Targeted rRNA depletion enables efficient mRNA sequencing in diverse bacterial species and complex co-cultures

Kellie A. Heom[a,b,*] , Chatarin Wangsanuwat[a,b,*], Lazarina V. Butkovich[a], Scott C. Tam[a], Annette R. Rowe[c], Michelle A. O'Malley[a,b,#] and Siddharth S. Dey[a,b,d,#]

[a] Department of Chemical Engineering, University of California Santa Barbara, Santa Barbara, CA 93106, USA.

[b] Biological Engineering Program, University of California Santa Barbara, Santa Barbara, CA 93106, USA.

[c] Biological Sciences, University of Cincinnati, Cincinnati, OH 45221, USA.

[d] Neuroscience Research Institute, University of California Santa Barbara, Santa Barbara, CA 93106, USA.

* These authors contributed equally to this work.

[#] Correspondence to: Michelle A. O'Malley (momalley@ucsb.edu) and Siddharth S. Dey (sdey@ucsb.edu).

ABSTRACT

Bacterial mRNA sequencing is inefficient due to the abundance of ribosomal RNA that is challenging to deplete. While commercial kits target rRNA from common bacterial species, they are frequently inefficient when applied to divergent species, including those from environmental isolates. Similarly, other methods typically employ large probe sets that tile the entire length of rRNAs; however, such approaches are infeasible when applied to many species. Therefore, we present EMBR-seq+, which requires fewer than ten oligonucleotides per rRNA by combining rRNA blocking primers with RNase H-mediated depletion to achieve rRNA removal efficiencies of up to 99% in diverse bacterial species. Further, in more complex microbial co-cultures between *F. succinogenes* strain UWB7 and anaerobic fungi, EMBR-seq+ depleted both bacterial and fungal rRNA, with a 4-fold improvement in bacterial rRNA depletion compared to a commercial kit, thereby demonstrating that the method can be applied to non-model microbial mixtures. Notably, for microbes with unknown rRNA sequences, EMBR-seq+ enables rapid iterations in probe design without requiring to start experiments from total RNA. Finally, efficient depletion of rRNA enabled systematic quantification of the reprogramming of the bacterial transcriptome when cultured in the presence of the anaerobic fungi *Anaeromyces robustus* or *Caecomyces churrovis.* We observed that *F. succinogenes* strain UWB7 downregulated several lignocellulose-degrading

carbohydrate-active enzymes in the presence of anerobic gut fungi, suggesting close interactions between two cellulolytic species that specialize in different aspects of biomass breakdown. Thus, EMBR-seq+ enables efficient, cost-effective and rapid quantification of the transcriptome to gain insights into non-model microbial systems.

INTRODUCTION

Microbes remain one of nature's largest and most diverse sources of biochemical transformations. Many bacterial processes, such as their ability to grow on unprocessed feedstocks, are interesting as their mechanisms for releasing sugars or other metabolites could lead to advances in bio-based fuels or other value-added chemicals (1–4). mRNA sequencing (mRNA-seq) provides an opportunity to gain insights into these biochemical reactions as it enables a quantitative and comprehensive view of the transcriptome, which underlies all biological activity. For example, mRNA-seq has been used to quantify the expression of proteins, transporters, and carbohydrate active enzymes (CAZymes) that are upregulated by *Fibrobacter succinogenes* strain S85 during cellulose utilization compared to when grown on glucose (5). Furthermore, *F. succinogenes* and other cellulolytic strains have been isolated in nature from complex communities, such as the rumen microbiome, and therefore it is also critical to characterize how interactions between different microbial species mutually tune their expression (6–8). Thus, efficient sequencing of mRNA from non-model bacteria in microbial consortia will enable a deeper understanding of the role of each member in these complex communities.

However, it is more challenging to perform mRNA-seq on bacteria, and especially those that exist in co-cultures. This is because mRNA-seq is inherently inefficient in bacteria, due to limitations associated with removing ribosomal RNA (rRNA). In eukaryotic species, the mRNA contains a poly-A tail at the 3'-end enabling poly-T primers to selectively reverse transcribe the mRNA and not the rRNA. However, prokaryotic mRNA does not have a 3'-end poly-A tail, making it more challenging to uniquely capture mRNA compared to rRNA or other non-coding RNA molecules. If left undepleted, rRNA can consume up to 90% of the sequencing data with uninformative reads, resulting in highly inefficient sequencing of mRNA (9). Therefore, efficient, high-throughput and cost-effective rRNA depletion is key to obtaining high-resolution maps of gene expression programs in bacterial species.

To accomplish this, several commercial kits for prokaryotic rRNA depletion have been developed, including the Ribo-Zero Plus Microbiome rRNA Depletion Kit (Illumina), RiboMinus™ Bacteria 2.0 Transcriptome Isolation Kit (Thermo Fisher), and the NEBNext rRNA Depletion Kit for Bacteria (NEB). These kits leverage sequence-specific rRNA removal, which is highly efficient at depleting the target rRNA. However, these kits are limited by the breadth of their probe sets, which restricts their application in diverse microbial systems. Further, the cost of these kits typically ranges from approximately $30-80 per sample. In response to the lack of specificity to varied bacterial species and the high cost, many research groups have also developed methods for rRNA depletion. A majority of these methods either employ biotinylated oligonucleotides that

hybridize with rRNA that are then removed using streptavidin-coated paramagnetic beads (10, 11), or oligonucleotides that hybridize with rRNA and act as a guide for RNase H-based enzymatic digestion (12–14). Similarly, CRISPR-Cas9 mediated degradation of rRNA has also been used to sequence bacterial mRNA (15). However, all these methods use large probe sets, typically up to hundreds of probes per bacterial species, to achieve depletion along the full length of each rRNA subunit. An alternative approach to large oligonucleotide or guide RNA arrays is to use the terminator 5'-phosphate-dependent exonuclease, which selectively degrades rRNA containing 5'-monophosphate ends while having minimal activity on mRNA molecules with 5'-triphosphate ends (16). However, this method has previously been shown to be less efficient at depleting rRNA compared to the other techniques discussed above (17, 18) (Supplementary Table S1).

To overcome challenges associated with the high cost and the time intensive activity of developing large probe sets for every new bacterial species profiled, we previously published a method, EMBR-seq (Enriched mRNA by Blocked rRNA) (18) that only requires 3-4 30-nt primers per rRNA species (approximately 9 primers per bacterial species) for rRNA depletion. EMBR-seq used a combination of 3'-end polyadenylation of prokaryotic mRNA together with specifically designed blocking primers to prevent the downstream amplification of rRNA molecules. In this study, we have developed a significantly improved method, EMBR-seq+, that integrates our previous method with RNase H-based digestion to deplete rRNA at high efficiencies while still employing fewer than ten oligonucleotides for each rRNA species. Further, we show that EMBR-seq+ can be easily applied to diverse bacterial species, and in particular the method enables rapid iterations in probe design to deplete rRNA for microbial species with poorly annotated genomes and unknown rRNA sequences. We demonstrate this in a co-culture system of anaerobic bacteria and fungi native to the rumen microbiome and show that EMBR-seq+ is instrumental in providing deeper coverage of the bacterial transcriptome. Finally, we also investigate transcriptional changes in the bacteria grown under monoculture or co-culture conditions to gain a deeper understanding of how cell-cell interactions between microbial species affect the CAZyme expression profile in these bacteria. In summary, EMBR-seq+ is a highly efficient, high-throughput and cost-effective platform for sequencing bacterial mRNA that also allows for rapid improvements in probe design to deplete rRNA from poorly annotated and non-model microbial species.

RESULTS

EMBR-seq+ combines blocking primers and RNase H digestion to efficiently deplete rRNA

EMBR-seq+ builds off the EMBR-seq method we have previously developed, in which rRNA depletion is achieved using a small set of short blocking primers designed to hybridize at the 3'-end of 16S, 23S, and 5S rRNA in the bacterial species of interest (Figure 1A). Briefly, in EMBR-seq, total RNA is polyadenylated in the presence of the rRNA blocking primers to preferentially polyadenylate mRNA. The blocking primers not only target the 3'-end of rRNA but also a few "hotspot" locations that are enriched with reads within the rRNA sequence (Supplementary Table S2). For *E. coli*, the hotspots present in the "No depletion" control library are targeted and efficiently depleted in the other libraries using hotspot-specific blocking primers (See inset in Figure 1B). Next, a poly-T primer containing a sample barcode, Illumina RA5 sequence, and T7 promoter in its overhang is used to reverse transcribe the polyadenylated molecules. Subsequently, the reverse transcription product is converted to cDNA and then amplified by *in vitro* transcription (IVT) using the T7 polymerase. As a majority of the rRNA molecules are reverse transcribed by the blocking primers instead of the poly-T primers, rRNA-derived cDNA molecules lack a T7 promoter and are therefore excluded from amplification during IVT. The resulting amplified RNA (aRNA) is converted to the final Illumina library by incorporating the Illumina RA3 sequence during a subsequent reverse transcription step followed by addition of the Illumina P5 and P7 adapters during PCR (Figure 1A).

In EMBR-seq+, we introduced an additional RNase H treatment step to specifically target remaining rRNA-derived molecules that are present in the aRNA after IVT. In this step, 50-nt ssDNA probes are hybridized near the junction of the Illumina RA5 adapter with the rRNA sequence, with additional probes targeting 3-4 hotspot locations along the rRNA-derived aRNA molecules. The probes form aRNA-DNA duplexes on the rRNA-derived aRNA molecules which are selectively degraded with RNase H, resulting in either the removal of the Illumina RA5 adapter at the 5' end of these molecules or the generation of very short fragments that are excluded during the downstream PCR amplification or size selection steps, respectively, thereby preventing their incorporation into the final Illumina library. As the RNase H treatment is performed on aRNA molecules to cleave the partial RA5 Illumina adapter, EMBR-seq+ requires only 3-4 ssDNA probes per rRNA species (approximately 9 per bacterial species) compared to most other RNase H-based depletion methods that typically need to tile entire rRNA molecules with up to hundreds of ssDNA probes per bacterial species (10–15). After RNase H treatment, the rRNA-depleted aRNA is reverse transcribed and then amplified with PCR to obtain the final Illumina libraries (Figure 1A). We tested two enzymes and reaction temperatures: (1) Invitrogen RNase H at 16°C

and (2) Hybridase Thermostable RNase H at 45°C. We found that for both the "RNase H" only and EMBR-seq+ libraries, the rRNA depletion efficiency was very similar with the two RNase H enzymes, and decided to proceed with the Invitrogen RNase H for all subsequent experiments (Supplementary Figure S1).

