

Ribosomal RNA-depletion provides an efficient method for successful dual RNA-Seq expression profiling of a marine sponge holobiont

Xueyan Xiang

University of Queensland

Davide Poli

University of Queensland

Bernard M. Degnan

University of Queensland

Sandie M. Degnan (✉ s.degnan@uq.edu.edu)

University of Queensland

Research Article

Keywords: Demosponge, Holobiont, Hologenome, Porifera, Transcriptomics

Posted Date: April 19th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1544807/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Investigations of host-symbiont interactions can benefit enormously from a complete and reliable holobiont gene expression profiling. The most efficient way to acquire holobiont transcriptomes is to perform RNA-Seq on both host and symbionts simultaneously. However, optimal methods for capturing both host and symbiont mRNAs are still under development, particularly when the host is a eukaryote and the symbionts are bacteria or archaea. Traditionally, poly(A)-enriched libraries have been used to capture eukaryotic mRNA, but the ability of this method to adequately capture bacterial mRNAs is unclear because of the short half-life of the bacterial transcripts. To determine the optimal RNA-Seq approach for animal host–bacterial symbiont analysis, we compared transcriptome bias, depth and coverage achieved by two different mRNA capture and sequencing strategies applied to the marine demosponge *Amphimedon queenslandica* holobiont, for which genomes of the animal host and three most abundant bacterial symbionts are available. The two strategies were: (i) poly(A) captured mRNA-Seq (Poly(A)-RNA-Seq); and (ii) ribosomal RNA depleted RNA-Seq (rRNA-depleted-RNA-Seq). For the host sponge, we find no significant difference in transcriptomes generated by the two different mRNA capture methods. However, for the symbiont transcriptomes, we confirm the expectation that the rRNA-depleted-RNA-Seq performs much better than the Poly(A)-RNA-Seq. Importantly, the availability of all genomes allowed us to accurately assess the nature of the Poly(A) bias compared to the rRNA-depleted transcriptomes. This comparison demonstrates that RNA-Seq by ribosomal RNA depletion is an effective and reliable method to analyse holobiont transcriptomes.

Introduction

Increasing interest in the microbes that live in association with multicellular organisms has revealed that symbioses are ubiquitous, and that plants and animals are best considered as holobionts comprising both a multicellular host and diverse microbiota (Bordenstein and Theis 2015; Rosenberg and Zilber-Rosenberg 2016). Host-associated microbes can have a critical influence on host metabolism, immunity, physiology, development and fitness (Engelstadter and Hurst 2009; Ezenwa et al. 2012; Leulier et al. 2017; McFall-Ngai et al. 2013; Pradeu 2011; Rosenberg and Zilber-Rosenberg 2016; Yen and Barr 1971). To investigate these influences, the mechanisms of host-symbiont interactions are increasingly being explored by multi-omics approaches (Heinken and Thiele 2015; Rowland et al. 2017).

With recent significant reductions in sequencing costs, RNA-sequencing (RNA-Seq) has become a widely used method to study host-symbiont interactions, and is especially valuable when used to profile host and symbiont transcriptomes simultaneously in the same biological samples. In the past decade, several studies have applied dual transcriptome analysis to reveal potential modes of cross-talk between symbiosis partners. For instance, early transcriptome analysis of the whitefly (*Bemisia tabaci*) complex revealed complementation in amino acid biosynthesis pathways between the whitefly and its bacterial endosymbiont *Candidatus Portiera aleyrodidarum* (Upadhyay et al. 2015). Dual RNA-Seq across the life cycle of the nematode *Brugia malayi*, and its bacterial endosymbiont *Wolbachia* identified developmental pathways involving both nematode and bacterial genes (Grote et al. 2017). Metatranscriptome analysis

contributed to the reconstruction of metabolic pathways of the sponge *Cymbastela concentrica* and its symbiotic microbes (Moitinho-Silva et al. 2017) and coral reef demosponges more broadly (Robbins et al. 2021).

However, the accurate and simultaneous analysis of the transcriptional state of both animal and bacterial partners in a symbiosis currently is not optimised for many holobiont systems. Transcriptomes generated via RNA-Seq tend to be dominated by highly abundant ribosomal (r)RNAs, making it necessary first to remove rRNA and enrich for coding sequence mRNAs. This need for rRNA removal creates challenges for studies wishing to capture both eukaryote and bacterial mRNAs, because of differences in eukaryote and bacterial transcript processing, lifespan and decay. In the eukaryote animal hosts, mRNA transcripts generally have a half-life of many hours (Sharova et al. 2009; Yang et al. 2003) and are stabilised by the addition of a long tail of adenines (poly(A) tails; ~250 long) at the 3' end of the transcript (Dendooven et al. 2020; Perez-Ortin et al. 2013; Westermann et al. 2012). These relatively stable mRNAs with long poly(A) tails can be easily captured and enriched away from rRNAs by the use of oligo (dT) primers, to generate libraries for sequencing; this is standard practice for eukaryote mRNA-Seq. In contrast, bacterial symbiont mRNAs generally have a much shorter half-life of only a few minutes on average (Selinger et al. 2003), and are transiently polyadenylated with short poly(A) tails (< 50 As) (Dendooven et al. 2020; Westermann et al. 2012) that tag the transcript for degradation, rather than for stabilisation (Dreyfus and Regnier 2002). Thus, the most common choice for rRNA depletion in bacterial studies is to use a subtractive hybridisation method via commercial kits (see Petrova et al. (2017) for comparison of different kits) before RNA-Seq analyses.

