applied a set of standard oligonucleotides to block amplification of *Arabidopsis thaliana* DNA during 16S rRNA gene or ITS amplification from leaf samples (Agler et al., 2016). Recognizing the potential broader usefulness of this approach, we have now extensively tested variations on it for effectiveness and potential biases. Our testing shows that blocking oligos increases the number of useful reads from a sequencing run and increases alpha diversity discovery without biasing beta diversity findings and are compatible with deriving quantitative bacterial load insights from host/microbe ratios in 16S rRNA gene data. Here, we greatly expand the blocking oligo toolbox with *A. thaliana* blocking oligos covering eight commonly used loci in the 16S rRNA gene, 18S rRNA gene, ITS1 and ITS2 regions and universal 16S and 18S rRNA gene plant blocking oligos that are effective in blocking most host plant species. We also develop blocking oligos for a nontarget microorganism, demonstrating that the approach can be extended well beyond plants. Since use in other systems could be hindered by the hurdle of designing new, effective oligos (Regalado et al., 2020; Lundberg et al., 2020), we developed an R package with three simple functions to rapidly and easily design effective oligos to block amplification of any specific nontarget DNA template. We show that dropping blocking oligos into two-step amplicon sequencing library preparation pipelines enables nearly complete characterization of hyperdiverse microbiomes in difficult systems. We expect that by enabling more researchers to cost-effectively increase diversity discovery, blocking oligos will help broaden the applicability and impact of amplicon sequencing experiments.

2 | MATERIALS AND METHODS

2.1 | Design of blocking oligos to avoid nontarget template amplification

Blocking oligos were previously designed for the *A. thaliana* chloroplast (16S rRNA gene V3–V4 region) or mitochondria (16S rRNA gene V5–V7 region) and *A. thaliana* ITS1 and ITS2 regions (fungal

and oomycete ITS) (Agler et al., 2016). Primers specific to a known, nontarget DNA template (blocking oligos) and nested inside the universal primer binding sites (Figure 1) were designed (see Supporting Information Files S1a and S1b for all oligo and primer sequences used in this study). To design oligomers with high specificity, we adapted the approach used by Lundberg et al., (2013) for PNA clamps. In short, the region of interest (chloroplast/mitochondria 16S rRNA gene or ITS) from *A. thaliana* was divided up into "k-mer" sequences of length 30. We then used BLAST to search the k-mers against a blast-formatted target database. The BLAST search used the following parameters which allow weak matches: percent identity 25, word size 7, e-value 100,000. Candidate 30-mer blocking oligos were selected that received a relatively low number of hits. For this study, we have developed an R package, "AmpStop", which automates this part of the process and suggests good candidates, and which is freely available on GitHub (https://github.com/magler1/AmpStop). We then selected candidates which had a high T_m (well above the universal primer binding temperatures) and which had low potential to form self-dimers or hairpins. Candidate oligomers were tested in single-step amplification of target and nontarget templates for nontarget specificity. The selected blocking oligomers (Supporting Information Files S1a and S1b) were always used in the first amplification step of library preparation (blocking cycles), resulting in shortened amplicons that could not be elongated with Illumina adapters in the second amplification step (Figure 1). All databases will be made publicly available prior to publication on Figshare.

2.2 | Design of 18S rRNA gene blocking oligos for host and microbial nontargets

To reduce amplification of both *Arabidopsis thaliana* and the oomycete pathogen *Albugo laibachii* amplification in the 18S rRNA gene regions we designed additional blocking oligos for both organisms (see also Supporting Information File S1a). We tested them by preparing 18S rRNA gene amplicon libraries from two mock communities consisting of *A. thaliana* (97% or 87%), *A. laibachii* (0% or 10%), *Sphingomonas* sp. (1.5%), *Bacillus* sp. (1.5%) and 0.001% to 1% of

target *Saccharomyces boulardii* (composition in Table S1). We then used qPCR with primers targeting the *Saccharomyces* sp. 18S rRNA gene (V4 Fwd/Rev: AACCTTGAGTCCTTGTG/AATACGCCTGCTTTG V9 Fwd/Rev: GTGATGCCCTTAGACG/ACAAGATTACCAAGACCTC) to relatively quantify target *S. boulardii* in the libraries generated with and without blocking oligos.

2.3 | Design of universal plant blocking oligos for 16S and 18S rRNA gene loci

To design blocking oligos that could be used for multiple plant species we used the same approach as described above. In short, we used chloroplast sequences (to design 16S rRNA gene blocking oligos) or plant sequences (to design 18S rRNA gene blocking oligos) from multiple plant species that spanned the phylogeny of plants (Table S2) as input for the AmpStop package. The 21 plant species were chosen because they are highly diverse, and many were available to us for testing the oligos. We checked where the results from the AmpStop package overlapped between species. Resulting blocking oligo candidates were tested in a one-step PCR protocol for their specificity against genomic DNA from various plant species and bacterial mixes (for 16S rRNA gene, see Figure S1 and Tables S3 and S4) or *Trichoderma* sp. (for 18S rRNA gene). The blocking oligo pairs BloO_16S_F5 and BloO_16S_R1 (16S rRNA gene) and BLc_18S_F5 and BLc_18_R5 (18S rRNA gene) were chosen for further analysis, since they hit most of the plant species tested but at the same time amplified none of the bacterial mixes or *Trichoderma* sp., respectively. These selected oligomers were used in the blocking cycles for library preparation from multiple plant species.

2.4 | Testing *A.thaliana* blocking oligos against mock communities

We tested blocking oligos designed to block *A. thaliana* in two loci from each of bacteria (16S rRNA gene V3–V4 and V5–V7), fungi (ITS1 and 2) and oomycetes (ITS1 and 2, for primer design see
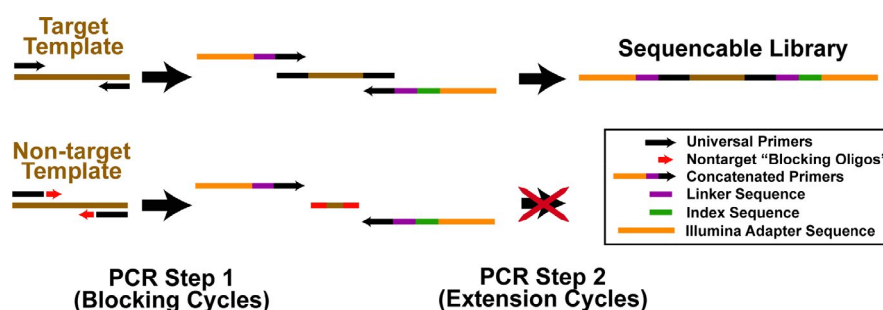
FIGURE 1 A strategy to reduce nontarget amplification in amplicon sequencing pipelines. A two-step amplification approach is used in which universal primers first amplify genomic templates, then indices and adapters are added in a second step. To prevent amplification of nontarget templates, blocking oligos complementary to nontarget genomic templates are employed in the first "blocking cycles" PCR step, resulting in short amplicons that cannot be amplified with concatenated primers in the second "extension cycles" step. Without addition of Illumina adapter sequences, these PCR products are not sequenced

Supporting Information Methods and Table S5) using mixed king-dom mock microbiomes (Table S6). The simulated host-associated microbiomes consisted of 5% of a mix of the mock microbiomes and 95% *A. thaliana* genomic DNA. For each template sample, six separate PCR reactions were prepared, one targeting each locus. Instead of preparing replicate libraries of each template mix, we prepared multiple template mixes including both even and uneven mock communities and prepared libraries for each in one replicate. We also tested the effect of variations on the amplicon sequencing library preparation method. We compared PCR performed in one step (35 cycles, no blocking) or two steps (10 then 25 cycles or 25 then 10 cycles). For two-step preparations, the primers used in the first step consisted of unmodified universal amplification primers (Figure S2). For single-step preparations and for the second step in two-step preparations, primers were a concatenation of the Illumina adapter P5 (forward) or P7 (reverse), an index sequence (reverse only), a linker region, and the universal primer for the region being amplified (Figure 1 and Figure S2). Sequences and details of all primers used can be found in Supporting Information File S1a and details on PCR, library preparation and sequencing, as well as the steps to generate operational taxonomic unit (OTU) tables and taxonomy from raw multilocus data can be found in the Supporting Information Methods (Protocol A). We summarized bacterial, fungal and oomycete OTU tables by taxonomic ranks, converted abundances to relative values and plotted the genus- and order-level taxonomic distribution directly from this data with the package ggplots2 in R. So that the plant fraction would stand out, the colour of chloroplast and mitochondria fractions in Figure 2a were altered to grey tones in Illustrator. To analyse the percent reduction in host plant-associated reads when blocking oligos were employed, we considered the relative abundance of reads associated with the class "Chloroplast" or the order "*Rickettsiales*" in the 16S rRNA gene OTU tables and reads in the kingdom "*Viridiplantae*" in the ITS OTU tables in samples with *A. thaliana* DNA and with and without blocking oligos. Separately, we also checked and confirmed using negative controls that the nonindexed amplification step of the two-step library preparation approach did not result in sample cross-contamination (see details in Supporting Information Notes and Figure S3).