As proof-of-principle, we first applied these methods to *E. coli* where rRNA was depleted to 12.1% of the sequenced reads when treated with RNase H alone, and to 3.7% with EMBR-seq, and 3% with EMBR-seq+, suggesting that the combined use of blocking primers and RNase H depletion in EMBR-seq+ can remove most of the rRNA-derived molecules from the sequencing libraries (Figure 1C). This efficiency of rRNA depletion in EMBR-seq+ is similar to that achieved by other methods and commercial kits (Supplementary Table S1). In the libraries in which only RNase H was used for rRNA depletion, no blocking primers were added during the poly-A tail synthesis and initial reverse transcription step. Similarly, in the "No depletion" control, neither blocking primers nor RNase H depletion was performed. Further, as expected, for the same sequencing depth, increased depletion of rRNA in EMBR-seq+ resulted in the detection of the most unique genes above every expression threshold over three orders of magnitude (Figure 1D). Finally, gene expression levels quantified by EMBR-seq+ were highly correlated to the "No depletion" libraries (Pearson r = 0.95), suggesting that the use of blocking primers and RNase H in EMBR-seq+ does not introduce technical biases (Figure 1E and Supplementary Table S3). Collectively, these experiments show that EMBR-seq+ can efficiently deplete rRNA from bacterial total RNA without introducing technical artifacts to quantify the transcriptome at high depth.

EMBR-seq+ can be easily applied in distant prokaryotic species

EMBR-seq+ requires fewer than 20 ssDNA oligonucleotides per bacterial species (fewer than 10 blocking primers and fewer than 10 probes for RNase H depletion) to deplete 16S, 23S, and 5S rRNA, and therefore it is easy and inexpensive to customize the method for any species of interest (Supplementary Table S2). We tested EMBR-seq+ on *Fibrobacter succinogenes* strain UWB7, a strain of cellulolytic bacteria isolated from cow rumen, and *Geobacter metallireducens*, a strain that reduces Fe(III) oxide and can exhibit syntropy with methanogens via direct interspecies electron transfer (19–21)*.* When employing EMBR-seq, rRNA was depleted to 10.3% in *F. succinogenes* strain UWB7 and 3.1% in *G. metallireducens* while EMBR-seq+ depleted rRNA to 7.6% in *F. succinogenes* strain UWB7 and 1.1% in *G. metallireducens*, highlighting the improved efficiency of rRNA depletion when combining blocking primers and RNase H digestion in EMBR-seq+ (Figure 2A). Further, as with *E. coli*, custom-designed blocking primers and RNase H probes were key to efficiently depleting rRNA reads from hotspot loci observed in the "No depletion"

controls for both *F. succinogenes* strain UWB7 and *G. metallireducens* (Figure 2B and Supplementary Figure S2A). Consistent with the improved rRNA depletion, an increased number of unique genes was detected in EMBR-seq+ above every expression threshold for both *F. succinogenes* strain UWB7 and *G. metallireducens* (Figure 2C and Supplementary Figure S2B). Finally, when compared to the "No depletion" controls, the blocking primers and RNase H probes did not introduce technical bias in the quantification of gene expression in EMBR-seq+ (Pearson r = 0.94 and r = 0.97 for *F. succinogenes* strain UWB7 and *G. metallireducens*, respectively) (Figure 2D, Supplementary Figure S2C, and Supplementary Table S3). These results show that the design of a small number of short probes in EMBR-seq+ can be used to easily extend the method with high efficiency to varied bacterial species.

EMBR-seq+ enables efficient prokaryotic mRNA sequencing in a fungal-bacterial co-culture system

Next, we aimed to apply EMBR-seq+ to more complex systems that mimic microbial communities in nature, such as co-cultures of bacteria with different fungal species. As an example, anaerobic organisms isolated from the rumen microbiome are particularly interesting because of their ability to grow on raw, unprocessed substrates such as grass and depolymerize lignocellulose, a feature attributed to the large numbers of carbohydrate-active enzymes (CAZymes) found in their genomes (5, 6, 22, 23). Moreover, as these species exist in consortia in nature rather than in isolation, understanding interactions between different species is key to gaining insights into their function in native environments and for engineering stable microbial systems *in vitro* (24–26). A previous study examined transcriptomic changes to the anaerobic gut fungi *Anaeromyces robustus* or *Caecomyces churrovis* when cultured together with *F. succinogenes* strain UWB7 compared to monoculture conditions and identified several novel components of the fungal secondary metabolism that were upregulated in co-culture (26). This study focused primarily on the remodeling of the fungal transcriptome induced by the presence of bacteria; however, this work did not characterize transcriptional changes in the bacteria *F. succinogenes* strain UWB7 in co-culture with *A. robustus* due to two technological limitations associated with (1) the low abundance of bacterial RNA compared to the fungi, and (2) the lack of efficient bacterial rRNA depletion.

To overcome these limitations, we focused on capturing the bacterial transcriptome by applying EMBR-seq+ to two independent co-cultures of *F. succinogenes* strain UWB7 with different anaerobic gut fungal partners, *A. robustus* or *C. churrovis* (23, 27, 28). In our initial co-culture experiments with *F. succinogenes* strain UWB7 and *A. robustus*, we applied EMBR-seq

to total RNA extracted from co-culture pellets (denoted as "Unprocessed total RNA") to find only 11.8% of the mapped reads derived from the bacteria with the remaining 88.2% of reads derived from the fungi (Figure 3A,B). We hypothesized that this large excess of fungal reads in our dataset derived from an initial overabundance of fungal RNA in the co-culture total RNA pellets, and a lack of targeting to deplete fungal RNA in EMBR-seq+. Therefore, to further enrich the bacterial RNA fraction, we first used poly-T magnetic beads to deplete polyadenylated fungal mRNA prior to applying EMBR-seq, resulting in a library (denoted as "Poly-A depleted") with 24.9% bacteria reads.

To increase the fraction of bacterial reads beyond the 2.1-fold achieved over "Unprocessed total RNA", we next attempted to deplete the fungal rRNA. However, as with many non-model microbes, a major challenge was that the rRNA sequences were not annotated for either of the fungal genomes. Therefore, we first mapped the sequencing data from the "Poly-A depleted" sample against the reference genome scaffolds and identified a series of adjacent pileups of mapped reads on the same scaffold that were spaced apart according to the expected sizes of the 18S, 28S and 5.8S eukaryotic rRNA subunits (Figure 3C). Thus, we inferred that the three genomic loci on this scaffold correspond to the fungal rRNA subunits and designed five RNase H probes to specifically deplete the RNA transcribed from these locations (Supplementary Table S2). We also aligned our probe sequences against the rRNA annotations of another closely related fungal species, *Orpinomyces sp OUS1*, and confirmed that they align to the 3'-ends of 18S, 28S, and 5.8S rRNA (29). By combining the upstream poly-A bead-based fungal mRNA depletion prior to EMBR-seq+ with the RNase H-based fungal rRNA depletion during EMBR-seq+ (denoted as "Poly-A depleted & RNase H treated"), we were able to successfully increase the percentage of bacterial reads in the final sequencing libraries to 56.8%, a 1.75-fold improvement over previous work (Supplementary Figure S3A) (26). Further, we also applied these two fungal depletion treatments to the co-culture of *F. succinogenes* strain UWB7 with *C. churrovis* to observe a similar enrichment of bacteria-derived reads to 41.3% in EMBR-seq+, suggesting that this is a robust approach to quantifying the bacterial transcriptome in complex microbial communities (Supplementary Figures S4 and S5). More generally, the fungal rRNA depletion strategy demonstrates that the RNase H digestion component of EMBR-seq+ enables rapid iterations of probe designs; this step is implemented on a small amount of amplified aRNA, an intermediate product in EMBR-seq+, thereby offering a convenient approach to circumvent the need to prepare new libraries from total RNA or re-grow the co-cultures.

After successfully enriching for bacterial RNA from bacteria-fungal co-cultures, we next quantified the depletion of bacterial rRNA within these co-cultures. As before, we observed that

EMBR-seq+ resulted in the highest depletion of *F. succinogenes* strain UWB7 rRNA compared to "RNase H" or "EMBR-seq" for both co-culture systems, demonstrating that the combination of blocking primers and RNase H digestion is highly effective in depleting bacterial rRNA (Figure 3D). Furthermore, the rRNA depletion efficiency of EMBR-seq+ was significantly improved over a previous transcriptomic study using the same co-culture pairings (5.1- and 3.9-fold increased rRNA depletion for co-cultures with *A. robustus* and *C. churrovis*, respectively) (Supplementary Figure S3). Similarly, compared to the other methods, we observed that EMBR-seq+ detected the most genes above all expression thresholds and did not introduce technical bias in quantifying gene expression levels (Supplementary Figure S6). Overall, these results show that EMBR-seq+ can be easily adapted to efficiently enrich for bacterial mRNA from complex microbial communities.

Differential gene expression between monocultures and co-cultures

After establishing that EMBR-seq+ can provide a detailed view of bacterial gene expression programs within co-cultures, we next wanted to characterize the transcriptional response of *F. succinogenes* strain UWB7 in the presence of the two different strains of anaerobic fungi. To accomplish this, we used DESeq2 to perform differential gene expression analysis for *F. succinogenes* strain UWB7 grown in monoculture *vs.* co-culture. While 140 genes were upregulated for *F. succinogenes* strain UWB7 in monoculture and 128 genes were upregulated in co-culture with *A. robustus* for the "No depletion" datasets, we identified 224 upregulated genes in monoculture and 374 upregulated genes in co-culture for the "EMBR-seq+" datasets (Figure 3E, Supplementary Figures S7 and S8). We observed similar robust results for experiments with *C. churrovis*, with 113 genes upregulated in monoculture and 158 genes upregulated in co-culture for the "No depletion" datasets, while bacterial RNA enriched and rRNA depleted datasets obtained from EMBR-seq+ displayed 247 upregulated genes in monoculture and 149 upregulated genes in co-culture (Supplementary Figures S7 and S9a, and Supplementary Table S4).