These features suggest that holobiont transcriptome analyses may require different methods to capture mRNAs from eukaryote and bacterial partners before RNA-Seq. In studies of marine sponges, well recognised as exemplar animal-bacterial holobionts (Hentschel et al. 2012; O'Brien et al. 2020; Steinert et al. 2020; Webster and Thomas 2016), numerous strategies have been used to try and capture transcriptomes from both host and symbionts. For example, to reconstruct metabolic networks linking the host sponge *Cymbastela concentrica*, a diatom and three proteobacterial symbionts, Moitinho-Silva et al. (2017) combined a eukaryote poly(A) capture with a separate bacterial rRNA depletion method. The need to treat eukaryote and bacterial RNA separately is labour- and cost-intensive. Thus, a single method that obtains accurate mRNA representation from host and microbe is preferred. To this end, both sponge host and bacterial symbiont transcriptomes were obtained from the giant barrel sponge *Xestospongia muta* holobiont by applying only a eukaryote rRNA depletion step (Fiore et al. 2015); it appears in that study that bacterial rRNAs were retained, but there is no analysis reported as to whether these interfered with the depth of mRNA reads acquired. In contrast, transcriptomes of both the sponge *Vaceletia* sp. and its bacterial symbionts appear to have been assessed by using only poly(A) capture method (Germer et al. 2017). Based on the differences in eukaryotic and bacterial mRNA stability and processing described above, an oligo (dT) enrichment method is likely to capture only a portion of the bacterial mRNAs, but the consequences of this for an accurate representation of symbiotic partners was not explored in Germer et al. (2017).

Here we provide, to the best of our knowledge for the first time, an empirical comparison of the completeness of transcriptomes captured by rRNA depletion compared to poly(A) capture RNASeq methods. Accurate determination of the fraction of expressed eukaryote and bacterial genes that are captured by poly(A) enrichment compared to rRNA depletion methods requires that transcripts can be mapped back to assembled genomes of all partners. When this is not possible because genomes are not available for both host and symbionts, it is unclear to what extent the transcriptomes represent an accurate picture of the contributions of various partners in the symbiosis. To quantitatively compare the efficacy of Poly(A)-RNA-Seq versus rRNA depletion-RNA-Seq for the capture of both animal host and bacterial symbiont transcriptomes, here we apply both strategies to an analysis of gene expression in the demosponge *Amphimedon queenslandica* holobiont. *A. queenslandica* hosts a low diversity microbiome that is dominated by three proteobacterial symbionts (*AqS1*, *AqS2* and *AqS3*) through most of the life cycle (Fieth et al., 2016). The genomes of both the host sponge (Fernandez-Valverde et al. 2015; Srivastava et al. 2010) and of the three primary symbionts have been assembled and annotated (Gauthier et al. 2016; Xiang 2021). These foundational genomic resources allow us to align transcript reads precisely to their associated genomes, making the *A. queenslandica* holobiont a useful system to compare the power of the two RNA-seq library construction methods.

Materials And Methods

Sample collection and sequencing

Adult *Amphimedon queenslandica* were collected from Heron Island Reef, Great Barrier Reef, Queensland, Australia (Latitude - 23.44, Longitude 151.92) as described previously (Leys et al. 2008). Six separate tissue biopsies, each approximately 3 cm³, were sampled immediately after collection as a source of RNA for this study. To increase the proportional representation of the low abundance bacterial symbionts in the biopsies, bacteria were enriched by a series of centrifugation and filtration steps following the protocol of Thomas et al. (2010). Bacterial cell enrichments were stained with 4',6-diamidino-2-phenylindole, dihydrochloride (DAPI) following the manufacturer's protocol and visualised by fluorescence microscopy to determine approximate proportions of sponge to bacteria cells.

Total RNA was extracted separately from the six replicated biopsies using TRIzol (Sigma-Aldrich) and DNA was removed with Deoxyribonuclease I (Invitrogen), both following the manufacturer's protocol. RNA was quantified by Qbit® 2.0 Fluorimeter (Invitrogen), and quality was assessed by 1% TAE agarose gel electrophoresis. Illumina libraries were constructed separately from each of the six total RNAs using an Illumina TruSeq Stranded mRNA Library Prep Kit.