## 2.5 | Testing universal plant blocking oligos in natural leaves and mock communities

The 16S and 18S rRNA gene universal plant blocking oligos were tested using five leaves from five plant species collected from different experiments. The plant species represent five plant orders spanning monocots and dicots (Table S3). All plant leaves were naturally grown outside without artificial addition of any microorganism. Details on the DNA extraction can be found in the Supporting Information Methods. The universal 16S rRNA gene blocking oligos were additionally tested for how they affect bacterial

diversity distribution by preparing one library with or without blocking oligos with a mixed microbial community standard as template (ZymoBIOMICS microbial community DNA standard) and against three different soil DNA extracts. For testing if bacterial load information is maintained, we quantified concentrations of DNA extracts of three plant species that were grown axenically and combined it with the Zymo standard to create genomic DNA mixes (0%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 15% and 25% microbial genomic DNA). The eight loads across three plants species were each prepared in one amplicon sequencing library (24 reactions total).

Libraries were prepared with either 10 or 15 blocking cycles and 25 or 20 extension cycles, respectively. In short, the templates were amplified in the blocking cycles including the universal 16S rRNA gene primers as well as the blocking oligos. The product of this first PCR was then purified and amplified in the extension cycles using concatenated primers. The extension step used concatenated primers similar to before but both primers had unique index sequences. Sequences and details of all primers used can be found in Supporting Information file S1b and details on PCR, library preparation and sequencing, as well as the steps to generate ASV tables and taxonomy from raw data can be found in the Supporting Information Methods (Protocol B).

2.6 | Eight-locus amplicon sequencing with blocking oligos to fully characterize *A.thaliana* leaf microbiome diversity

We next expanded the multilocus approach to cover eukaryotic microbial diversity more completely by including two additional 18S rRNA gene loci (V4–V5 and V8–V9, Figure 1a and Table S7 - primer sequences are available in Supporting Information File S1a). With the expanded target set, we characterized the phyllosphere microbiome of *A. thaliana* leaves infected with the oomycete pathogen *Albugo laibachii*. Whole leaves (defined as a single whole rosette) or endophytic fractions of leaves (defined as in Agler et al., 2016) were collected in the wild (a total of 18 samples: nine whole leaf, nine endophyte) and were immediately frozen on dry ice. DNA extraction was performed as described previously (Agler et al., 2016). Library preparation, sequencing and analysis was performed as in the optimized protocol with blocking oligos. To provide a complete and concise picture of the diversity of microbiota inhabiting *A. thaliana*, we combined the data from all samples. To visualize data, we assigned taxonomy to OTUs and generated two phylogenetic trees where branches represent unique genera. Trees were generated based on the taxonomic lineages (not phylogenetic relatedness of OTUs or genera) with the ape package in R and output as newick files (Paradis et al., 2004). The trees were uploaded to iTOl v3.1 (Letunic & Bork, 2016) to colour branches by taxonomy or by targeted regions. The first tree, for eukaryotes, includes data from the 18S rRNA gene and ITS targeted regions. The second tree includes data from the 16S rRNA gene targeted regions.
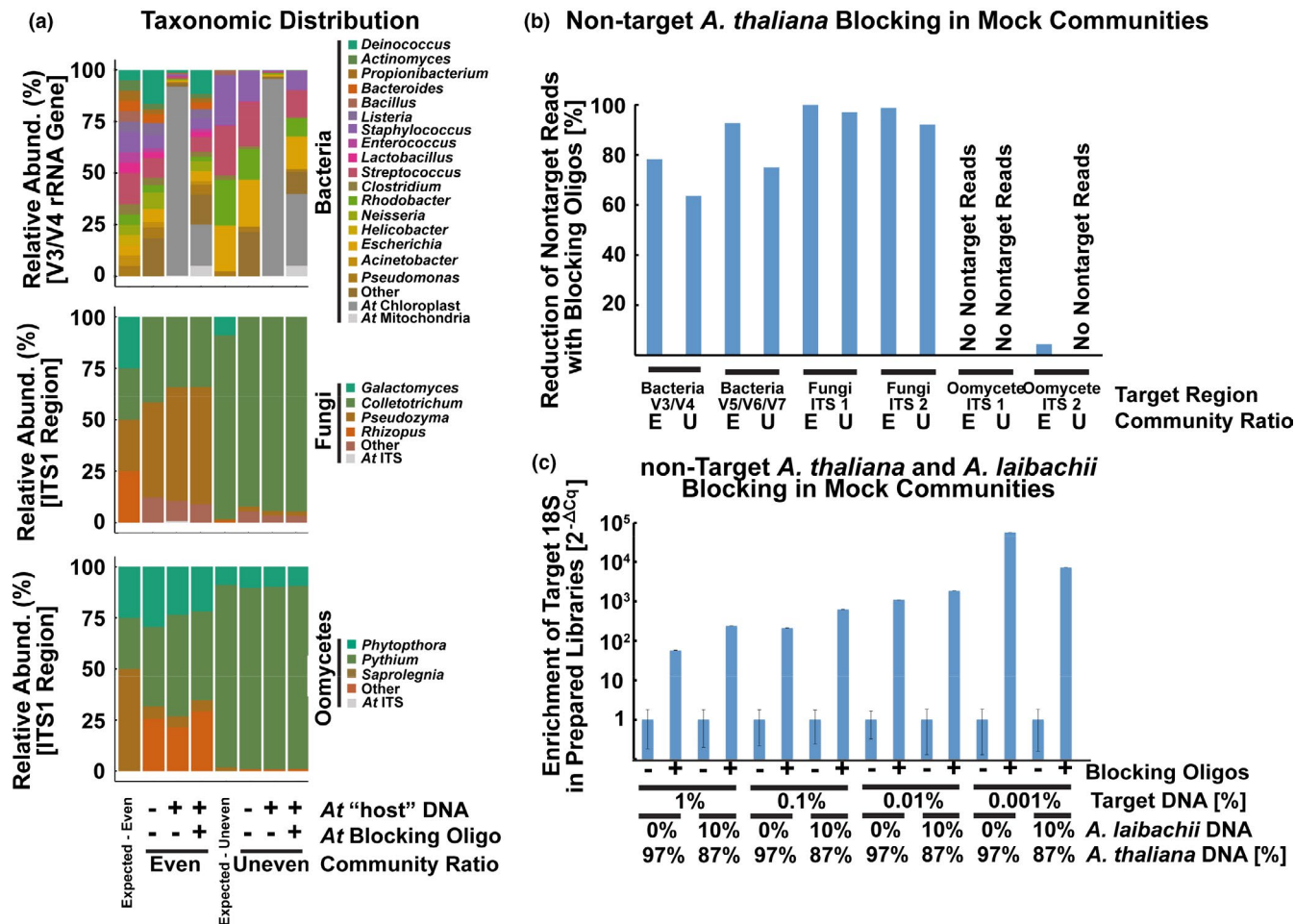
FIGURE 2 Reproducible and accurate characterization of mock communities of bacteria, fungi, and oomycetes by amplicon sequencing. (a) Observed taxa at the genus level in sequenced mock communities closely matched expected communities. Each bar represents the results of one prepared amplicon sequencing library. Additional conditions and loci are found in Figure S4 and S5. (b) Near-complete reduction of amplification of *A. thaliana* "host" nontarget plastid 16S rRNA gene or ITS by employing blocking oligos in preparation of mock community libraries. Each bar represents the results of one prepared amplicon sequencing library. Replicate bars in each locus are labelled "E" and "U", indicating mock communities with even or uneven distribution of taxa, respectively. (c) Relative increase of target (*Saccharomyces* sp.) 18S rRNA gene V4–V5 region amplicons (qPCR $2^{-\Delta Cq}$ values relative to measurement without blocking oligomers) in mock community libraries prepared with blocking oligomers to reduce *A. thaliana* and *A. laibachii* nontarget amplification. Error bars represent the standard error based on three replicate qPCR reactions

3 | RESULTS

3.1 | Blocking oligos reduce nontarget amplification by universal primers

Host or other nontarget amplicons are not useful to assess microbial diversity and are therefore often discarded, wasting sequencing depth. Therefore, we developed blocking oligos to reduce amplification of nontarget DNA. Blocking oligos are standard oligonucleotides whose binding site is nested inside the binding site of universal primers for a locus of interest and are highly specific for a nontarget organism (Figure 1 and Figure S2). During the first PCR step (blocking cycles), their nested binding location physically blocks the nontarget elongation by the polymerase at the universal primer site, resulting in only short nontarget amplicons.