To better understand the deconstruction of carbon substrates by *F. succinogenes* strain UWB7, we used dbCAN2 (30, 31) to predict CAZyme annotations in the bacterial transcriptome. Within dbCAN2, we performed three predictive searches: (1) HMMER against the dbCAN database, (2) DIAMOND against the CAZy database, and (3) HMMR against the dbCAN_sub database. 191 genes were predicted to be CAZymes in at least two out of the three tools, and we found that 189 of these genes were expressed with an RPM of at least 2 in all the monocultures and co-culture replicates. Notably, in co-cultures with both *A. robustus* and *C. churrovis*, CAZymes represented a substantial fraction of the differentially regulated genes. For example, of

the 191 CAZyme genes, 32 (17%) were differentially expressed, with 17 upregulated in monoculture and 15 upregulated in co-culture with *A. robustus* (Figure 3F). In addition to CAZymes, *F. succinogenes* strain UWB7 also upregulated several molecular chaperones and components of toxin-antitoxin systems when grown in co-culture with either fungal species (Supplementary Table S5). Finally, many of the transporters found by Swift *et al.* to be upregulated in co-culture were also upregulated in our EMBR-seq+ datasets (Supplementary Table S6) (26). Overall, this expression profile highlights elements of stress response in the bacteria in the presence of the anaerobic gut fungi.

Interestingly, in monoculture and co-culture, the bacteria upregulated different subclasses of CAZymes. In monoculture, *F. succinogenes* strain UWB7 upregulated a number of lignin-active CAZymes in the glycoside hydrolase (GH), polysaccharide lyase (PL), carbohydrate esterase (CE), and carbohydrate binding module (CBM) subclasses (Figure 3F and Supplementary Table S7). Conversely, in co-cultures, the bacteria did not upregulate as many CAZymes in these classes and instead upregulated CAZymes in the glycosyl transferase (GT) subclass. While the GH, PL, and CE classes represent enzymes that hydrolyze or cleave different types of bonds found in lignocellulose, the functions of GT enzymes include forming new glycosidic bonds to modify or polymerize sugars (32). The GT enzymes found to be upregulated in the bacteria in co-culture have predicted functions for cell wall biosynthesis, suggesting that the cells could be incorporating sugars in their peptidoglycan layers. Thus, the dramatic transcriptomic shift away from lignocellulose breakdown observed in *F. succinogenes* strain UWB7 in the presence of anaerobic fungi highlights that the interaction between these two species, which are both known to be efficient lignocellulose degraders, substantially alters the bacterial response. Further, out of the thousands of microbial members identified in the cattle rumen, members of the Fibrobacteres phylum have been found to devote the largest proportion of their proteome towards carbohydrate metabolism; therefore, the downregulation of several lignocellulose-modifying CAZymes in the presence of fungi potentially suggests that *A. robustus* releases sugars in such great excess that the *F. succinogenes* strain UWB7 redirect their efforts away from sugar hydrolysis (6, 33). Overall, these findings suggest that *F.* succinogenes strain UWB7 may not need to reach their full lignocellulose-degrading potential in this co-culture system, and from an engineering perspective, a more efficient bioreactor might pair either the *F. succinogenes* strain UWB7 or the *A. robustus* with another organism that has complementary behavior, such as the ability to ferment sugars or produce fatty acid products (34–36).

In the co-culture experiments with *C. churrovis*, a similar but less pronounced shift in CAZyme expression profile was observed for *F. succinogenes* strain UWB7 (Supplementary

Table S7, Supplementary Figure S9B). For example, compared to co-culture experiments with *A. robustus*, fewer GH CAZymes were upregulated in monoculture and a few GH genes were also upregulated in co-culture. These observations potentially reflect that co-cultures with different fungal strains result in altered substrate availability for *F. succinogenes* strain UWB7. This is consistent with recent work showing that *C. churrovis* does not release glucose and xylose in concentrations nearly as high as *A. robustus*, and the lower levels of excess reducing sugars could result in distinct signaling and transcriptome remodeling in the bacteria (36). This is likely due to the fact that *C. churrovis* lacks the rhizoid system found in most other anaerobic fungi, which limits its ability to degrade fibrous biomass (27, 28). Additionally, *F. succinogenes* strain UWB7 in co-culture upregulated a few xylan esterase genes CE6 and CBM+CE6 fusion genes. As these genes were not differentially expressed in co-culture with *A. robustus*, this result agrees with previous findings that *C. churrovis* is unique from other anaerobic fungi because it is a xylan-degrading specialist. (27, 37, 38). Fibrobacteres do not have genes for xylan transport or metabolism, and previous work has suggested that their xylan machinery is used to gain access to cellulose (19); therefore, our result suggests a relationship wherein when grown on switchgrass with *C. churrovis*, there is either a direct or indirect benefit to the bacteria for expressing xylan esterase genes. These transcriptional responses in the bacteria highlight how high-quality mRNA-seq from EMBR-seq+ provides a detailed perspective of complex co-culture systems and contributes to our understanding of how these species interact both with each other and also with the substrate present.

DISCUSSION

In this study, we have developed EMBR-seq+, a method that depletes rRNA in prokaryotic species at high efficiencies to enable cost-effective and improved mRNA sequencing. To accomplish this, EMBR-seq+ uses a combination of selective polyadenylation on mRNA molecules, blocking primers that prevent downstream amplification of rRNA molecules, and RNase H digestion targeted at rRNA-derived amplified RNA. Importantly, the method is designed to only require a single blocking primer per rRNA sequence at the 3'-end, together with 3-4 additional blocking primers at "hotspots" and 3-4 probes to direct RNase H digestion of rRNA. By requiring blocking primers and RNase H probe targets at only a few strategic locations, EMBR-seq+ is significantly easier to implement, less expensive, and can be rapidly adapted to any bacterial species, compared to alternate methods that require large primer sets to tile the entire length of rRNA molecules (12–15).

A major change that we introduced in EMBR-seq+ is the inclusion of RNase H digestion downstream of IVT amplification. A key advantage of the RNase H treatment in EMBR-seq+ over other RNase H-based methods is that it is performed on IVT-amplified RNA rather than the starting total RNA; therefore, if the sequencing results from a previous run indicate the continued presence of one or more highly abundant rRNA fragments, additional RNase H probes can be designed and applied to the same IVT-amplified RNA without having to start again from total RNA, thereby significantly reducing experimental time. We demonstrate this approach in the depletion of fungal rRNA in the co-culture experiments (Figure 3). Additionally, a limitation of current single-cell bacterial RNA-seq methods is that they do not deplete rRNA; therefore, in the future we anticipate that this RNase H digestion strategy can be used to readily deplete rRNA from single-cell RNA-seq methods that have an intermediate IVT amplification step (39–43).

In this study, we also showed that EMBR-seq+ can be applied to complex microbial communities, such as a co-culture system of anaerobic bacteria and fungi native to the rumen microbiome to deplete rRNA efficiently from both species. In a previous study, the Ribo-Zero Gold rRNA Removal Kit (Epidemiology) was used for rRNA depletion from the same bacterial and fungal co-cultures; however, rRNA reads contributed to more than 90% of the sequencing data, thereby requiring significantly higher sequencing depths to obtain sufficient coverage of the fungal and bacterial transcriptomes (26). This reduced depletion efficiency potentially arises from the kit's lack of specificity towards non-model microbes due to sequence divergence, highlighting the importance of methods such as EMBR-seq+ that can be rapidly adapted to any species. Further, in the aforementioned study, co-cultures of *A. robustus* and *F. succinogenes* strain UWB7 were sequenced to a depth of ~15 million reads to obtain only 20,000 bacterial mRNA reads, making it

impractical to perform differential expression analysis on the bacterial RNA. In contrast, EMBR-seq+ libraries were sequenced to a depth of ~12 million reads to obtain ~1.9 million *F. succinogenes* strain UWB7 mRNA reads, representing a dramatic improvement in quantifying the bacterial transcriptome (Supplementary Figure S3). Overall, this demonstrates that the small number of probes required in EMBR-seq+ enables rapid application of this method to varied microbial species while retaining high rRNA depletion efficiency.

Finally, to gain new insights into how *F. succinogenes* strain UWB7 is affected in co-culture with anaerobic gut fungi, we used EMBR-seq+ to quantify transcriptional changes in the bacteria when transitioned from monoculture to co-culture conditions. In previous work, Swift *et al.* designed this co-culture system for the bacteria to exert pressure on the fungi and induce the production of fungal secondary metabolites; however, we found that the bacteria also exhibit a substantial stress response and changes to its gene expression program (26). We observed a transcriptional shift in the bacterial CAZyme profile, marked by a decrease in the expression of lignocellulose-active CAZymes in the presence of the anaerobic fungi. In agreement with this, a previous study showed that microbiome samples enriched in anaerobic fungi (including *A. robustus*) release excess reducing sugars in the growth supernatant that potentially acts as a nutrient source for other organisms in its environment (33, 36). Similarly, this is consistent with another previous finding where in fungal-dominated microbiome cultures (achieved by antibiotic treatment), antibiotic-resistant bacteria were found to survive on reed canary grass, despite the absence of CAZymes in their genomes (37). Thus, in general, efficient rRNA depletion using EMBR-seq+ enables a deeper view of the bacterial transcriptome to gain functional insights into the response of bacteria in complex cultures.