Three of the samples were subjected to the standard TruSeq Stranded mRNA-seq workflow, using the poly-A pulldown and standard protocol with 1 µg total RNA as input. The other three samples (2.5 µg total RNA each) were first rRNA-depleted using the Ribo-Zero Gold ribosomal RNA Removal Kit (Epidemiology, supplied by Illumina; <https://sapac.illumina.com/products/by-type/molecular-biology-reagents/ribo-zero-gold-rrna-removal-epidemiology.html>), which removes animal, Gram-positive and Gram-negative bacterial

cytoplasmic and mitochondrial rRNA. 100 ng of the depleted RNA was then used as input to the mRNA-seq protocol, but omitting the poly-A pulldown and instead of starting at the RNA fragmentation step.

From the fragmentation step onwards, the TruSeq protocol was followed for all six samples, using random hexamer priming for the generation of cDNA. We used 12 PCR cycles for the amplification. Each library was sequenced on the Illumina NextSeq500 platform in four runs. Both library preparation and sequencing were performed at Ramaciotti Centre for Genomics, Sydney, Australia. Thus, all six samples were treated identically throughout except for their random assignment to either poly(A) capture or rRNA depletion.

Read processing and alignment

Raw paired-end sequences of 75 base pairs (bp) in length were processed using Trimmomatic (version 0.36) (Bolger et al. 2014) to crop the first 10 bp of each read, and to trim the reads using a 4 bp sliding window and an average quality threshold of 20. Both unpaired reads and resulting reads smaller than 60 bp were discarded. The remaining quality-filtered, paired-end reads were aligned to the *A. queenslandica*, *AqS1*, *AqS2* and *AqS3* genomes (Fernandez-Valverde et al. 2015; Srivastava et al. 2010; Gauthier et al. 2016; Xiang 2021) using HISAT2 (version 2.0.5) with default parameters (Pertea et al., 2016). Both unaligned reads, and reads that aligned to more than one of these four genomes, were discarded. The read mapping bam file of each sample was split into four bam files based on the scaffold id of each species by SAMtools (version 1.3) (Li et al. 2009).

Comparative assessment of gene depth and gene coverage

The number of reads that aligned to the sponge or bacterial protein-coding sequences (CDS) were counted by htseq-count (v. 0.11.2.) from the HTSeq framework, using default parameters except for –stranded = reverse (Anders et al. 2015). These aligned reads were then converted to transcript counts per gene as a measure of gene depth. In addition, for each genome, gene coverage was estimated as the percentage of genes that were represented in the RNA-Seq data (that is, expressed gene number divided by total gene number). These two measures were compared within and between libraries prepared by each of the two different methods, and with a previously generated adult transcriptome to provide further context, as described below.

Comparison with an RNA-Seq dataset generated previously from non-bacterial-enriched sponge sample

To specifically assess the effect of bacterial enrichment on the ability to recover holobiont transcriptomes, we compared the data generated in this study with previously published deep RNA-Seq data (Fernandez-Valverde et al. 2015), which were stranded Poly(A)-RNA libraries derived from a single adult sponge with no bacterial enrichment step. We directly compared results from this published data with that of bacterial-enriched Poly(A)-RNA-Seq data from this study. To do so, the raw paired-end reads were filtered by Trimmomatic (parameters: SLIDINGWINDOW:4:15 MINLEN:60 HEADCROP:7) (version 0.36) (Bolger et al. 2014). The remaining high-quality pair-end reads were aligned to the holobiont

genomes as described above. Gene counts and gene coverage were also calculated from these data as described above.

Comparison of gene function among different data sets

To investigate the possibility of functional biases resulting from differential capture of transcripts, Gene Ontology (GO) analyses were performed for the genes captured by the three different library preparation methods, rRNA-depleted-RNA-Seq, Poly(A)-RNA-Seq and bacterial-unenriched Poly(A)-RNA-Seq. GO term annotations of the holobiont gene models were performed using Blast2GO (version 5.2.4) (Conesa and Gotz 2008), and the number of genes assigned to each annotated GO term was calculated for each species. Expressed CDS in each sample were identified as those to which at least one read pair mapped and expressed GO terms were scored when there was at least one gene expressed that could be attributed to the GO term. To identify any differences in the GO functions among the different RNA-Seq libraries, the expressed GO percentages were estimated in each sample for the sponge host and for each of the three bacterial symbionts, separately. That is, for each of the four species, the number of expressed GO terms in a given sample was divided by the total number of GO terms annotated in that species; the three replicated samples per RNA-Seq dataset were averaged. As several genes would likely be annotated with the same GO function, for each expressed GO term, the number and proportion of expressed genes were also calculated for each RNA-Seq data set; the three replicated samples per RNA-Seq dataset were averaged.