In the second PCR step (extension cycles), concatenated primers are used to add indices and Illumina sequencing adapters. Since the concatenated primer binding site is not present on nontarget products, they are not amplified, and the sequencing library becomes enriched with target amplicons (Figure 1). They can be dropped into the first step of any standard two-step amplicon library preparation pipeline.

We previously designed blocking oligos to reduce amplification of plant chloroplast (16S V3–V4 rRNA gene), mitochondria (16S V5-V7 rRNA gene) and plant ITS (fungal and oomycete ITS regions 1 and 2) (Table S7) (Agler et al., 2016). We only targeted chloroplast in the V3–V4 and mitochondria in the V5–V7 regions because these are the most problematic host sequences in each region. Here, we thoroughly tested them by checking how much they reduced host amplification compared to a "standard" library preparation without

blocking, and whether they biased beta diversity estimates. We used a mock community that simulated a host-associated microbiome (95% *A. thaliana*/5% microbial genomic DNA). With blocking oligos, useful read depth was largely recovered by eliminating 60%–90% of nontarget contamination in bacterial 16S rRNA gene data sets and nearly all of the small amount of contamination in fungal ITS data (Figure 2a,b). Blocking oligos were slightly more effective in even (target microbes all in equal abundance) than in uneven communities (target microbes in unequal abundance), but much of this difference was either caused by poor taxonomic annotation or universal primer bias (see Supporting Information Notes). Importantly, in all three kingdoms, replicates that only differed in the use of blocking oligos had very similar differences from the expected distribution of the mock microbial communities (Figure 2a, Figures S4A and Figures S5A and at the order level in Figure S6), demonstrating that blocking oligos do not change the recovered taxa distribution (p > .2 in the *t* test of Bray-curtis distance to the expected mock community structure with versus without blocking oligos (Figures S4B and Figures S5B). Variations of the library preparation protocol had little or no effect on the results (see Supporting Information Note for additional details on testing the library preparation protocol).

Sequencing 18S rRNA gene libraries in addition to ITS (Table S7) should recover more leaf eukaryotic microbial diversity (George et al., 2019). This diversity would be obscured, however, by the host and occasionally by sporulating pathogens such as *Albugo laibachii* that are efficiently amplified by universal 18S rRNA gene primers. Therefore, we designed and tested blocking oligos to overcome nontarget amplification of both *A. laibachii* and *A. thaliana* in the 18S rRNA gene region (Table 1). For testing, we generated mock genomic DNA templates containing nontarget DNA (*Bacillus* sp., *Sphingomonas* sp., *A. thaliana, A. laibachii)* and target (*Saccharomyces boulardii*) genomic DNA (Table S1). We then prepared 18S rRNA gene V4–V5 region amplicon libraries from the template samples with or without *A. thaliana* and *A. laibachii* blocking oligos in the first PCR step. Finally, we quantified the levels of target (*S. boulardii*) amplicons

in the prepared libraries using qPCR. Indeed, blocking both nontargets increased target levels between ~57× (1% target template) and ~57,000× (0.001% target template) (Figure 2c), demonstrating that both host and nontarget microbial taxa can be efficiently simultaneously blocked.

3.2 | Universal plant blocking oligos enable profiling of microbiota in many host species

A. thaliana blocking oligos are not effective against every plant host, so users would need to design and test new oligos for their purposes (Giangacomo et al., 2020a). Thus, we expanded to multiple hosts by designing a set of oligos to block amplification of chloroplast 16S rRNA gene and plant 18S rRNA gene using a highly diverse set of plant species (Table S2). To identify regions where blocking oligos would be most effective across plant species, we used our R package AmpStop (see methods and the AmpStop section below). Candidates were first tested for specificity to plants by amplifying DNA from 21 plant species (Table S3) and mixed bacterial DNA and then visualizing bands on a gel. We selected primer sets that amplified most plants but avoided amplification of bacteria or fungi (For 16S rRNA gene see Figure S1). Next, we tested the oligos by preparing sequencing libraires (Protocol B, see Supporting Information notes for comparison with Protocol A) with DNA templates from leaves of plant species representing five orders spanning monocots and dicots (*Amaranthus spec., Arabidopsis thaliana, Bromus erectus, Lotus corniculatus* and *Plantago lanceolata*) (Table S3). Although *A. thaliana* blocking oligos were very efficient in mock leaf microbiomes (Figure 2), in real leaf samples they only sometimes helped recover higher microbial diversity (Figures S7D/E and S8D/E). Microbial loads on leaves are typically very low (Regalado et al., 2020), so we reasoned that more blocking cycles may be required in real leaves. Therefore, for testing universal 16S rRNA gene blocking oligos, we compared 10 versus 15 blocking cycles.

TABLE 1 Overview of the loci and the nontarget regions for which blocking oligos were previously designed and for this study. The blocking oligo approach of Agler et al. (2016) was extended here to eight loci to characterize bacteria, fungi, oomycetes and other eukaryotic microbiota while avoiding nontarget *A. thaliana* and *Albugo* sp. amplification. Libraries were prepared and sequenced with "Protocol A", similar to Agler et al., (2016). Single sets of blocking oligos that block amplification of DNA of most plant hosts were designed for the 16S and 18S rRNA gene regions. An alternative protocol B was used for sequencing and library preparation. All primer sequences are available in Files S1a and S1b.

Target region	Nontarget blocking oligo set	Library protocol	Blocking oligo origin
Bacteria - 16S rRNA gene V3/V4	*A. thaliana* Chloroplast 16S rRNA gene V3/V4	A	Agler et al., 2016
Bacteria – 16S rRNA gene V5-V7	*A. thaliana* Mitochondria 16S rRNA gene V5-V7		
Oomycete ITS1/ITS2	*A. thaliana* ITS1/ITS2		
Fungi ITS1/ITS2			
Eukaryote – 18S rRNA gene V4/V5	*A. thaliana* 18S rRNA gene V4/V5		This study
	Albugo sp. 18S rRNA gene V4/V5		
Eukaryote – 18S rRNA gene V9	*A. thaliana* 18S rRNA gene V9		
	Albugo sp. 18S rRNA gene V9		
Bacteria - 16S rRNA gene V3/V4	Universal Plant Chloroplast 16S rRNA gene V3/V4	B	
Eukaryote – 18S rRNA gene V4/V5	Universal Plant 18S rRNA gene V4/V5		

When blocking for 10 cycles, 1%–25% of target (nonchloroplast) reads were recovered from *Arabidopsis thaliana*, *Bromus erectus* and *Lotus corniculatus* (Figure 3a, the other two species had no usable reads with 10 cycles). We found that 15 blocking cycles increased the amount of retrieved target reads by at least 2.5-fold compared to 10 cycles (Figure 3a), increasing the fraction of useful reads 8–16x compared to without blocking oligos in all five plant species (Figure 3a and Figure S9). 18S rRNA gene blocking oligos were only tested with 10 blocking cycles but in four plant species we observed an increase from <5% target (nonplant) reads without blocking oligos to up to 57% target reads with blocking oligos (Figure S10). Next, we again checked whether blocking oligos bias recovered beta diversity (differences between samples). In the 16S rRNA gene data, we observed no significant effects on leaf samples (Figure 3b), a microbial community standard (Figure S11), nor in three different samples with soil DNA as template (Figure S12). We only tested the 18S rRNA gene oligos on leaf samples and observed that they resulted in recovery of more diverse communities. However, without further testing using mock communities we cannot say to which extent, if any, the 18S rRNA gene oligos introduce bias to the measurements. Overall, all universal blocking oligos can be used with practically any sample to increase useful data recovery and 16S rRNA gene blocking oligos do this without biasing beta diversity patterns.