Overall, EMBR-seq+ enables low-cost and efficient bacterial mRNA sequencing that is accomplished through rRNA depletion using only a few short blocking primers and RNase H targeting probes that are simple to optimize for any bacterial species of interest. Further, we also demonstrate that EMBR-seq+ can efficiently sequence bacterial mRNA from non-model microbial co-culture systems containing eukaryotic species, thereby enabling detailed quantification of the interaction between the two species. Finally, given the low cost, ease of design and efficiency of this method, we anticipate that high-throughput screens employing EMBR-seq+ offer a strategy for rapidly understanding microbial interactions.

MATERIALS AND METHODS

E. coli cell culture

Overnight cultures of *Escherichia coli MG1655* (ATCC: 700926) were inoculated into fresh LB medium at 1:1000 and grown at 37°C with shaking (225 rpm). After 4 hours the cells reached the exponential growth phase (verified by OD_{600} measurement) and the culture was centrifuged at 3000 g and 4°C for 10 mins. The media was removed and the pellet was resuspended in PBS to a concentration of 10^7 cells/µL. For RNA extraction, 5×10^7 cells were added to 750 µL Trizol, mixed, and stored at -20°C.

G. metallireducens culture

Cultures of *Geobacter metallireducens* GS-15 (DSM 7210) were grown at 30°C in an anaerobic basal salt media (1.3 mM KCl, 4.6 mM NH_4Cl, 4.4 mM NaH_2PO_4, and 30 mM $NaHCO_3$) amended with 56 mM iron citrate and 20 mM acetate. Wolf's vitamin mix (from DSMZ protocol 141) and metal reducer mineral mix (described by Bretschger *et al*.) were added from 100X stock to 1X final concentrations (44). All cultures were grown in 20 mL volumes in anaerobic serum vials from 1% culture transfers. Cultures were maintained in DMSO at -80°C when not in use. For harvesting the cells, 1 mL of culture was centrifuged at 8000 g and 4°C for 5 mins to obtain a pellet. The media was removed and the pellet was resuspended in 750 µL of Trizol, mixed, and stored at -20°C.

F. succinogenes strain UWB7 monoculture & co-culture with anaerobic gut fungi

Fibrobacter succinogenes strain UWB7 monoculture and co-culture with anaerobic gut fungi were prepared based on the experimental protocol described by Swift *et al.* (26). *F. succinogenes* strain UWB7 was previously gifted by Garret Suen at the University of Wisconsin-Madison, with isolation of the strain described by Neumann and Suen (45). The anaerobic gut fungi *Anaeromyces robustus* and *Caecomyces churrovis* were previously isolated via reed canary grass enrichment from fecal pellets of a Churro sheep at the Santa Barbara Zoo (22, 23, 27).

F. succinogenes strain UWB7, *A. robustus*, and *C. churrovis* were cultivated at 39°C using standard anaerobic techniques (46). Cultures were prepared with a CO_2 headspace in Hungate tubes containing 9.0 mL of a modified formulation (MC-) of complex medium C (47), containing 0.25 g/L yeast extract, 0.5 g/L Bacto Casitone, and 7.5 vol% clarified rumen fluid, along with 0.1 g/L of substrate: Avicel PH-101 (Sigma-Aldrich, St. Louis, MO) or milled switchgrass (gift from U.S. Department of Agriculture). Avicel PH-101 substrate was used for bacterial passaging, whereas switchgrass substrate was used for fungal passaging. After autoclaving, the MC-

medium was supplemented with 0.1% (v/v) vitamin solution (ATCC, catalogue no. MD-VS) per 10 mL of final culture volume, as described by Teunissen (48). Each passaging Hungate tube was inoculated with 1.0 mL of cryostock or live inoculum culture, with cultures passaged anaerobically using sterile syrince needle technique every 2 to 4 days.

In order to inoculate the *F. succinogenes* strain UWB7 monocultures and co-cultures, seed cultures of *F. succinogenes* strain UWB7, *A. robustus*, and *C. churrovis* were prepared. The *F. succinogenes* strain UWB7 seed culture was prepared by adding 1.0 mL of passaging tube inoculum to a serum bottle containing 40 mL of MC- and 0.4 g of Avicel PH-101. The seed culture of *F. succinogenes* strain UWB7 was incubated at 39°C for 24 hrs for co-culture inoculum and incubated for a total of 48 hrs for UWB7 monoculture inoculum. *A. robustus* and *C. churrovis* seed cultures were prepared by adding 1.0 mL of appropriate passaging tube inoculum to serum bottles containing 40 mL of MC- and 0.4 g of switchgrass. The fungal seed cultures were incubated at 39°C for 24 hrs before 1.0 mL inocula were used to inoculate co-culture Hungate tubes with 8.0 mL MC- and 0.1 g switchgrass ($t = 0$ hrs) in triplicate. The gut fungi were allowed to grow at 39°C for 24 h ($t = 24$ hrs) before 1.0 mL of *F. succinogenes* strain UWB7 seed inoculum was introduced. The *A. Robustus-F. succinogenes* strain UWB7 co-cultures and *C. churrovis-F. succinogenes* strain UWB7 co-cultures were incubated for an additional 48 h ($t = 72$ hrs) before samples were harvested for RNA-seq. These experiments were performed in triplicate. *F. succinogenes* strain UWB7 monocultures were prepared by adding 1.0 mL seed culture inocula to Hungate tubes with 9.0 mL MC- and 0.1 g switchgrass. *F. succinogenes* strain UWB7 monocultures were then incubated at 39°C for 24 hrs before samples were harvested for RNA-seq. This experiment was performed in triplicate. The final contents of each co-culture and monoculture Hungate tube were transferred to 15-mL Falcon tubes (Fisher Scientific, Waltham, MA) and centrifuged at 3,200 g and 4°C using a swinging-bucket rotor (Eppendorf A-4-81) for 5 mins. Supernatants were decanted, and 1.0 mL of RNA preservative was added (49, 50). Cell pellets were flash-frozen and stored at -80°C until RNA extraction.

Trizol RNA extraction for *E. coli* and *G. metallireducens* cultures
Trizol (Thermo Fisher Scientific, Cat. # 15596018) RNA extractions were performed following the manufacturer's protocol. Briefly, 750 μL Trizol containing cell lysates were thawed and mixed with 150 μL chloroform. After centrifugation, the clear aqueous layer was recovered and precipitated with 375 μL of isopropanol and 0.67μL of GlycoBlue (Thermo Fisher Scientific, Cat. # AM9515). The pellet was washed twice with 75% ethanol and after the final centrifugation, the resulting pellet was air-dried and resuspended in nuclease-free water.

Liquid N$_2$ grinding & RNeasy RNA extraction for *F. succinogenes* strain UWB7 monoculture & co-culture

For RNA extraction, cell pellets were thawed on ice and centrifuged at 3,200 g and 4°C using a swinging-bucket rotor for 10 mins. RNA was extracted from *F. succinogenes* strain UWB7 monoculture & co-culture pellets using the RNeasy Mini Kit (Qiagen, Germantown, MD) following the protocol for purification of total RNA from plant cells and tissues and filamentous fungi, using liquid N$_2$ method of cell lysis, QIAshredder homogenization, on-column Dnase I digestion, and elution in 30 µL of RNAse-free water.

EMBR-seq+

Overview

EMBR-seq libraries were prepared according to the protocol described in Wangsanuwat *et al.* (18). 100 ng of total RNA were used as input material for all libraries. For the libraries denoted "EMBR-seq" and "EMBR-seq+", blocking primers were added during the polyadenylation step and reverse transcription step. For the libraries denoted "No depletion" and "RNase H", water was added instead of blocking primers during these two steps. In the libraries denoted "RNase H" and "EMBR-seq+", the RNase H treatment was performed as described below. In the libraries denoted "No depletion" and "EMBR-seq", the RNase H treatment was skipped; the amplified RNA (aRNA) product of *in vitro* transcription (IVT) was used directly for the library reverse transcription step as described in Wangsanuwat *et al.* (18).

Blocking primer design for G. metallireducens and F. succinogenes strain UWB7

All the primers used in this study are listed in Supplementary Table S2. The 16S, 23S, and 5S rRNA sequences were downloaded for *G. metallireducens GS-15* (NCBI database ASM1292v1) and *F. succinogenes* strain UWB7 (JGI IMG database Taxon ID 2703719236). Primers were designed according to the following specifications: (1) Reverse complement to the reference sequences at the 3'-end of each rRNA subunit; (2) Initial primer length of approximately 30 nucleotides; and (3) Initial primer length modified to ensure a T$_m$ greater than 65°C.

Initial EMBR-seq libraries were prepared with a 50 µM blocking primer mix consisting only of the three 3'-end primers (16S, 23S, and 5S). The resulting sequencing data was used to identify the locations of the most frequently sequenced rRNA fragments. Hotspot primers were designed to bind the selected locations for additional rRNA depletion. Hotspot primers were designed according to the following specifications: (1) Reverse complement to the reference sequence for each rRNA subunit; (2) Located approximately 100 nucleotides downstream (towards the 3'-end)

from the start of the hotspot sequence; (3) Initial primer length of approximately 30 nucleotides; and (4) Initial primer length modified and/or initial primer position shifted upstream or downstream by a few bases to reach a T_m greater than 65°C.

RNase H probe design

RNase H probes bind to aRNA, the IVT product, which has a sequence complementary to the original template transcriptome. Therefore, RNase H probes are designed to have the same sequence as the rRNA references. The locations of the RNase H probes were selected in the neighborhood of the 3'-end and hotspot blocking primers with a length of 50 nucleotides. The RNase H probe design can also be easily modified to target any additional abundant fragments found in prior sequencing libraries.

Preparation of blocking primer mix

Blocking primers for the species of interest (Supplementary Table S2) were individually resuspended to 100 μM. The primers were mixed such that the three 3'-end primers together were 50 μM in the final primer mix, and all the hotspot primers together were 50 μM in the final primer mix, for a total primer mix concentration of 100 μM.