Results

Sequencing and alignment profile

Using four lanes of Illumina NextSeq500, we generated 388 and 403 million reads for the three replicate rRNA-depleted-RNA-Seq samples (A, B, C) and the three Poly(A)-RNA-Seq samples (D, E, F), respectively. After filtering low-quality reads with Trimmomatic (Bolger et al., 2014), more than 86% of reads (~ 336 and ~ 354 million reads for rRNA-depleted and Poly(A) samples, respectively) remained for subsequent analysis. For the additional bacterial-unenriched Poly(A)-RNA-Seq transcriptomes, 341 million reads were generated for the four samples, and 320 million reads remained after the quality filter step. All these high-quality reads were then aligned to updated versions of *A. queenslandica* and its proteobacterial symbionts *AqS1*, *AqS2* and *AqS3* genomes (Xiang 2021) by HISAT2 (version 2.0.5) (Pertea et al. 2016). On average, 69.3% rRNA-depleted-RNA-Seq reads, 81.2% Poly(A)-RNA-Seq reads, and 63.1% bacterial-unenriched Poly(A)-RNA-Seq reads mapped to the four genomes. Of these mapped reads, 4.8% rRNA-depleted-RNA-Seq reads, 0.1% Poly(A)-RNA-Seq and few bacterial-unenriched Poly(A)-RNA-Seq reads mapped to more than one genome, and were discarded because they could not be unambiguously assigned. 222 million mapped rRNA-depleted-RNA-Seq reads, 287 million mapped Poly(A)-RNA-Seq reads, and 51 million mapped bacterial-unenriched Poly(A)-RNA-Seq reads remained for further analysis.

For the bacterial-enriched rRNA-depleted-RNA-Seq libraries, 77% of the mapped reads were attributed to the sponge host *Aq*, while 19.2%, 2.54%, and 1.3% were attributed to *AqS1*, *AqS2*, and *AqS3*, respectively

(Fig. 1A). By comparison, for the bacterial-enriched Poly(A)-RNA-Seq libraries, almost all the aligned reads (99.2%) were attributed to the sponge host; only 0.6%, 0.2%, and 0.1% reads were attributed to bacterial symbiont genomes *AqS1*, *AqS2*, and *AqS3*, respectively (Fig. 1B). For the bacterial-unenriched Poly(A)-RNA-Seq library, 99.9% of mapped reads were attributed to the sponge host, and only 0.03% to the three symbiotic bacteria in total.

Correlation between samples

The reads aligned to each gene were considered as a measure of gene expression. Correlation of the gene expression between the technical triplicates was estimated by gene read counts (gene depth) between individuals (biological replicates) within each experimental group. All read counts were included in calculating the Spearman correlation efficiencies. For the host sponge *A. queenslandica*, the gene expression correlation was strong among the triplicates from both rRNA-depleted- and Poly(A)-RNA-Seq groups, with Spearman correlation efficiencies ranging from 0.907 to 0.969 (Fig. 1C, D). For the three symbiotic bacteria, the correlation of gene expression varied between the biological replicates from each experimental group. For the rRNA-depleted-RNA-Seq data, the Spearman correlation efficiencies of *AqS1* biological replicates ranged from 0.967 to 0.987, with *AqS2* from 0.916 to 0.941, and *AqS3* from 0.826 to 0.889. These were higher than those correlation efficiencies of the Poly(A)-RNA-Seq replicates, with *AqS1* ranged from 0.915 to 0.920, *AqS2* from 0.843 to 0.855, and *AqS3* from 0.695 to 0.733. Overall, the strong correlations between the biological replicates from each experimental group suggested the gene expression profiles of the four species in the *A. queenslandica* holobiont were comparable within the rRNA-depleted- and the Poly(A)-RNA-Seq samples.

Gene coverage

Transcript quantification is one of the most common applications of RNA-Seq; it estimates the gene expression levels based on the number of reads mapped to each gene. This approach is widely used to identify differential expression (DE) of genes between different treatments and contexts (Conesa et al., 2016). Many DE software packages filter genes with < 5 reads before identifying DE genes because of both biological and statistical concerns, removing genes with low read counts prior to downstream analysis (Chen et al. 2016). To better understand the influence of RNA-Seq library preparation method for capture of both animal host and bacterial symbiont transcriptomes, we compared the percentage of genes to which RNA-Seq reads were mapped across the different methods (Fig. 2A). Specifically, we filtered the genes with two thresholds (5 and 1 read pairs) and then calculated the gene coverage.

When we used the minimal threshold of 1 read pairs mapped, the percent of *A. queenslandica* genes covered by the rRNA-depleted-RNA-Seq data (53.93%) was very similar to that of the Poly(A)-RNA-Seq data (57.38%) (two-tailed t-test, p-value 0.308), and both of these were lower than the bacterial unenriched-Poly(A)-RNA-Seq data (66.62%) (Fig. 2A). For the bacterial symbiont genomes, on average, 89.70% (± 0.0082) *AqS1* genes, 95.11% (± 0.0029) *AqS2* genes, and 80.36% (± 0.0260) *AqS3* genes were captured by rRNA-depleted-RNA-Seq libraries. These gene coverages were much higher than those captured by Poly(A)-RNA-Seq libraries, which, on average, were 70.07% (± 0.0119) for *AqS1*, 65.99% (\pm

0.0181) for *AqS2*, and 44.46% (± 0.0259) for *AqS3* (one-tailed t-test, p-values were 2.54e-05 in *AqS1*, 4.98e-04 in *AqS2*, and 3.56e-05 in *AqS3*). The coverages also were higher than that of the bacterial unenriched-Poly(A)-RNA-Seq samples, which were 18.14% in *AqS1*, 24.06% in *AqS2*, and 1.13% in *AqS3*. This gene coverage comparison indicated that rRNA-depleted-RNA-Seq and Poly(A)-RNA-Seq could capture the animal host transcriptomes at the same level, but rRNA-depleted-RNA-Seq was better at capturing the bacterial symbiont transcriptomes. The unenriched-Poly(A)-RNA-Seq was less likely to comprehensively capture bacterial symbiont transcriptomes compared to the rRNA-depleted-RNA-Seq and Poly(A)-RNA-Seq.