3.3 | Plant blocking oligos increase recovered alpha diversity

When the majority of reads retrieved from amplicon sequencing are nontarget, the effective sequencing depth is drastically decreased. Thus, an important question is whether this actually obscures the microbial diversity recovered and whether blocking oligos allow recovery of higher alpha diversity. We calculated Shannon and Simpson diversity indices (observed species richness and diversity, respectively) and ACE and Chao1 (which estimate total species richness) alpha diversity of bacterial 16S rRNA gene data from the three naturally grown plant species amplified with 10 or 15 blocking cycles with and without universal blocking oligos. To be able to carry out a fair comparison the data sets for both protocols were combined and rarefied to the same number of reads. With 10 blocking cycles, observed richness and diversity were marginally higher (Figure 4a,b) and estimates of total species richness were unchanged (Figure 4c,d) with blocking. Blocking for 15 cycles, on the other hand, resulted in significantly higher observed and estimated richness and diversity ($p < .01$ for Shannon and Simpson and $p = .11$ and $p = .09$ for Chao1 and ACE, respectively) (Figure 4). The difference between 10 and 15 cycles is again most likely due to low bacterial loads in real leaf samples (Figures S7D,E and Figures S8D,E). Thus, >10 blocking cycles are recommended to consistently recover complete diversity.

3.4 | Leaf bacterial loads can be estimated using 16S rRNA gene amplicon data

One limitation of amplicon sequencing is that it is compositional, such that the quantitative bacterial load information is lost. Recently it has been demonstrated that nontarget host to target bacterial ratios can be used to roughly estimate bacterial loads in plant samples (Humphrey & Whiteman, 2020). Losing this information would be a downside of implementing blocking oligos. We observed that after blocking chloroplast amplification with the universal blocking oligos, host mitochondrial reads still made up a significant part of the data (Figure 3). Therefore, we checked whether information required for quantitative insights into microbial load was still contained in the data generated with chloroplast blocking oligos. We tested this using the same plant species as previously, which we grew axenically,
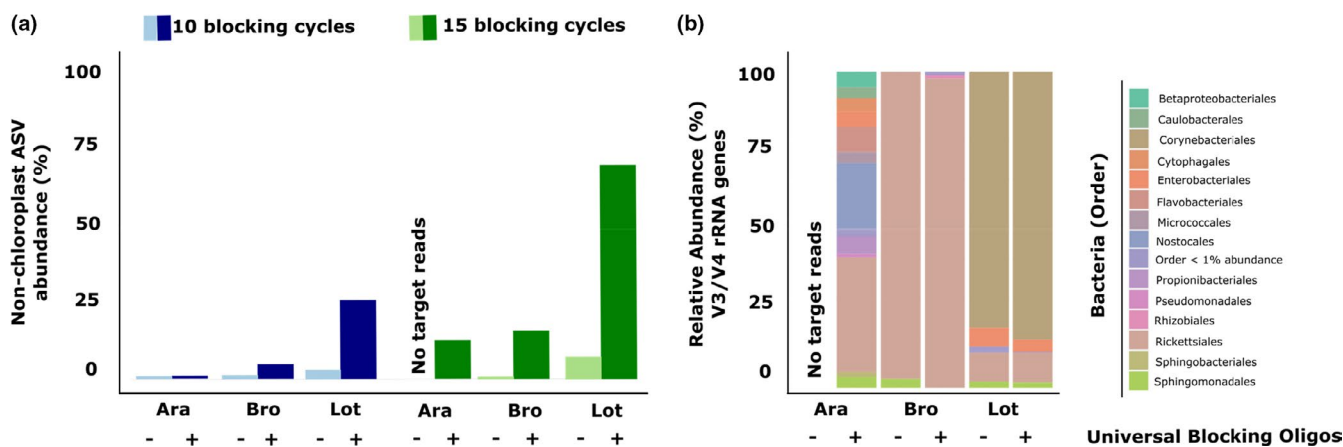
FIGURE 3 Universal blocking oligos successfully block undesired reads therefore increasing sequencing depth. (a) Percentage of reads assigned to ASVs other than chloroplast (nonchloroplast ASVs) with 10 versus 15 blocking cycles. The use of blocking oligos leads to higher recovery of bacterial ASVs. When the number of blocking cycles is increased, the fraction of blocked ASVs increases as well. (b) Taxonomic distribution in samples of different plant species with and without blocking oligos (15 blocking cycles). The use of universal blocking oligos does not significantly change the identity of retrieved ASVs. Each bar represents the results of one prepared amplicon sequencing library. Results for other plant species and mock communities are shown in Figure S9 and Figure S11

FIGURE 4 The use of blocking oligos increases the bacterial alpha diversity recovered. Comparison of alpha diversity measures between samples with 10 or 15 blocking cycles. We calculated the alpha diversity indices Shannon (a), Simpson (b), Chao1 (c) and ACE (d) for the plant samples shown in Figure 3 (n = 3, 1 of each species). Shannon and Simpson diversity indices combine richness and diversity (they measure both the number of species as well as the inequality between species abundances), whereas Chao1 and ACE estimate the total species richness. The use of blocking oligos for 10 cycles showed an increase in only Shannon and Simpson alpha diversity indices. However, when blocking for 15 cycles all indices show an increase in alpha diversity. Each box represents three measurements from the three plant species shown in Figure 3. p-values are calculated with a one-sided paired t test
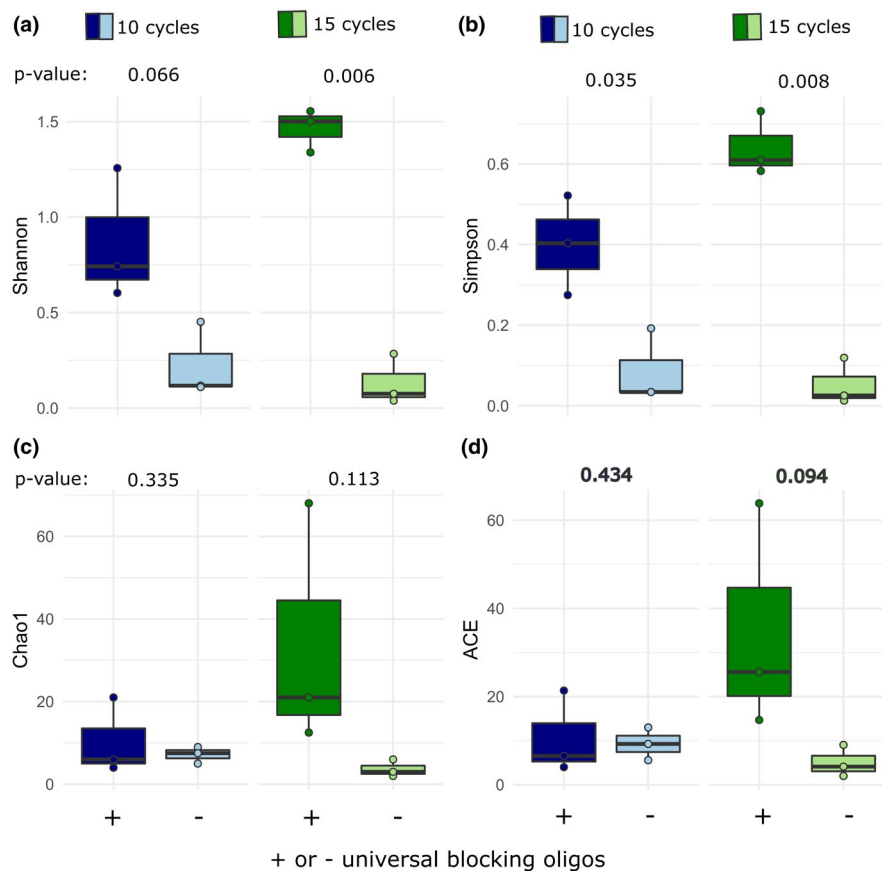
harvested DNA, then combined with specific amounts of a bacterial DNA mix (Zymo Research Europe). After removing remaining chloroplast reads but leaving mitochondrial reads, we calculated the fraction of bacterial 16S rRNA gene sequences (Figure 5).

Up to a 1% fraction of bacterial DNA in the template, we observed a nearly linear increase in the fraction of sequenced reads assigned to bacteria (Figure 5). For a given load, the fraction of bacterial reads was similar for *A. thaliana* and *L. corniculatus* and was higher for *B. erectus*. Therefore, we conclude that within samples of the same species, blocking oligos not only increase recovered bacterial diversity but can be applied so that bacterial load information is maintained.