Polyadenylation

2 μL of total RNA (50 ng/μL) was mixed with 1 μL of 5x first strand buffer [250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM MgCl$_2$, comes with Superscript II reverse transcriptase, Invitrogen Cat. # 18064014], 1 uL of blocking primer mix, 0.1 μL 10 mM ATP, and 0.1 μL *E. coli* poly-A polymerase (New England Biolabs, Cat. # M0276S). The samples were incubated at 37°C for 10 mins. In the "No depletion" and "RNase H" libraries, 1 μL of nuclease-free water was added instead of the blocking primer mix.

Reverse transcription

The polyadenylation product was mixed with 0.5 μL 10 mM dNTPs (New England Biolabs, Cat. # N0447L), 1 μL reverse transcription primers (25 ng/μL, Supplementary Table S2), and 1.3 μL blocking primer mix, and heated to 65°C for 5 mins, 58°C for 1 min, and then quenched on ice. In the "No depletion" and "RNase H" libraries, 1.3 μL of nuclease-free water was added instead of the blocking primer mix. This product was then mixed with 1.2 μL 5x first strand buffer, 1 μL 0.1 M DTT, 0.5 μL RNaseOUT (Thermo Fisher Scientific, Cat. #10777019), and 0.5 μL Superscript II

reverse transcriptase, and then incubated at 42°C for 1 hr. Immediately afterwards, the samples were heat-inactivated at 70°C for 10 mins.

Second strand synthesis

The reverse transcription product was mixed with 33.5 µL nuclease-free water, 12 µL 5x second strand buffer [100 mM Tris-HCl (pH 6.9), 23 mM MgCl$_2$, 450 mM KCl, 0.75 mM β-NAD, 50 mM (NH$_4$)$_2$SO$_4$, Invitrogen, Cat. # 10812014], 1.2 µL 10 mM dNTPs, 0.4 µL *E. coli* ligase (Invitrogen, Cat. # 18052019), 1.5 µL DNA polymerase I (Invitrogen, Cat. # 18010025), and 0.4 µL RNase H (Invitrogen, Cat. # 18021071), and incubated at 16°C for 2 hrs. The cDNA was purified with 1x AMPure XP DNA beads (Beckman Coulter, Cat. # A63881) and eluted in 24 µL nuclease-free water that was subsequently concentrated to 6.4 µL.

In vitro transcription

The product from the previous step was mixed with 9.6 µL of *in vitro* transcription mix (1.6 µL of each ribonucleotide, 1.6 µL 10x T7 reaction buffer, 1.6 µL T7 enzyme mix, MEGAscript T7 Transcription Kit, Thermo Fisher Scientific, Cat. # AMB13345) and incubated at 37°C for 13 hrs. The IVT product was mixed with 6 µL ExoSAP-IT™ PCR Product Cleanup Reagent (Thermo Fisher Scientific, Cat. # 78200.200.UL) and incubated at 37°C for 15 mins. Next, it was treated with 5.5 µL fragmentation buffer (200 mM Tris-acetate (pH 8.1), 500 mM KOAc, 150 mM MgOAc) at 94°C for 3 mins and immediately quenched with 2.75 µL stop buffer (0.5 M EDTA) on ice. The fragmented aRNA was size-selected with 0.8x AMPure RNA beads (RNAClean XP Kit, Beckman Coulter, Cat. # A63987) and eluted in 18 µL nuclease-free water. In the libraries denoted "No depletion" and "EMBR-seq", the fragmented and cleaned-up *in vitro* transcription product was used directly for Illumina library preparation by reverse transcription and PCR, as described previously (18, 43). In the "RNase H" and "EMBR-seq+" libraries, the RNase H treatment was performed before proceeding to Illumina library preparation.

RNase H probe mix and buffer preparation

RNase H probes (Supplementary Table S2) were individually resuspended to 100 µM, mixed equimolarly, and the mixture was diluted to 10 µM (~155 ng/µL). RNase H buffer mix was prepared by mixing 10 µL of 5x first strand buffer with 2.7 µL of 5x second strand buffer or scaling up as needed.

RNase H treatment for monocultures

After ExoSAP treatment, RNA fragmentation, and bead-based size selection, the aRNA concentration was measured on a NanoDrop One spectrophotometer. 1000 ng of aRNA was mixed with 500 ng (3.2 µL) of RNase H probe mix, 3.2 µL of RNase H buffer mix, and nuclease-free water to reach a total volume of 15 µL for each reaction. This mixture was preheated at 65°C for 5 mins and quenched on ice. 1 µL of RNase H (Invitrogen, Cat. # 18021071) was added and the reaction was incubated at 16°C for 30 mins. For tests with the thermostable RNase H, 1 µL of Hybridase Thermostable RNase H at 45°C (Lucigen (now Biosearch Technologies), Cat. # H39500) was added and the reaction was incubated at 45°C for 30 mins. Finally, the RNase H product was size-selected with 1x AMPure RNA beads and eluted in 15 µL nuclease-free water, and then concentrated to 5 µL volume for Illumina library preparation.

Enrichment of *F. succinogenes* strain UWB7 RNA from co-cultures

Fungal mRNA removal

Total RNA extracted from co-culture pellets was concentrated to 30 µL (40-60 ng/µL) and treated with Dynabeads™ Oligo(dT)$_{25}$ (Invitrogen Cat # 61002), according to the manufacturer's protocol. Briefly, each sample was mixed with 30 µL of "Binding Buffer", incubated at 65°C for 2 mins, chilled on ice, mixed with 30 µL of washed oligo(dT)$_{25}$ beads, and mixed at 300 rpm on a thermomixer at room temperature for 15 mins. To prevent the beads from settling in the thermomixer, the samples were mixed by pipetting every 5 mins. The reaction tubes were placed on a magnetic stand and the supernatants (90-95 µL) were transferred to new tubes. The supernatants were purified with 1.8x AMPure RNA beads and eluted in 30 µL of nuclease-free water. This remaining RNA was used for preparing the "Poly-A depleted" EMBR-seq libraries.

Determination of fungal rRNA sequences for RNase H depletion

To determine potential rRNA sequences for *C. churrovis*, FastQC was used to identify the most overrepresented sequences in the "Poly-A depleted" EMBR-seq sample, and the top hits were aligned with BLAST against the full *C. churrovis* genome. The BLAST results suggested that the most overrepresented reads were aligned on scaffold 672.

Next, all the RNA-seq reads from the "Poly-A depleted" EMBR-seq libraries were mapped against scaffold 672 of the *C. churrovis* genome with the Burrows-Wheeler Aligner software package (BWA) (51). This scaffold contained four repeats of three hotspots which appeared to correspond to rRNA subunits, based on the spacing between hotspots (Supplementary Figure S5). The four repeats were observed because non-unique mapping was allowed. To produce

Figure 3C, the mapping was repeated on a version of scaffold 672 that is truncated to only the first 10,000 bases (1 repeat of the predicted rRNA pattern). RNase H probes were designed according to the locations of the predicted hotspots for the 3'-ends of the 28S, 18S, and 5.8S rRNA subunits in *C. churrovis*. For RNase H treatment in co-culture EMBR-seq+ libraries, the *C. churrovis* probes were applied to co-cultures with *A. robustus* as well. Although they are different fungal species, we verified that the same RNase H target sequences are highly abundant in both *A. robustus* and *C. churrovis* genomes, and the 3'-ends of their rRNA are similar enough such that fungal rRNA depletion was efficient in both co-cultures (Supplementary Figures S4 and S5).

RNase H reaction for co-cultures

For co-culture samples, the RNase H probe mix was altered to include probes for depleting fungal rRNA along with bacterial rRNA. 1000 ng of aRNA derived from co-culture samples was mixed with 850 ng (5.5 µL) of co-culture RNase H probe mix, which consisted of: 100 ng each of the two fungal 28S probes with 50 ng each of the two fungal 18S, fungal 5.8S, and ten bacterial RNase H probes. The aRNA and probes were mixed with 3.2 µL RNase H buffer mix and nuclease-free water to reach a total volume of 15 µL for each reaction. The reactions were preheated at 65°C for 5 mins and quenched on ice. 1 µL of RNase H was added and the reaction was incubated at 16°C for 30 mins. The RNase H product was size-selected with 1x AMPure RNA beads and eluted in 15 µL nuclease-free water, and then concentrated to 5 µL for Illumina library preparation.

RNA sequencing and bioinformatics analysis

Paired-end sequencing of all libraries was performed on an Illumina NextSeq 500 or HiSeq 4000 platform. All sequencing data has been deposited to Gene Expression Omnibus under the accession number GSE223404. In the sequencing libraries, read 1 contains the sample barcode (Supplementary Table S2). Read 2 is mapped to the bacterial or fungal transcriptome. Prior to mapping, read 1 was trimmed to 25 bases, read 2 was trimmed to 51 bases, and only reads containing valid sample barcodes on read 1 were retained. Next, the reads were mapped to the appropriate reference transcriptome using Burrows-Wheeler Aligner (BWA) version 0.7.15 with default parameters. The "CDS from genomic FASTA" reference files for *Escherichia coli K12* substr. *MG1655* (ASM584v2 cds) and *Geobacter metallireducens GS-15 (*ASM1292v1 cds) were obtained from the NCBI database. The reference transcriptome for *Fibrobacter succinogenes* strain UWB7 (Taxon ID 2703719236) was obtained from the JGI IMG database (52). The reference genomes for *Anaeromyces robustus* and *Caecomyces churrovis* were obtained from the JGI MycoCosm portal (53)). Each bacterial reference transcriptome was modified by

appending the 16S, 23S, and 5S rRNA sequences. To determine rRNA content for each sample, the number of reads mapped to the rRNA was taken as a fraction of total reads mapped to the bacteria.

To compare the results of this work to Swift *et al.*, raw sequencing reads (fastq files) were retrieved from the Sequencing Read Archive (SRA) for NCBI BioProject PRJNA666900 as specified in Swift *et al.* (26). Reads 1 and 2 were trimmed to 51 bases and mapped to the reference transcriptomes for *F. succinogenes* strain UWB7, *A. robustus* and *C. churrovis* using the same BWA pipeline described above.