When the threshold was raised to at least 5 read pairs per gene, on average, 50.94% and 49.92% *A. queenslandica* genes were covered with in the rRNA-depleted-RNA-Seq data and Poly(A)-RNA-Seq data, respectively, with 58.66% of the genes represented in bacterial unenriched-Poly(A)-RNA-Seq data (Fig. 2A). These values are consistent with the estimated proportion of genes in the genome that are expressed in adult *A. queenslandica* (Fernandez-Valverde et al. 2015). For the symbiont bacterial genes, on average, 86.23% (± 0.0079) *AqS1* genes, 93.65% (± 0.0087) *AqS2* genes and 75.63% (± 0.0344) *AqS3* genes were captured in rRNA-depleted-RNA-Seq data. BY comparison, in the Poly(A)-RNA-Seq data, on average, only 54.22% (± 0.0016) *AqS1* genes, 41.07% (± 0.0123) *AqS2* genes, and 19.57% (± 0.0130) *AqS3* genes were captured. In the bacterial unenriched-Poly(A)-RNA-Seq samples, only 3.22% *AqS1* genes, 5.43% *AqS2* genes, and 0.20% *AqS3* genes were captured. These gene coverages indicate that rRNA-depleted-RNA-Seq captured a similar proportion of sponge genes and a higher proportion of bacterial genes than Poly(A)-RNA-Seq. For the unenriched-Poly(A)-RNA-Seq sample, the coverage was representative only for the sponge transcriptome but not for the symbiotic bacterial transcriptomes.

Distribution of read depth in gene coding regions

Sequence depth is a crucial measure of the robustness of RNA-Seq data. We calculated the gene depth by counts of aligned reads per gene, and compared the gene depth distribution of the rRNA-depleted-, Poly(A)- and bacterial unenriched-Poly(A)-RNA-Seq samples (Fig. 2B). For the host sponge *A. queenslandica*, on average, the median gene depths were 110 (± 15.95) and 113 (± 1.73) reads per gene for the rRNA-depleted- and Poly(A)-RNA-Seq libraries, and 83 for the bacterial unenriched-Poly(A)-RNA-Seq adult. For each bacterial symbiont, the rRNA-depleted-RNA-Seq data showed the highest average gene depths, which were higher than these of the Poly(A)-RNA-Seq data and bacterial unenriched-Poly(A)-RNA-Seq data. On average, the rRNA-depleted-RNA-Seq median gene depths were 446 (± 89.37) in *AqS1*, 162 (± 84.50) in *AqS2*, and 47 (± 13.75) in *AqS3*; the Poly(A)-RNA-Seq median gene depths were 14 (± 0.58) in *AqS1*, 7 (± 0.58) in *AqS2*, and 4 (± 0) in *AqS3* (Fig. 2B). The median gene depth of all the three primary symbiotic bacterial was 2 for the bacterial unenriched-Poly(A)-RNA-Seq sample. The similar read depth distributions of *A. queenslandica* for rRNA-depleted- and Poly(A)-RNA-Seq samples implied that there was no noteworthy gene depth difference among the host sponge transcriptome captured by both methods, but sponge transcript depths were lower than obtained by bacterial unenriched-Poly(A)-RNA-Seq (Fig. 2B). The symbiont gene depths were much higher in the rRNA-depleted-RNA-Seq data compared to both the Poly(A)-RNA-Seq and the bacterial unenriched-Poly(A)-RNA-Seq data.

Comparison of the expressed biological functions in the different RNA-Seq data sets

A first step towards understanding gene functions in host and symbionts is to annotate expressed and enriched genes in RNA-Seq data sets using Gene Ontology (GO) (Ashburner et al., 2000). GO analyses were performed and compared between rRNA-depleted-RNA-Seq, Poly(A)-RNA-Seq and bacterial unenriched Poly(A)-RNA-Seq using Blast2GO (version 5.2.4) (Conesa & Gotz, 2008) to determine if any of the RNA-Seq approaches resulted in a bias in the types of genes captured in the transcriptomes. In total, 28,570 *A. queenslandica* genes, 2,417 *AqS1* genes, 1,193 *AqS2* genes, and 1,977 *AqS3* genes were assigned to 6,357, 1,317, 904, and 1,030 GO terms, respectively.