## 3.5 | An expanded multikingdom view of leaf microbial diversity

We tested using the blocking oligo system to generate as broad of a microbial diversity profile as possible from leaves. We amplified and sequenced the eight target loci in 12 wild *A. thaliana* leaf samples, including leaves with sporulating *A. laibachii* infections, and employed the *A. thaliana* and *A. laibachii* specific blocking oligos. We then analysed the diversity insights gained with this broad approach. The addition of the 18S rRNA gene primers broadly targeting eukaryotes increased diversity recovery by nearly 50% compared to ITS primers alone (observed genera, Figure 6). This included red and green algae, cercozoa and amoebozoa and even suggested signs of metazoa

like insects and helminths (Figure 6 and Supporting Information File S2). The fungal and oomycete ITS data sets complemented the broader 18S rRNA gene data with more specificity in those groups – together, these two accounted for 44% of observed eukaryotic genera (Figure 6a). Prokaryote data sets further demonstrate complementarity for primer sets targeting the same groups of microbes (Figure 6b). Here, 42% of observed genera were discovered by both primer sets, with complementary diversity discovery especially in the phyla *Cyanobacteria* (V3–V4 data set) and *Firmicutes* (V5–V7 data set). For all groups, both tested primer sets tended to recover similar diversity at high taxonomic levels with differences at the genus level. A full list of which primer sets recovered which taxa are provided in Supporting Information File S2. Thus, blocking of overabundant host and microbial amplicons allows deep diversity characterization in leaves using multiple loci.

3.6 | AmpStop: An R package for quick design of blocking oligos for any nontarget organism

A key advantage of using standard oligomers as a tool to block amplification is that many design options can be tested rapidly and at low cost using standard PCR techniques. A limit on rapid implementation in laboratories could be the design step, where some computational know-how is required. To reduce this burden, we created AmpStop, an R package to automate design of blocking oligos. AmpStop can be used by anyone with R and BLAST+installed on their computer.
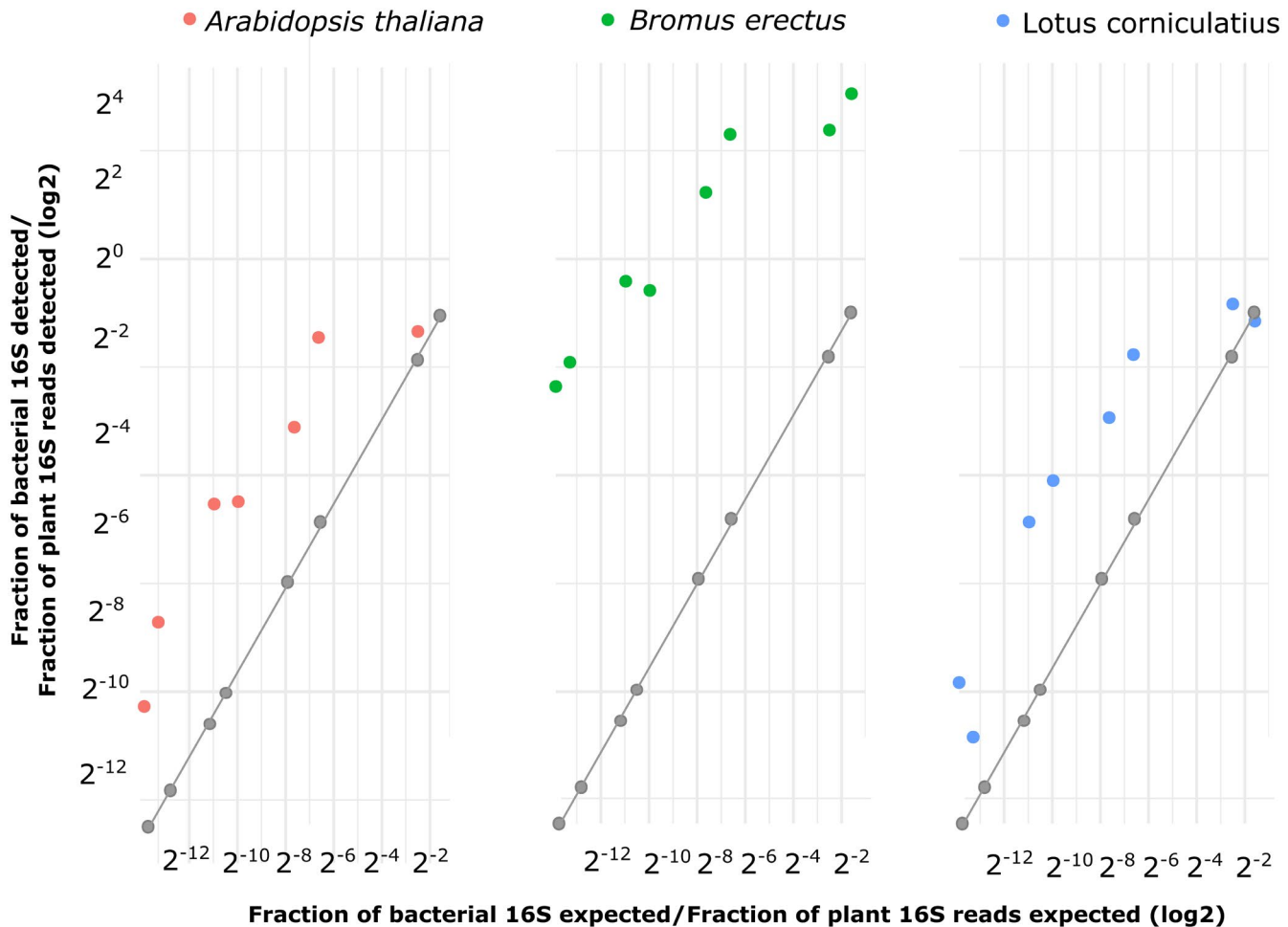
FIGURE 5 The fraction of bacterial reads can be used to gain quantitative microbial load information. The fraction of bacterial reads in sequenced amplicon libraries increases as the load of mixed bacterial gDNA increases in the template. The axenic plant gDNA used in the mixes were *A. thaliana* (red), *B. erectus* (green) and *L. corniculatus* (blue). The ratio of bacterial to plant reads in the amplicon sequencing libraries (y-axis) was plotted against the expected ratio (x-axis). The grey points represent the expected linear relationship. Each point on the plot represents one replicate amplicon sequencing library

It requires as input only the amplified nontarget region (for example the host ITS1 sequence) and a target sequence database that is BLAST-formatted. Three functions enable users within minutes to generate a list of all possible blocking oligos, a figure showing how many times each oligo "hits" target templates and other useful metrics of specificity (Figure 7a–7c) and a list of the most promising blocking oligo pairs. Since the design of peptide nucleic acid clamps follows practically identical steps (Lundberg et al., 2013), the package can also be used to design them. AmpStop and detailed instructions on its use and interpretation of results is freely available on GitHub (https://github.com/magler1/AmpStop).

4 | DISCUSSION

Amplicon sequencing of phylogenetically or functionally informative loci has become an indispensable technique in a variety of biology-related fields because its targeted approach (compared to untargeted approaches like metagenomics) enables in-depth diversity characterization with accurate annotation using specialized databases (Quast et al., 2013). It has revealed that microbial community structuring is more complex than previously thought and suggested extensive interactions between (a) biotic factors and microbes (de Menezes et al., 2015) and between microbes even across kingdoms (Agler et al., 2016; Lima-Mendez et al., 2015). As we have shown here, recent advances have made sequencing up to eight loci in parallel possible, drastically increasing throughput and diversity resolution. This will be important to add certainty to system scale investigations of factors contributing to microbial community structures. On the other hand, the use of universal primers has the disadvantage that highly abundant microorganismal or host DNA are often strongly amplified, sacrificing read depth and masking diversity (Hanshew et al., 2013).

A previously described method to address this problem are peptide nucleic acid clamps that are highly specific to nontarget templates and which physically block their amplification (Lundberg et al., 2013). These clamps work efficiently even in single-step amplifications, but their production is relatively expensive, which would