CAZyme annotations were predicted for the *F. succinogenes* strain UWB7 transcriptome using dbCAN2 (31). Three tools were run: (1) HMMER: dbCAN, (2) DIAMOND: CAZy, and (3) HMMR: dbCAN_sub, and CAZyme annotations were accepted if they were predicted with two out of the three tools.

Differential gene expression analysis was performed using DEseq2 version 1.36.0 (54), with a minimum absolute $\log_2$(Fold Change) of 0.8, statistical significance threshold for adjusted P value of less than 0.05, and a minimum gene expression level of 2 reads per million.

To account for biases introduced by EMBR-seq+ in the *F. succinogenes* strain UWB7 monocultures (Figure 2D), genes determined to be differentially expressed between the "No depletion" and "EMBR-seq" dataset were filtered out and excluded from all analyses between monoculture and co-culture (Figures 3E,F, Supplementary Figures S7-S9, Supplementary Tables S4-S7). DE-Seq2 thresholds for the bias determination were $|\log_2(Fold\ Change)| > 1.5$, $P_{adj} < 0.05$, and RPM > 2. Filtering results are included in Supplementary Table S4.

DATA AVAILABILITY

RNA-seq data: Gene Expression Omnibus GSE223404
(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE223404).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

K.A.H, C.W., and S.S.D. conceived the method and designed the experiments. K.A.H., C.W. and S.C.T. participated in data collection for EMBR-seq+. L.V.B. cultivated and performed RNA extractions for all monocultures and co-cultures of *F. succinogenes* strain UWB7, *A. robustus*, and *C. churrovis*. A.R.R. cultivated *G. metallireducens*. K.A.H. and C.W. analyzed the data. K.A.H., L.V.B., A.R.R., M.A.O., and S.S.D. wrote the manuscript. S.S.D guided experimental design and data analysis. All authors have read and approved the final version of the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

1. Hess M, Sczyrba A, Egan R, Kim T-W, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM. 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331:463–467.

2. Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, Eloe-Fadrosh EA, Pavlopoulos GA, Hadjithomas M, Varghese NJ, Paez-Espino D, Perry R, Henderson G, Creevey CJ, Terrapon N, Lapebie P, Drula E, Lombard V, Rubin E, Kyrpides NC, Henrissat B, Woyke T, Ivanova NN, Kelly WJ. 2018. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. Nat Biotechnol 36:359–367.

3. Peng X "Nick," Gilmore SP, O'Malley MA. 2016. Microbial communities for bioprocessing: lessons learned from nature. Curr Opin Chem Eng 14:103–109.

4. Lillington SP, Leggieri PA, Heom KA, O'Malley MA. 2020. Nature's recyclers: anaerobic microbial communities drive crude biomass deconstruction. Curr Opin Biotechnol 62:38–47.

5. Burnet MC, Dohnalkova AC, Neumann AP, Lipton MS, Smith RD, Suen G, Callister SJ. 2015. Evaluating nodels of cellulose degradation by Fibrobacter succinogenes S85. PLoS One 10:e0143809.

6. Stewart RD, Auffret MD, Warr A, Walker AW, Roehe R, Watson M. 2019. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. Nat Biotechnol 37:953–961.

7. Venturelli OS, Carr AV, Fisher G, Hsu RH, Lau R, Bowen BP, Hromada S, Northen T, Arkin AP. 2018. Deciphering microbial interactions in synthetic human gut microbiome communities. Mol Syst Biol 14:e8157.

8. Schink B. 1997. Energetics of syntrophic cooperation in methanogenic degradation. Microbiol Mol Biol Rev 61:262–280.

9. Giannoukos G, Ciulla DM, Huang K, Haas BJ, Izard J, Levin JZ, Livny J, Earl AM, Gevers D, Ward DV, Nusbaum C, Birren BW, Gnirke A. 2012. Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. Genome Biol 13:r23.

10. Kraus AJ, Brink BG, Siegel TN. 2019. Efficient and specific oligo-based depletion of rRNA. Sci Rep 9:12281.

11. Culviner PH, Guegler CK, Laub MT. 2020. A simple, cost-effective, and robust method for rRNA depletion in RNA-sequencing studies. mBio 11:e00010-20.

12. Huang Y, Sheth RU, Kaufman A, Wang HH. 2020. Scalable and cost-effective ribonuclease-based rRNA depletion for transcriptomics. Nucleic Acids Res 48:e20.

13. Engelhardt F, Tomasch J, Häussler S. 2020. Organism-specific depletion of highly abundant RNA species from bacterial total RNA. Access Microbiol 2:acmi000159.

14. Choe D, Szubin R, Poudel S, Sastry A, Song Y, Lee Y, Cho S, Palsson B, Cho B-K. 2021. RiboRid: A low cost, advanced, and ultra-efficient method to remove ribosomal RNA for bacterial transcriptomics. PLoS Genet 17:e1009821.

15. Prezza G, Heckel T, Dietrich S, Homberger C, Westermann AJ, Vogel J. 2020. Improved bacterial RNA-seq by Cas9-based depletion of ribosomal RNA reads. RNA 26:1069–1078.

16. Kang Y, Norris MH, Zarzycki-Siek J, Nierman WC, Donachie SP, Hoang TT. 2011. Transcript amplification from single bacterium for transcriptome analysis. Genome Res 21:925–935.

17. He S, Wurtzel O, Singh K, Froula JL, Yilmaz S, Tringe SG, Wang Z, Chen F, Lindquist EA, Sorek R, Hugenholtz P. 2010. Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. Nat Methods 7:807–812.

18. Wangsanuwat C, Heom KA, Liu E, O'Malley MA, Dey SS. 2020. Efficient and cost-effective bacterial mRNA sequencing from low input samples through ribosomal RNA depletion. BMC Genomics 21:717.

19. Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, Deneke J, Drinkwater C, Ivanova NN, Mikhailova N, Chertkov O, Goodwin LA, Currie CR, Mead D, Brumm PJ. 2011. The complete genome sequence of Fibrobacter succinogenes S85 reveals a cellulolytic and metabolic specialist. PLoS One 6:e18814.

20. Childers SE, Ciufo S, Lovley DR. 2002. Geobacter metallireducens accesses insoluble Fe(iii) oxide by chemotaxis. Nature 416:767–769.

21. Rotaru A-E, Shrestha PM, Liu F, Markovaite B, Chen S, Nevin KP, Lovley DR. 2014. Direct interspecies electron transfer between Geobacter metallireducens and Methanosarcina barkeri. Appl Environ Microbiol 80:4599–4605.

22. Solomon KV, Haitjema CH, Henske JK, Gilmore SP, Borges-Rivera D, Lipzen A, Brewer HM, Purvine SO, Wright AT, Theodorou MK, Grigoriev IV, Regev A, Thompson DA, O'Malley MA. 2016. Early-branching gut fungi possess a large, comprehensive array of biomass-degrading enzymes. Science 351:1192–1195.

23. Haitjema CH, Gilmore SP, Henske JK, Solomon KV, de Groot R, Kuo A, Mondo SJ, Salamov AA, LaButti K, Zhao Z, Chiniquy J, Barry K, Brewer HM, Purvine SO, Wright AT, Hainaut M, Boxma B, van Alen T, Hackstein JHP, Henrissat B, Baker SE, Grigoriev IV, O'Malley MA. 2017. A parts list for fungal cellulosomes revealed by comparative genomics. Nat Microbiol 2:17087.

24. Gilmore SP, Lankiewicz TS, Wilken StE, Brown JL, Sexton JA, Henske JK, Theodorou MK, Valentine DL, O'Malley MA. 2019. Top-down enrichment guides in formation of synthetic microbial consortia for biomass degradation. ACS Synth Biol 8:2174–2185.

25. Swift CL, Brown JL, Seppälä S, O'Malley MA. 2019. Co-cultivation of the anaerobic fungus Anaeromyces robustus with Methanobacterium bryantii enhances transcription of carbohydrate active enzymes. J Ind Microbiol Biotechnol 46:1427–1433.

26. Swift CL, Louie KB, Bowen BP, Hooker CA, Solomon KV, Singan V, Daum C, Pennacchio CP, Barry K, Shutthanandan V, Evans JE, Grigoriev IV, Northen TR, O'Malley MA. 2021. Cocultivation of anaerobic fungi with rumen bacteria establishes an antagonistic relationship. mBio 12:e01442-21.

27. Henske JK, Gilmore SP, Knop D, Cunningham FJ, Sexton JA, Smallwood CR, Shutthanandan V, Evans JE, Theodorou MK, O'Malley MA. 2017. Transcriptomic characterization of Caecomyces churrovis: a novel, non-rhizoid-forming lignocellulolytic anaerobic fungus. Biotechnol Biofuels 10:305.

28. Brown JL, Swift CL, Mondo SJ, Seppälä S, Salamov A, Singan V, Henrissat B, Drula E, Henske JK, Lee S, LaButti K, He G, Yan M, Barry K, Grigoriev IV, O'Malley MA. 2021. Co-cultivation of the anaerobic fungus Caecomyces churrovis with Methanobacterium

bryantii enhances transcription of carbohydrate binding modules, dockerins, and pyruvate formate lyases on specific substrates. Biotechnol Biofuels 14:234.

29. Nicholson MJ, Theodorou MK, Brookman JL. 2005. Molecular analysis of the anaerobic rumen fungus Orpinomyces – insights into an AT-rich genome. Microbiology 151:121–133.

30. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445–W451.

31. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101.

32. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. 2022. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res 50:D571–D577.

33. Henske JK, Wilken StE, Solomon KV, Smallwood CR, Shutthanandan V, Evans JE, Theodorou MK, O'Malley MA. 2018. Metabolic characterization of anaerobic fungi provides a path forward for bioprocessing of crude lignocellulose. Biotechnol Bioeng 115:874–884.

34. Scarborough MJ, Lawson CE, Hamilton JJ, Donohue TJ, Noguera DR. 2018. Metatranscriptomic and thermodynamic insights into medium-chain fatty acid production using an anaerobic microbiome. mSystems 3:e00221-18, /msystems/3/6/msys.00221-18.atom.