To identify any differences in GO functional annotations among the different RNA-Seq libraries, the percentage of expressed GO terms in each sample was calculated for each species. That is, for a given species, if any gene in the genome that had been assigned to a particular GO term was expressed with at least 1 read pair in a given sample, that GO term was considered as an expressed GO term. The expressed GO percentage was estimated as the percent of total GO terms for the genome that were represented in the RNA-Seq data, with replicates averaged (Fig. 3A; Supplementary Table S1). For *A. queenslandica*, almost all the GO terms were present in the rRNA-depleted-RNA-Seq data (95.88% \pm 0.0085), Poly(A)-RNA-Seq data (97.43% \pm 0.0036) and bacterial unenriched-Poly(A)-RNA-Seq data (98.47%). For the bacterial symbionts, the expressed GO term percentages were 98.96% \pm 0.002 in *AqS1*, 99.34% \pm 0.001 in *AqS2* and 98.32% \pm 0.006 in *AqS3* for the rRNA-depleted-RNA-Seq samples (Fig. 3A). These percentages of expressed GO terms in rRNA-depleted-RNA-Seq samples were higher than in Poly(A)-RNA-Seq samples, which was 92.63% \pm 0.006 in *AqS1*, 86.32% \pm 0.01 in *AqS2*, and 79.58% \pm 0.014 in *AqS3* (one-tailed t-test, p-values were 3.68e-04 in *AqS1*, 9.37e-04 in *AqS2*, and 1.91e-04 in *AqS3*); and also higher than that of bacterial unenriched-Poly(A)-RNA-Seq sample, which was 38.72% in *AqS1*, 46.02% in *AqS2*, and 2.91% in *AqS3* (Fig. 3A). These GO term expression percentages indicate that all the different libraries captured a similar functional transcriptional profile of the host sponge *A. queenslandica*. In contrast, while rRNA-depleted-RNA-Seq provides a near-complete functional representation of the symbiotic bacteria, this is not true for the Poly(A)-RNA-Seq and bacterial unenriched-Poly(A)-RNA-Seq.

GO functional profiles of expressed and over-represented (enriched) genes in the three RNA-Seq data sets, with replicates averaged, were compared to determine if GO categories were equally represented in the different types of RNA-Seq libraries (Fig. 3B; Supplementary Table S1). This approach revealed differences in the representation of specific GO functions in each of the libraries. For the sponge *A. queenslandica*, on average, the expressed gene percentage of each GO term was higher in the bacterial unenriched-Poly(A)-RNA-Seq data (93.64%) compared to the rRNA-depleted-RNA-Seq data (86.57%) and the Poly(A)-RNA-Seq data (90.62%). For each bacterial symbiont, the rRNA-depleted-RNA-Seq data showed the highest GO term expressed gene percentages compared to the Poly(A)-RNA-Seq data and bacterial unenriched-Poly(A)-RNA-Seq data (Fig. 3B). The rRNA-depleted-RNA-Seq average GO term expressed gene percentages were similar, 95.57, 98.62% and 95.62% for *AqS1*, *AqS2*, and *AqS3*, respectively. In contrast, the average GO term expressed gene percentages of Poly(A)-RNA-Seq and bacterial unenriched-Poly(A)-RNA-Seq for the three symbionts varied considerably. For the Poly(A)-RNA-

Seq, the average GO term expressed gene percentages decreased from 84.49% in *AqS1*, to 77.45% in *AqS2* and 67.85% in *AqS3*. For the bacterial unenriched-Poly(A)-RNA-Seq, this percentage decreased from 21.88% in *AqS1* and 31.27% in *AqS2* to 0.43% in *AqS3* (Fig. 3B). These GO term percentages suggest that for the symbionts, the Poly(A)-RNA-Seq method produces bias in the functional types of genes that are captured, as indicated by GO enrichment results.

We used pairwise comparisons of expressed gene numbers to identify which functional groups of genes, as indicated by GO terms, were most differently captured by the different library types. For each pairwise comparison, the most differentially represented 30 GO terms were identified by the biggest differences of the expressed gene numbers. The most differentially represented GO terms included four biological process GOs (oxidation-reduction process, proteolysis, transmembrane transport, and translation), four cellular component GOs (integral component of membrane, membrane, cytoplasm, and plasma membrane) and 13 molecular function GOs (i.e. catalytic activity, hydrolase activity, transferase activity, ATP binding, and DNA binding) (Fig. 3C). Overall, compared to unenriched-Poly(A)-RNA data, the rRNA-depleted-, and Poly(A)-RNA data sets under-represented sponge genes involved in integral component of membrane, membrane, ATP binding, and protein binding (Fig. 3C). For the three symbiotic bacteria, compared to the rRNA-depleted RNA data, the Poly(A)- and bacterial unenriched-Poly(A)-RNA data underrepresented bacterial genes involved in oxidation-reduction, transmembrane transport, integral component of membrane, membrane, and ATP binding (Fig. 3C).