FIGURE 6 A comprehensive overview of highly diverse *A. thaliana* leaf microbiomes revealed by parallel amplicon sequencing of eight loci targeting eukaryotic and prokaryotic microbes. Tree branches represent recovered genera and are coloured by taxonomy (left, "Recovered Diversity") and loci from which they were recovered (right, "Target loci specificity"). The percentage of genera found in each data set or by multiple data sets is presented (% of Tips). (a) Eukaryotes were targeted in six loci: Two regions of the 18S rRNA gene (V4–V5 and V8–V9), two regions of the fungal ITS (ITS 1 and 2) and two regions of the oomycete ITS (ITS 1 and 2). The 18S rRNA gene loci revealed the broadest diversity but was complemented by fungi and oomycete-specific primer sets which had more detailed resolution within these groups. (b) Two loci targeting prokaryotes: Two regions of the 16S rRNA gene (V3–V4 and V5–V7) that amplify mostly bacteria revealed a largely overlapping diversity profile complemented by unique discovery of taxa from each of the two target regions
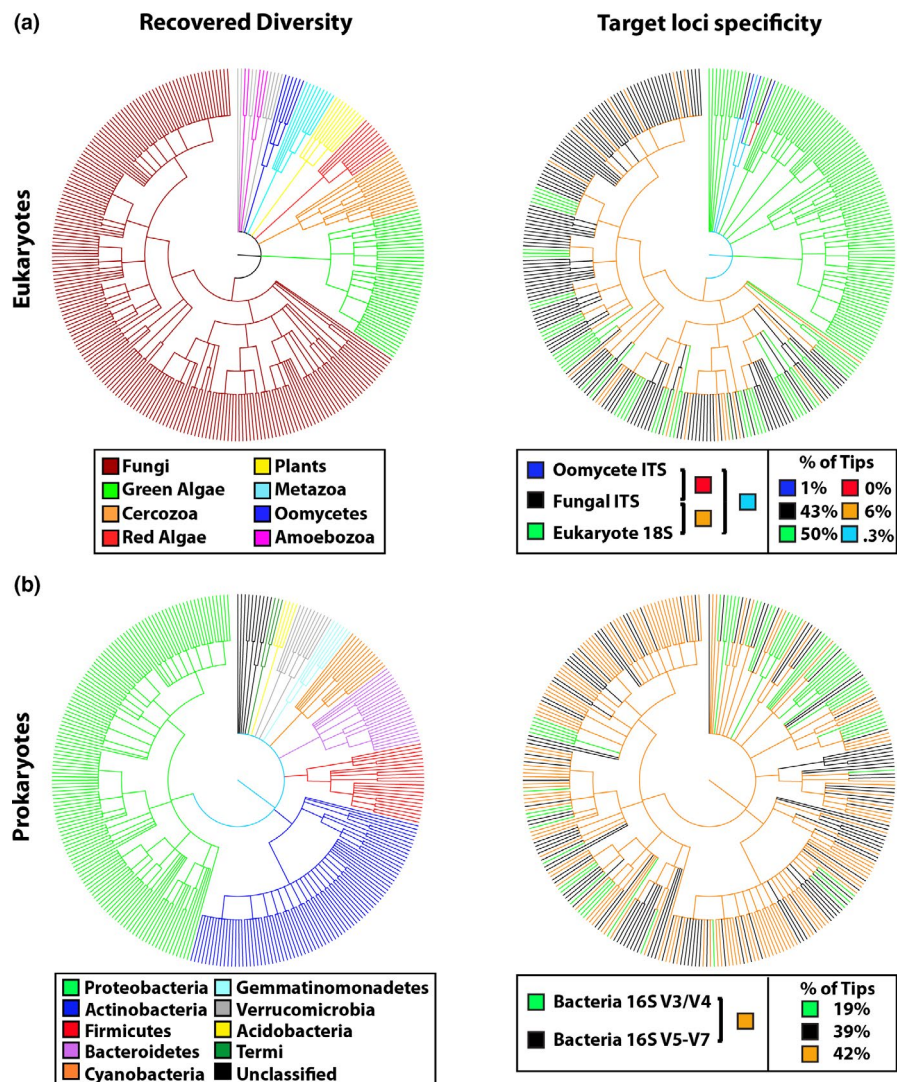

limit rapid development and deployment of multiple clamps for new loci, for blocking multiple nontargets and add major costs for high-throughput projects. For example, our current library preparation costs are estimated at about 2 Euros per library. PNAs, at about 4500 Euro/µmol would add 1.14 Euro or 57% per library. This would only be for host blocking and does not include costs of design and testing of new PNAs for new loci and new nontargets. Other approaches, like using oligonucleotide clamps modified with a C3 spacer (Vestheim and Jarman, 2008) are also costly and work best when they block the universal primer binding site. For many highly conserved target regions, the target and nontarget binding sites are therefore too similar to design specific clamps.

Blocking oligos, which are cheap and flexible, therefore fill an important need for a tool that can be quickly designed and employed for different purposes (e.g., host or microbe blocking). Blocking can also be dropped into practically any pipeline and does not bias results, so it is beneficial to include them when relative abundance of target and nontarget DNA is unknown. Several different blocking techniques have been previously placed under blocking oligos or blocking primers as umbrella terms (Vestheim

and Jarman, 2008). However, we suggest using this term specifically for the blocking oligos we present here, as it most accurately describes their function.

An important question in addition to price and flexibility is whether blocking oligos work as well as PNAs and other methods. Giangacomo et al., (2020a) tested blocking oligos that they designed for maize versus other methods, including PNA clamps and discriminating primers. The blocking oligo pairs they designed blocked organelle DNA as efficiently as discriminating primers and more efficiently than PNA clamps. Most concerning, they found that their blocking oligos distorted microbiota profiles in soil samples, which were used as a diverse template to check for bias. Additionally, they argue that blocking oligos were too species-specific with complicated design and testing steps required for new applications compared to other methods. To address these concerns, we designed and tested new universal plant 16S rRNA gene blocking oligos in leaves, in mock communities and in soil samples and observed no discernible effects on beta diversity. We did observe a desirable increase in alpha diversity in real leaf samples due to blocking the host and recovering more microbial reads. Thus, universal 16S and 18S rRNA gene plant
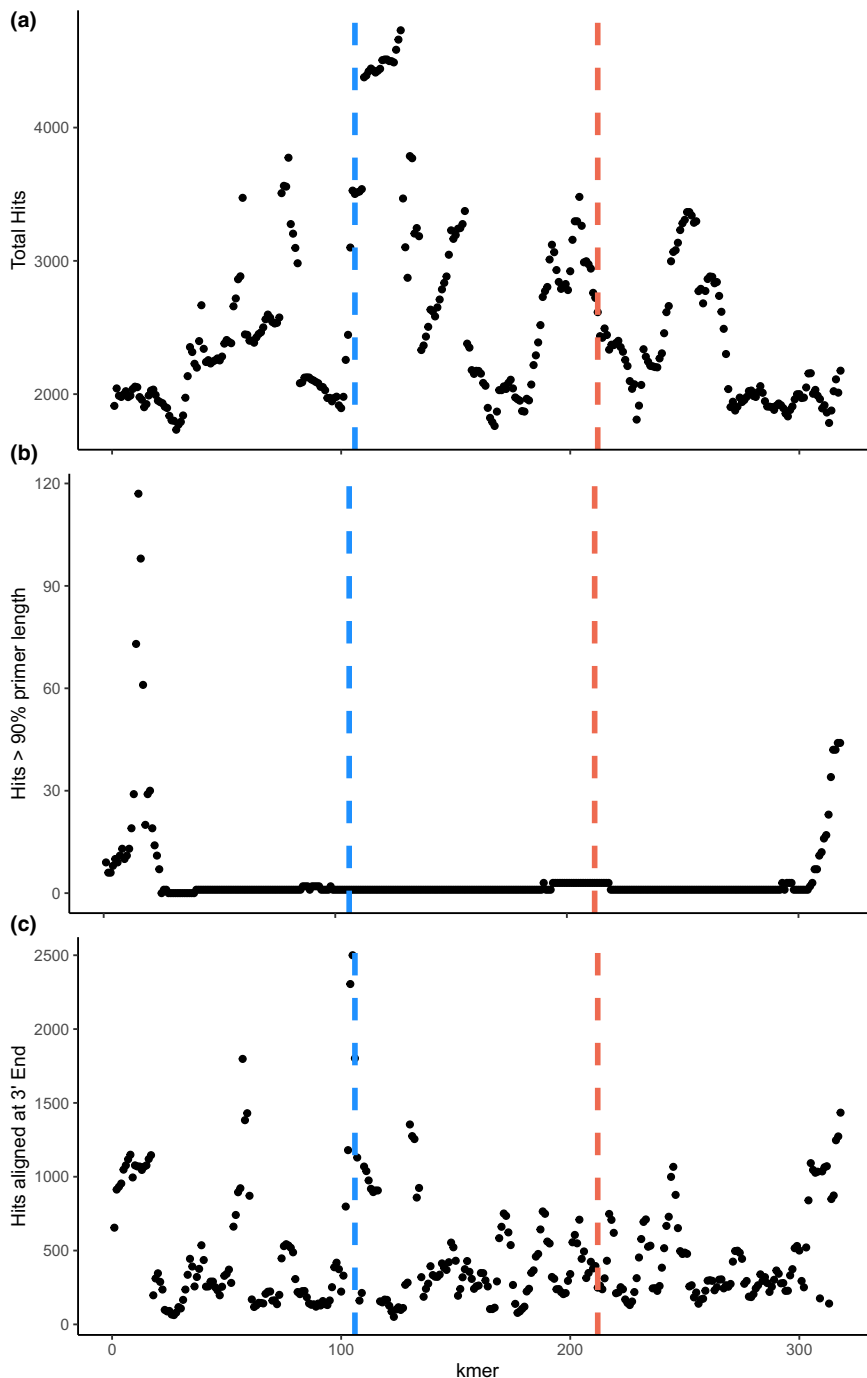
FIGURE 7 Example output from the R package AmpStop when given as input the *A. thaliana* ITS1 sequence (nontarget) and the UNITE fungal ITS database (target). The x-axis represents all possible 30-base oligomers (candidate blocking oligos) along the length of the nontarget sequence. The sequences of the oligos are produced in a separate file. The y-axes show three complementary measures of how likely each candidate oligo is to amplify target sequences, where a hit represents an alignment of the oligo a target. The best candidates will minimize hits and thus be highly specific to nontargets. (a) The total number of hits of each candidate to the target database. Those least likely to amplify targets will have few hits. However, not all hits are equally problematic. Thus, (b) shows oligos that hit a sequence in the target database along >90% of its length, which would increase the chance of amplifying a target organism and (c) shows hits aligned at the 3' end, which are especially problematic because this is the site of polymerase binding. The blue and red guidelines represent one- and two-thirds of the length of the nontarget, respectively, and forward and reverse blocking oligos are probably best chosen before and after these lines, respectively

blocking oligos are effective blockers of all plant species we tested and should also work in maize (Figure S1) and other species, representing an affordable alternative to other methods.