35. Lawson CE, Harcombe WR, Hatzenpichler R, Lindemann SR, Löffler FE, O'Malley MA, García Martín H, Pfleger BF, Raskin L, Venturelli OS, Weissbrodt DG, Noguera DR, McMahon KD. 2019. Common principles and best practices for engineering microbiomes. Nat Rev Microbiol 17:725–741.

36. Brown JL, Perisin MA, Swift CL, Benyamin M, Liu S, Singan V, Zhang Y, Savage E, Pennacchio C, Grigoriev IV, O'Malley MA. 2022. Co-cultivation of anaerobic fungi with *Clostridium acetobutylicum* bolsters butyrate and butanol production from cellulose and lignocellulose. J Ind Microbiol Biotechnol 49.

37. Peng X, Wilken StE, Lankiewicz TS, Gilmore SP, Brown JL, Henske JK, Swift CL, Salamov A, Barry K, Grigoriev IV, Theodorou MK, Valentine DL, O'Malley MA. 2021. Genomic and functional analyses of fungal and bacterial consortia that enable lignocellulose breakdown in goat gut microbiomes. Nat Microbiol 6:499–511.

38. Leggieri PA, Valentine MT, O'Malley MA. 2022. Biofilm disruption enhances growth rate and carbohydrate-active enzyme production in anaerobic fungi. Bioresour Technol 358:127361.

39. Blattman SB, Jiang W, Oikonomou P, Tavazoie S. 2020. Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. Nat Microbiol 5:1192-1201.

40. Kuchina A, Brettner LM, Paleologu L, Roco CM, Rosenberg AB, Carignano A, Kibler R, Hirano M, DePaolo RW, Seelig G. 2020. Microbial single-cell RNA sequencing by split-pool barcoding. Science 371.

41. Imdahl F, Vafadarnejad E, Homberger C, Saliba A-E, Vogel J. 2020. Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria. Nat Microbiol https://doi.org/10.1038/s41564-020-0774-1.

42. Hashimshony T, Wagner F, Sher N, Yanai I. 2012. CEL-Seq: Single-cell RNA-seq by multiplexed linear amplification. Cell Rep 2:666–673.

43. Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, Anavy L, Gennert D, Li S, Livak KJ, Rozenblatt-Rosen O, Dor Y, Regev A, Yanai I. 2016. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol 17:77.

44. Bretschger O, Obraztsova A, Sturm CA, Chang IS, Gorby YA, Reed SB, Culley DE, Reardon CL, Barua S, Romine MF, Zhou J, Beliaev AS, Bouhenni R, Saffarini D, Mansfeld F, Kim B-H, Fredrickson JK, Nealson KH. 2007. Current production and metal oxide reduction by *Shewanella oneidensis* MR-1 wild type and mutants. Appl Environ Microbiol 73:7003–7012.

45. Neumann AP, Suen G. 2018. The phylogenomic diversity of herbivore-associated *Fibrobacter* spp. is correlated to lignocellulose-degrading potential. mSphere 3:e00593-18.

46. Haitjema CH, Solomon KV, Henske JK, Theodorou MK, O'Malley MA. 2014. Anaerobic gut fungi: Advances in isolation, culture, and cellulolytic enzyme discovery for biofuel production. Biotechnol Bioeng 111:1471–1482.

47. Davies DR, Theodorou MK, Lawrence MIG, Trinci APJ. 1993. Distribution of anaerobic fungi in the digestive tract of cattle and their survival in faeces. J Gen Microbiol 139:1395–1400.

48. Teunissen MJ, Op den Camp HJM, Orpin CG, Huis in 't Veld JHJ, Vogels GD. 1991. Comparison of growth characteristics of anaerobic fungi isolated from ruminant and non-ruminant herbivores during cultivation in a defined medium. J Gen Microbiol 137:1401–1408.

49. Erster O, Shkedi O, Benedek G, Zilber E, Varkovitzky I, Shirazi R, Oriya Shorka D, Cohen Y, Bar T, Yechieli R, Tepperberg Oikawa M, Venkert D, Linial M, Oiknine-Djian E, Mandelboim M, Livneh Z, Shenhav-Saltzman G, Mendelson E, Wolf D, Szwarcwort-Cohen M, Mor O, Lewis Y, Zeevi D. 2021. Improved sensitivity, safety, and rapidity of COVID-19 tests by replacing viral storage solution with lysis buffer. PLoS One 16:e0249149.

50. Malmstrom R. 2015. RNAlater recipe v1. preprint.

51. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

52. Chen I-MA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, Huntemann M, Varghese N, White JR, Seshadri R, Smirnova T, Kirton E, Jungbluth SP, Woyke T, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC. 2019. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Res 47:D666–D677.

53. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res 42:D699–D704.

54. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.
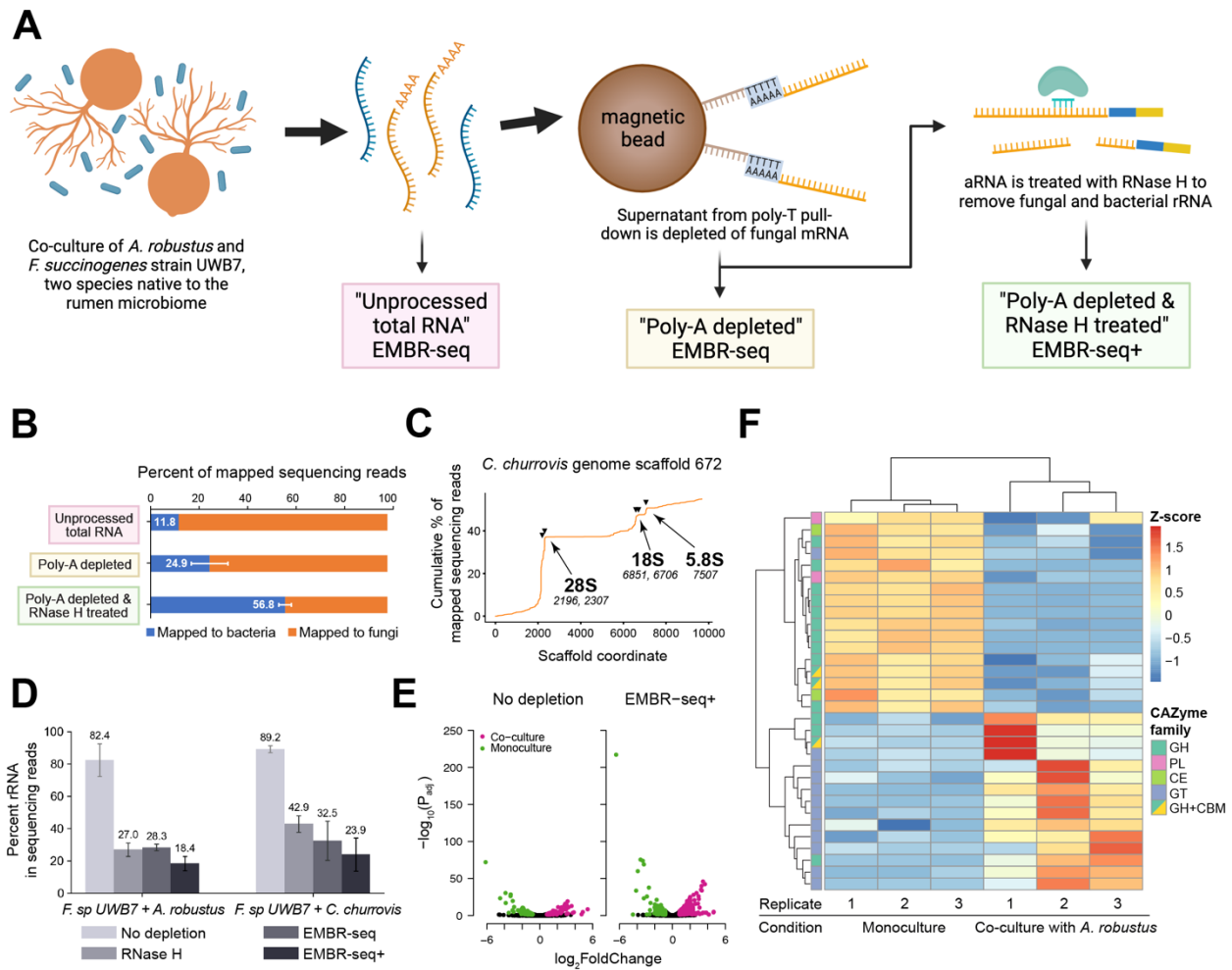
FIGURES
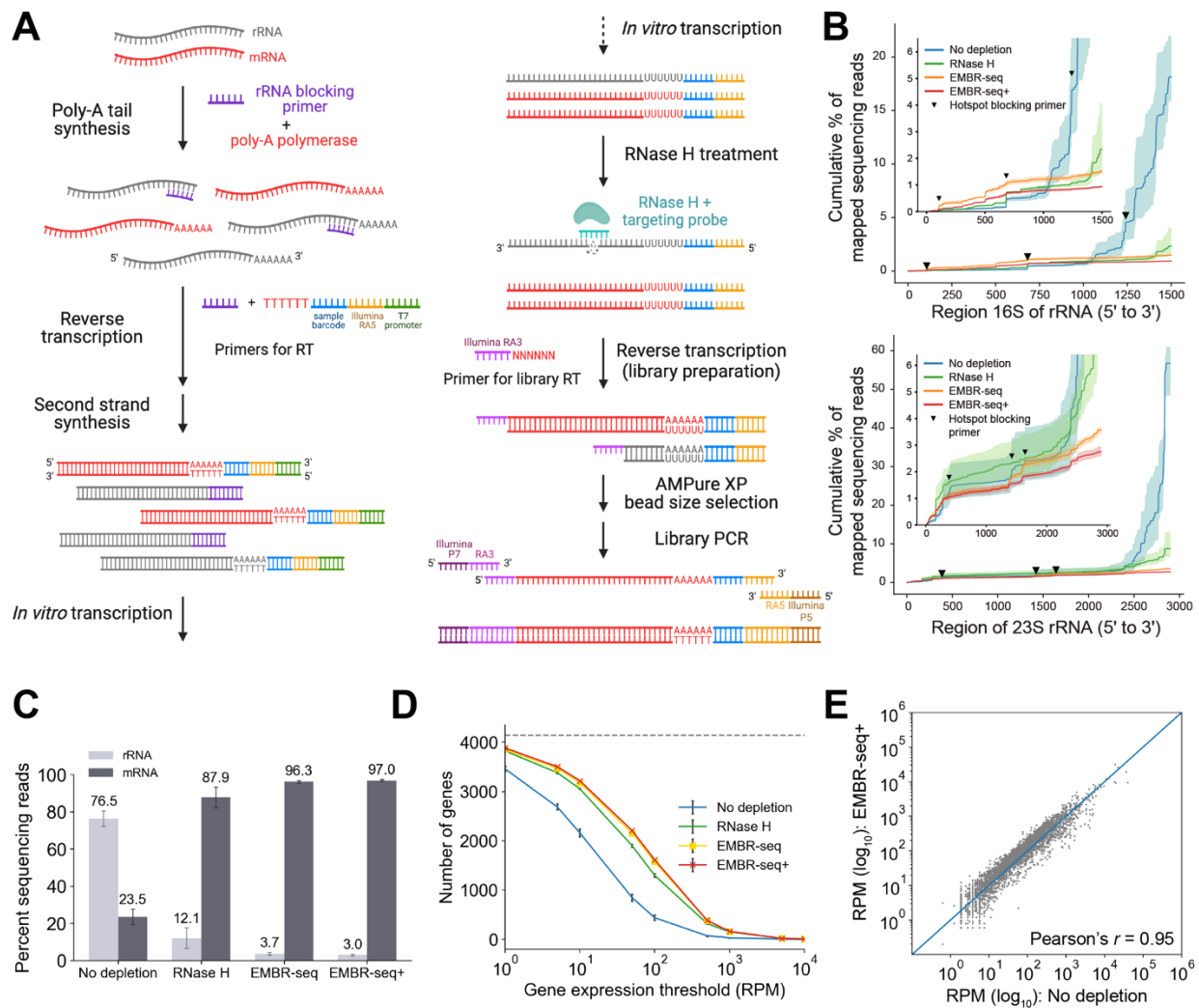
Figure 1. EMBR-seq+ depletes rRNA from *E. coli* total RNA at high efficiency without introducing technical biases using a small number of rRNA-specific probes. (**A**) Schematic for EMBR-seq+ shows that a combination of blocking primers and RNase H-based targeting is used to deplete rRNA. (**B**) Cumulative percentage of reads mapping along *E. coli* 16S and 23S rRNA for different depletion methods. Bold lines indicate mean values and shaded regions indicate the minimum and maximum over three independent experiments. Inverted triangles indicate location of hotspot regions targeted by blocking primers. Insets highlight the depletion of reads from hotspot locations along the 16S and 23S rRNA sequence. (**C**) Percent of sequencing reads mapping to rRNA and mRNA in *E. coli*. Bars indicate mean values and error bars indicate standard deviations over three independent experiments. (**D**) Number of genes detected above different gene expression thresholds for *E. coli*. Points indicate mean values and error bars indicate standard deviations over three independent experiments. (**E**) Correlation of gene