Discussion

Transient polyadenylation in bacteria reduces mRNA capture in Poly(A)-RNA-Seq

Sequencing depth and coverage are central to assessing the efficacy of RNA-Seq (Sims et al. 2014). In this transcriptome analysis of the four dominant partners in the marine sponge *A. queenslandica* holobiont, we sought to identify an RNA-Seq approach that could yield high depth and coverage for genes expressed in both the sponge host and its three proteobacterial symbionts. Under conditions used in this study – i.e., libraries from bacterial-enriched cell preparations from adult sponges – we found that both rRNA-depleted- and Poly(A)-RNA-Seq capture sponge host transcripts at similar levels with at least 5 read pairs aligned (on average, 50.94% and 49.92% of genes in the genome). In contrast, only the rRNA-depleted-RNA-Seq captured a large proportion of the expressed genes in the symbionts (86.23% *AqS1*, 93.65% *AqS2*, and 75.63% *AqS3* of genes per genome), compared to much smaller proportions obtained by Poly(A)-RNA-Seq (54.22% *AqS1*; 41.07% *AqS2*; 19.57% *AqS3* of genes per genome). This difference reflects the different capacities of these two RNA-Seq approaches to efficiently capture bacterial mRNAs, which are relatively unstable (Selinger et al. 2003) and are polyadenylated just prior to degradation (Dreyfus & Regnier, 2002). The Poly(A)-RNA approach captures a smaller number of bacterial mRNAs at any one sampling time, because it can only capture those in the transient state of polyadenylated immediately prior to degradation.

Interestingly, however, the subset of bacterial mRNAs captured by Poly(A)-RNA-Seq method does not appear to be a random snapshot of the bacterial transcriptome profile. Combining all the three bacterial enriched Poly(A)-RNA-Seq samples and the bacterial unenriched Poly(A)-RNA-Seq sample, 502, 396, and 29 GO terms were represented in the *AqS1*, *AqS2*, and *AqS3* expressed genes, respectively; 23 of these were expressed in all the three symbionts. 59, 50, and 93 of the annotated GO terms were not detected in *AqS1*, *AqS2*, and *AqS3*, respectively; and one of these, GO:0002100 (tRNA wobble adenosine to inosine editing), was not detected in any of the three symbionts. This bias appears to be in contrast to bacterial polyadenylation being a ubiquitous post-transcriptional modification (Maes et al. 2017) and bacterial poly(A) polymerases having no mRNA sequence specificity (Haugel-Nielsen et al. 1996). Given that both exert a general RNA destabilising function in bacteria (Dreyfus and Regnier 2002; Maes et al. 2017), the differences of the detected and undetected GO terms in the poly(A)-RNA samples in this study might be explained by differential mRNA turnover rate. Bacterial mRNA turnover is modulated by ribonucleases (RNases), RNA binding proteins and small noncoding RNAs (sRNA) (Anderson & Dunman 2009), and could be facilitated by poly(A) polymerase (Hajnsdorf and Kaberdin 2018). These regulatory processes can play a vital role in bacterial stress adaptation, cell growth and virulence factor production (Anderson and Dunman 2009; Li et al. 2017). Of total *Escherichia coli* RNA content, only 110 genomic regions were detected as being polyadenylated; other polyadenylated RNAs might degrade too rapidly to be detected (Maes et al. 2017). Thus, variable mRNAs decay rates (Hui et al. 2014) could yield biased subsets of bacterial mRNAs as captured by poly(A)-RNA-Seq in this study.

rRNA-depleted-RNA-Seq best captures the holobiont transcriptome

We found that the rRNA-depleted-RNA-Seq simultaneously captures a high proportion of the expressed genes in the sponge host and in the bacterial symbionts, which is in contrast to Poly(A) – based RNA-Seq approaches. Importantly, there was not a marked difference between sponge host transcriptome representation in rRNA-depleted- and Poly(A)-RNA-Seq data sets that were derived from bacterial cell enrichment. The output from both rRNA-depleted- and Poly(A)-RNA-Seq datasets indicate that (i) a sufficient number of sponge cells are captured using this bacterial cell enrichment method to generate highly representative RNA-Seq libraries and (ii) there is no substantial loss of host transcript abundance or representation in the rRNA-depleted-RNA-Seq dataset. However, all three bacterial symbiont transcriptomes are significantly better covered in the rRNA-depleted-RNA-Seq. This high bacterial gene coverage in the rRNA-depleted-RNA-Seq libraries is similar to the human-pathogen system where an rRNA-depleted-RNA-Seq method captured 88% of the *Haemophilus influenzae* genes from *in vitro* human bronchial epithelium; the human transcriptome was simultaneously captured (Baddal et al. 2015). In addition, rRNA-depleted-RNA-Seq also improves detection of low abundance transcripts (Kim et al. 2019; Petrova et al. 2017) and noncoding RNA (Westermann et al. 2016), which could reveal a more complex host-symbiont transcriptome profile.