A downside common to blocking oligos, PNAs and other methods is that they need to be designed and tested for different nontargets, which can be cumbersome (Giangacomo et al., 2020a). The R package AmpStop, which we make available here should ease the design process. AmpStop can also be used by researchers who do choose PNAs, as blocking oligos design essentially follows the same procedure[13]. The availability of universal plant blocking oligos that block amplification of most host species will further reduce the need

to make new designs. Notably, we were not able to design universal blocking oligos for the ITS region because diversity between different plant species made it impossible to find a universal blocking oligo set. On the other hand, we and others have observed that the host ITS is not efficiently amplified when there is significant target microbial DNA (Ihrmark et al., 2012). Thus, universal ITS blocking oligos are not as urgently needed as universal 16S rRNA gene blocking oligos.

Studies of leaves of the wild plant *A. thaliana* found that bacterial fraction of extracted DNA are typically very low but range up to about 25% (Regalado et al., 2020). We found that more blocking cycles (15

vs. 10) were necessary to efficiently block nontarget amplification in real leaf samples, but not in mock communities with 5% bacterial DNA. More cycles also lead to discovery of more alpha diversity in real leaf samples. This effect is most likely due to low bacterial loads in real leaf samples, so too few blocking cycles results in libraries that still contain relatively high levels of nontarget contamination. This occurs because nontarget template DNA can be carried over to the extension PCR and are amplified because the concatenated primers contain the universal primer sequence as the binding site (Figure 2). Thus, we recommend increasing the number of blocking cycles when following protocol A or B used here and when no prior information about bacterial loads is available. Alternatively, a linker sequence could be used as the binding site for concatenated primers in the second step Lundberg et al., 2020). This only amplifies amplicons from the first step, not leftover template DNA. Thus, blocking could be minimized to only a few cycles. Lundberg et al. (2020) did apply our ITS blocking oligos in their protocol, demonstrating that blocking oligos can be dropped into most two-step library pipelines.

Some host amplification can be advantageous because it can be used to estimate bacterial loads (Humphrey & Whiteman, 2020), which can change inferred ecological relationships between plant microbiota (Regalado et al., 2020). When we amplified axenic plant samples to which we added known amounts of a bacterial mock community and only blocked chloroplast amplification (not mitochondria), the fraction of bacterial reads was proportional to the bacterial DNA load in mock communities (Figure 5). At our target read depths in the V3–V4 region, mitochondria amplification did not overwhelm the bacterial diversity signal, but this will be locus-specific (in our hands in the V5–V7 region, mitochondrial amplification is more problematic). Generally, estimating bacterial loads from 16S rRNA gene data is not perfect because 16S rRNA gene copy numbers can vary drastically between bacterial species and plastid abundance per cell varies between eukaryotic species (Cole, 2016). HamPCR (Lundberg et al., 2020) is an alternative method utilizing single-copy host genes to gain accurate quantitative insights into microbial load. While that approach is more precise, it does require more steps and may not be suitable for extremely high-throughput studies. Direct estimates from 16S rRNA gene data has the advantage of simplicity and throughput – we have designed dual-indexing primer sets to parallelize up to 500 samples (Supporting Information File S1b). Gaining approximate microbial load information here would allow users to quickly scan for plants, conditions and microbial interactions affecting bacterial load. Thus, we advise using fraction of bacterial reads in 16S rRNA gene data as an initial approximation to gain insight in many samples and then to design specific experiments using more precise measures.

5 | CONCLUSIONS

The realization of the immense complexity of biological systems – and our inability to adequately describe them - has led to many important, unresolved issues. For example, there is ongoing debate about what it means to view macroorganisms as holobionts, since symbiotic microbiota affect host health and fitness (Brucker & Bordenstein, 2013; Sharma et al., 2014). Unanswered questions also remain, such as what causes host genotype-independent taxonomic conservation of plant root microbiomes over broad geographic distances (Hacquard, 2016). Blocking oligos will help researchers to deeply and accurately resolve microbial community diversity when nontarget contamination is problematic, addressing some of the current barriers to progress. Although other challenges remain, we expect this approach to enable researchers to formulate better hypotheses and address currently intractable questions. These advances will thereby assist in increasing discovery of the important roles of microbiota.

AUTHOR CONTRIBUTIONS

Matthew T. Agler and Eric M. Kemen conceptualized the project. Teresa Mayer, Matthew T. Agler, Alfredo Mari, Juliana Almario, Hafiz Syed M. Abdullah and Nina Dombrowski contributed experimental work. Teresa Mayer and Matthew T. Agler designed and wrote the scripts to process and analyse the data and generated figures with input from Juliana Almario, Alfredo Mari, Nina Dombrowski and Stephane Hacquard. Teresa Mayer, Matthew T. Agler and Eric M. Kemen wrote the manuscript. All authors edited and approved the manuscript.

DATA AVAILABILITY STATEMENT

Universal primer sequences, sequencing primers, blocking oligos and concatenated primer sequences are all provided in File S1a and File S1b. Scripts used to generate ASV tables from the raw data, as well as OTU/ASV tables and metadata files to recreate the main figures are being made publicly via Figshare: https://figshare.

com/projects/Obtaining_deeper_insights_into_microbiome_diver sity_using_a_simple_method_to_block_host_and_non-targets_ in_amplicon_sequencing_/89504 (Mayer & Agler, 2020). Raw se quencing data has been made publicly available as a NCBI projects (Agler et al., 2017; PRJNA420016 [https://www.ncbi.nlm.nih.gov/ bioproject/PRJNA420016] and Mayer et al., 2020, PRJNA663775 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA663775/]). AmpStop is freely available on GitHub (https://github.com/magle r1/AmpStop).

ORCID

Teresa Mayer https://orcid.org/0000-0002-6841-2402
Alfredo Mari https://orcid.org/0000-0002-3871-088X
Juliana Almario https://orcid.org/0000-0002-1475-7819
Mariana Murillo-Roos https://orcid.org/0000-0001-9964-7004
Nina Dombrowski https://orcid.org/0000-0003-1917-2577
Stephane Hacquard https://orcid.org/0000-0003-2293-3525
Eric M. Kemen https://orcid.org/0000-0002-7924-116X
Matthew T. Agler https://orcid.org/0000-0003-0732-094X

REFERENCES

Agler, M. T., Mari, A., Almario, J., Dombrowski, N., Hacquard, S., & Kemen, E. M. (2017). Mock communities and Arabidopsis thaliana leaves, Amplification of, 16thS 18S ITS; ed. PRJNA420016: NCBI Sequence Read Archive.