expression between the "No depletion" and "EMBR-seq+" conditions (Pearson's *r* = 0.95). RPM is computed after removal of rRNA reads. The x- and y-coordinates of each point indicate mean values over three independent experiments in the two conditions. For consistent comparison across methods, panels b and d show data that has been downsampled to 1 million sequencing reads.
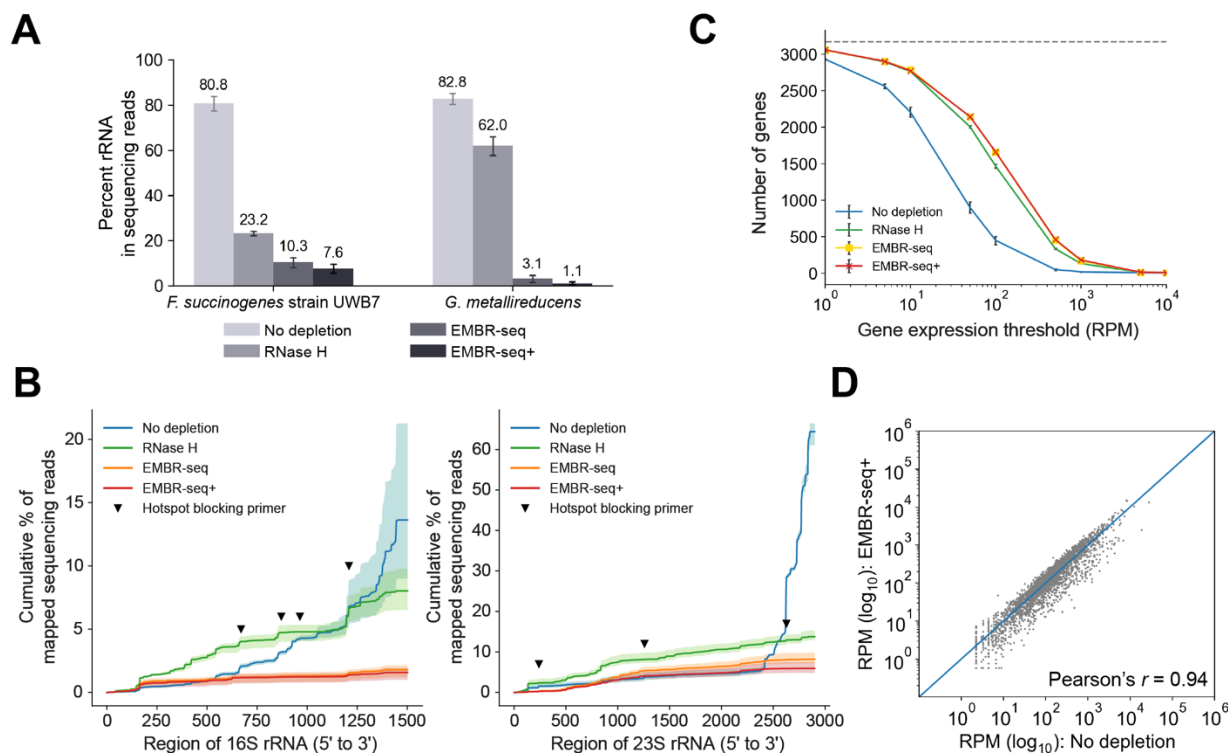
Figure 2. EMBR-seq+ can be easily extended to diverse bacterial species. (**A**) Percent of sequencing reads mapping to rRNA in *F. succinogenes* strain UWB7 and *G. metallireducens* for different rRNA depletion methods. Bars indicate mean values and error bars indicate standard deviations over three independent experiments. (**B**) Cumulative percentage of reads mapping to 16S and 23S rRNA of *F. succinogenes* strain UWB7 for different rRNA depletion methods. Bold lines indicate mean values and shaded regions indicate the minimum and maximum over three independent experiments. Inverted triangles indicate the location of hotspots targeted by blocking primers. (**C**) Number of genes detected above different gene expression thresholds for *F. succinogenes* strain UWB7. Points indicate mean values and error bars indicate standard deviations over three independent experiments. (**D**) Correlation of gene expression between the "No depletion" and "EMBR-seq+" conditions for *F. succinogenes* strain UWB7 (Pearson's *r* = 0.94). RPM is computed after removal of rRNA reads. The x- and y-coordinates of each point indicate mean values over three independent experiments in the two conditions. For consistent comparison across methods, panels b and c show data that has been downsampled to 1 million sequencing reads.

Figure 3. EMBR-seq+ efficiently depletes rRNA of *F. succinogenes* strain UWB7 grown in co-culture with *A. robustus*. (**A**) Schematic of fungal mRNA and rRNA removal from co-culture-derived total RNA samples for more efficient bacterial mRNA sequencing. (**B**) Percentage of reads mapping to the bacterial and fungal genomes using the three strategies described in panel A: total RNA isolated from co-culture pellets treated with EMBR-seq to remove bacterial rRNA ("Unprocessed total RNA"), fungal poly-adenylated mRNA depleted total RNA treated with EMBR-seq to remove bacterial rRNA ("Poly-A depleted"), and "Poly-A depleted" EMBR-seq library additionally treated with RNase H to remove fungal and bacterial rRNA ("Poly-A depleted & RNase H treated"). For the "Unprocessed total RNA" library, $n = 1$. For "Poly-A depleted" and "Poly-A depleted & RNase H treated" libraries, $n = 3$. Bars indicate mean values and error bars indicate standard deviation from the mean. (**C**) Cumulative percentage of sequencing reads from the "Poly-A depleted" EMBR-seq library that align to the *C. churrovis* genome scaffold 672. The three rRNA hotspots were inferred to be the 3'-ends of 18S, 28S, and 5.8S rRNA for *C. churrovis*. Black inverted triangles and italicized text indicate target locations of the five RNase H probes

used for fungal rRNA depletion in the EMBR-seq+ libraries. (**D**) Percent of sequencing reads mapping to the rRNA of *F. succinogenes* strain UWB7 for different depletion methods when grown in co-culture with *A. robustus* or *C. churrovis*. All four samples are poly-A depleted to remove fungal mRNA. Bars indicate mean values and error bars indicate standard deviation over three independent experiments. (**E**) Volcano plot of differentially expressed genes of *F. succinogenes* strain UWB7 grown in monoculture *vs.* co-culture with *A. robustus*, for the "No depletion" condition and after rRNA depletion using EMBR-seq+. *n* = 3 for both culture conditions and for both the library preparation conditions. Colored points indicate differentially expressed genes, determined by thresholds of $|\log_2(Fold\ Change)| > 0.8$, $P_{adj} < 0.05$, and RPM > 2. (**F**) Heatmap of differentially expressed CAZymes in *F. succinogenes* strain UWB7 grown in monoculture *vs.* co-culture with *A. robustus*. Genes are colored on the left side of the heatmap based on the CAZyme family they are associated with.