Agler, M. T., Ruhe, J., Kroll, S., Morhenn, C., Kim, S. T., Weigel, D., & Kemen, E. M. (2016). Microbial hub taxa link host and abiotic fac tors to plant microbiome variation. *PLoS Biology*, *14*(1), e1002352. https://doi.org/10.1371/journal.pbio.1002352

Brucker, R. M., & Bordenstein, S. R. (2013). The hologenomic basis of speciation: Gut bacteria cause hybrid lethality in the genus naso nia. *Science*, *341*(6146), 667–669. https://doi.org/10.1126/scien ce.1240659

Cole, L. W. (2016). The evolution of per-cell organelle number. *Frontiers in Cell and Developmental Biology*, *4*, 85. https://doi.org/10.3389/ fcell.2016.00085

de Menezes, A. B., Prendergast-Miller, M. T., Richardson, A. E., Toscas, P., Farrell, M., Macdonald, L. M., Baker, G., Wark, T., & Thrall, P. H. (2015). Network analysis reveals that bacteria and fungi form modules that correlate independently with soil parame ters. *Environmental Microbiology*, *17*(8), 2677–2689. https://doi. org/10.1111/1462-2920.12559

George, P. B. L., Creer, S., Griffiths, R. I., Emmett, B. A., Robinson, D. A., & Jones, D. L. (2019). Primer and database choice affect fungal func tional but not biological diversity findings in a national soil survey. *Frontiers in Environmental Science*, *7*, 173. https://doi.org/10.3389/ fenvs.2019.00173

Giangacomo, C., Mohseni, M., Kovar, L. L., & Wallace, J. (2020a). Comparing DNA extraction and 16S rRNA gene amplification methods for plant-associated bacterial communities. *Phytobiomes Journal*, https://doi.org/10.1094/PBIOMES-07-20-0055-R

Gould, A. L., Zhang, V., Lamberti, L., Jones, E. W., Obadia, B., Korasidis, N., Gavryushkin, A., Carlson, J. M., Beerenwinkel, N., & Ludington, W. B. (2018). Microbiome interactions shape host fitness. *Proceedings of the National Academy of Sciences USA*, *115*(51), E11951–E11960. https://doi.org/10.1073/pnas.18093 49115

Hacquard, S. (2016). Disentangling the factors shaping microbiota com position across the plant holobiont. *New Phytologist*, *209*(2), 454– 457. https://doi.org/10.1111/nph.13760

Hanshew, A. S., Mason, C. J., Raffa, K. F., & Currie, C. R. (2013). Minimization of chloroplast contamination in 16S rRNA gene py rosequencing of insect herbivore bacterial communities. *Journal of Microbiological Methods*, *95*(2), 149–155. https://doi.org/10.1016/j. mimet.2013.08.007

Henriques, S. F., Dhakan, D. B., Serra, L., Francisco, A. P., Carvalho- Santos, Z., Baltazar, C., Elias, A. P., Anjos, M., Zhang, T., Maddocks, O. D. K., & Ribeiro, C. (2020). Metabolic cross-feeding in imbal anced diets allows gut microbes to improve reproduction and alter host behaviour. *Nature Communications*, *11*(1), 4236. https://doi. org/10.1038/s41467-020-18049-9

Humphrey, P. T., & Whiteman, N. K. (2020). Insect herbivory reshapes a native leaf microbiome. *Nature Ecology and Evolution*, *4*(2), 221–229. https://doi.org/10.1038/s41559-019-1085-x

Ihrmark, K., Bödeker, I. T. M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., Strid, Y., Stenlid, J., Brandström-Durling, M., Clemmensen, K. E., & Lindahl, B. D. (2012). New primers to amplify the fungal ITS2 region-evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology and Ecology*, *82*(3), 666–677. https://doi.org/10.1111/j.1574-6941.2012.01437.x

Johnson, E. L., Heaver, S. L., Waters, J. L., Kim, B. I., Bretin, A., Goodman, A. L., Gewirtz, A. T., Worgall, T. S., & Ley, R. E. (2020). Sphingolipids produced by gut bacteria enter host metabolic pathways impacting ceramide levels. *Nature Communications*, *11*(1), 2471. https://doi. org/10.1038/s41467-020-16274-w

Letunic, I., & Bork, P. (2016). Interactive tree of life (ITOL) v3: An on line tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, *44*(W1), W242–245. https://doi. org/10.1093/nar/gkw290

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J. C., Roux, S., Vincent, F., Bittner, L., Darzi, Y., Wang, J., Audic, S., Berline, L., Bontempi, G., Cabello, A. M., Coppola, L., Cornejo-Castillo, F. M., … Raes, J. (2015). Ocean plankton. Determinants of community structure in the global plankton interactome. *Science*, *348*(6237), 1262073. https://doi. org/10.1126/science.1262073

Lundberg, D. S., Na Ayutthaya, P. P., Strauß, A., Shirsekar, G., Lo, W.-S., Lahaye, T., & Weigel, D. (2020). Measuring both microbial load and diversity with a single amplicon sequencing library. *bioRxiv*, https:// doi.org/10.1101/2020.05.19.103937

Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D., & Dangl, J. L. (2013). Practical innovations for high-throughput ampli con sequencing. *Nature Methods*, *10*(10), 999–1002. https://doi. org/10.1038/nmeth.2634

Mayer, T., & Agler, M. T. (2020). Obtaining deeper insights into microbi ome diversity using a simple method to block host and non-targets in amplicon sequencing. Figshare, https://figshare.com/proje cts/Obtaining_deeper_insights_into_microbiome_diversity_us ing_a_simple_method_to_block_host_and_non-targets_in_ampli con_sequencing_/89504

Mayer, T., Murillo-Roos, M., Abdullah, H. S. M., & Agler, M. T. (2020). Development of blocking oligos for avoiding non-target ampli fication in amplicon sequencing; NCBI Sequence Read Archive; PRJNA663775.

Mera, H., & Bourne, D. G. (2018). Disentangling causation: Complex roles of coral-associated microorganisms in dis ease. *Environmental Microbiology*, *20*(2), 431–449. https://doi. org/10.1111/1462-2920.13958

Panke-Buisse, K., Poole, A. C., Goodrich, J. K., Ley, R. E., & Kao-Kniffin, J. (2015). Selection on soil microbiomes reveals reproducible im pacts on plant function. *ISME Journal*, *9*(4), 980–989. https://doi. org/10.1038/ismej.2014.196

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phyloge netics and evolution in R language. *Bioinformatics*, *20*(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412

Prince, D. C., Rallapalli, G., Xu, D., Schoonbeek, H.-J., Cevik, V., Asai, S., Kemen, E., Cruz-Mireles, N., Kemen, A., Belhaj, K., Schornack, S., Kamoun, S., Holub, E. B., Halkier, B. A., & Jones, J. D. G. (2017). Albugo-imposed changes to tryptophan-derived antimicrobial metabolite biosynthesis may contribute to suppression of non-host resistance to phytophthora infestans in Arabidopsis thaliana. *BMC Biology*, 15(1), 20. https://doi.org/10.1186/s12915-017-0360-z

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA Ribosomal RNA Gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41 (Database issue), D590–596. https://doi.org/10.1093/nar/gks1219

Regalado, J., Lundberg, D. S., Deusch, O., Kersten, S., Karasov, T., Poersch, K., Shirsekar, G., & Weigel, D. (2020). Combining whole-genome shotgun sequencing and RRNA gene amplicon analyses to improve detection of microbe-microbe interaction networks in plant leaves. *ISME Journal*, 14(8), 2116–2130. https://doi.org/10.1038/s41396-020-0665-8

Sharma, R., Mishra, B., Runge, F., & Thines, M. (2014). Gene loss rather than gene gain is associated with a host jump from monocots to dicots in the smut fungus melanopsichium pennsylvanicum. *Genome Biology and Evolution*, 6(8), 2034–2049. https://doi.org/10.1093/gbe/evu148

Simhadri, R. K., Fast, E. M., Guo, R., Schultz, M. J., Vaisman, N., Ortiz, L., Bybee, J., Slatko, B. E., & Frydman, H. M. (2017). The gut commensal microbiome of drosophila melanogaster is modified by the endosymbiont wolbachia. *mSphere*, 2(5). https://doi.org/10.1128/mSphere.00287-17

Vestheim, H., & Jarman, S. N. (2008). Blocking primers to enhance PCR amplification of rare sequences in mixed samples - a case study on prey DNA in antarctic krill stomachs. *Frontiers in Zoolgy*, 5, 12. https://doi.org/10.1186/1742-9994-5-12

Werner, J. J., Knights, D., Garcia, M. L., Scalfone, N. B., Smith, S., Yarasheski, K., Cummings, T. A., Beers, A. R., Knight, R., & Angenent, L. T. (2011). Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proceedings of the National Academy of Sciences USA*, 108(10), 4158–4163. https://doi.org/10.1073/pnas.1015676108

Xu, L., Naylor, D., Dong, Z., Simmons, T., Pierroz, G., Hixson, K. K., Kim, Y.-M., Zink, E. M., Engbrecht, K. M., Wang, Y. I., Gao, C., DeGraaf, S., Madera, M. A., Sievert, J. A., Hollingsworth, J., Birdseye, D., Scheller, H. V., Hutmacher, R., Dahlberg, J., ... Coleman-Derr, D. (2018). Drought delays development of the sorghum root microbiome and enriches for monoderm bacteria. *Proceedings of the National Academy of Sciences USA*, 115(18), E4284–E4293. https://doi.org/10.1073/pnas.1717308115

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.