We provide here the first, to the best of our knowledge, quantitative comparison of RNASeq methods that deal with the high proportion of rRNAs in both eukaryotic and prokaryotic transcriptomes by either selecting for polyadenylated mRNAs or depleting rRNAs. In doing so, we empirically demonstrate that the

latter approach – depleting rRNAs prior to construction of an RNASeq library – can be applied to a marine invertebrate animal holobiont to successfully achieve simultaneous and comprehensive capture of transcripts of both the animal host and its bacterial symbionts. Although selection for polyadenylated mRNAs, which is the approach routinely used for animal transcriptomes, does indeed capture some of the bacterial symbiont transcripts, it does so in a biased manner that renders it unsuitable for symbiosis studies where the goal is to study gene expression in both host and symbionts. Instead, targeting rRNA depletion to both host and symbionts in a single step provides a time- and cost-efficient approach that can proceed through to simultaneous capture of transcriptomes from all partners using the same biological samples. The approach should be relatively easily applied to any holobiont comprising a eukaryotic host and bacterial symbionts.

Declarations

Supplementary Information The online version contains supplementary material available at [https:// xxx](https://xxx).

Acknowledgments

We thank Mr Nick Rhodes, QCIF, for computational support, and Mr Chris Challen for aquarium maintenance at UQ. The Heron Island Research Station provided logistical support for collection of animals from the field.

Funding Information

We are very appreciative of funding provided by Australian Research Council grants DP170102353 and DP190102521, to SMD and BMD.

Competing Interests: The Authors have no financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

Author Contributions: The study was conceptualised by SMD and BMD. Biological material was collected and prepared by DP. All data analysis and preparation of figures was conducted by XX. The original manuscript draft was prepared by XX, and was edited by SMD and BMD. All authors reviewed the manuscript. Funding was acquired by SMD and BMD.

Data Availability: Genomes of the marine sponge *A. queenslandica* and of the three proteobacterial symbionts, *AqS1*, *AqS2* and *AqS3*, used in this study are available at NCBI under Bioproject accession PRJNA668660t. RNASeq datasets generated in this study are available at NCBI under accession ##### (to be advised prior to publication, as we are currently awaiting finalisation of the data submission).

References

1. Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169

2. Anderson KL, Dunman PM (2009) Messenger RNA turnover processes in *Escherichia coli*, *Bacillus subtilis*, and emerging studies in *Staphylococcus aureus*. *Int J Microbiol.* 2009:525491
3. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25–9
4. Baddal B, Muzzi A, Censini S, Calogero RA, Torricelli G, Guidotti S, Taddei AR, Covacci A, Pizza M, Rappuoli R, Soriani M, Pezzicoli A (2015) Dual RNA-seq of nontypeable *Haemophilus influenzae* and host cell transcriptomes reveals novel insights into host-pathogen cross talk. *mBio* 6:e01765-15
5. Bolger, M, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120
6. Bordenstein SR, Theis KR (2015) Host biology in light of the microbiome: Ten principles of holobionts and hologenomes. *PLoS Biol* 13:1002226
7. Chen Y, Lun AT, Smyth GK (2016) From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000 Res* 5:1438.
8. Conesa A, Götz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics* 2008:619832
9. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesnia MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A (2016) A survey of best practices for RNA-seq data analysis. *Genome Biol* 17:13
10. Dendooven T, Luisi BF, Bandyra KJ (2020) RNA lifetime control, from stereochemistry to gene expression. *Curr Opin Struct Biol*:59–70
11. Dreyfus M, Regnier P (2002) The poly(A) tail of mRNAs: Bodyguard in eukaryotes, scavenger in bacteria. *Cell* 111: 611–613
12. Engelstadter J, Hurst GDD (2009) The ecology and evolution of microbes that manipulate host reproduction. *Ann Rev Ecol Evol Syst* 40:127–149
13. Ezenwa VO, Gerardo NM, Inouye DW, Medina M, Xavier JB (2012) Animal behavior and the microbiome. *Science*, 338:198–199
14. Fernandez-Valverde SL, Calcino AD, Degnan BM (2015) Deep developmental transcriptome sequencing uncovers numerous new genes and enhances gene annotation in the sponge *Amphimedon queenslandica*. *BMC Genomics* 16:387
15. Fieth, R. A., Gauthier, M.-E. A., Bayes, J., Green, K. M., & Degnan, S. M. (2016) Ontogenetic changes in the bacterial symbiont community of the tropical demosponge *Amphimedon queenslandica*: Metamorphosis is a new beginning. *Front Mar Sci* 3:228
16. Fiore CL, Labrie M, Jarett JK, Lesser MP (2015) Transcriptional activity of the giant barrel sponge, *Xestospongia muta* holobiont: molecular evidence for metabolic interchange. *Front Microbiol* 6:364

17. Gauthier M-EA, Watson JR, Degnan SM (2016) Draft genomes shed light on the dual bacterial symbiosis that dominates the microbiome of the coral reef sponge *Amphimedon queenslandica*. *Front Mar Sci* 3: 196.
18. Germer J, Cerveau N, Jackson DJ.(2017) The holo-transcriptome of a calcified early branching metazoan. *Front MarSci* 4:00081
19. Grote A, Voronin D, Ding T, Twaddle A, Unnasch TR, Lustigman S, Ghedin E (2017) Defining *Brugia malayi* and *Wolbachia* symbiosis by stage-specific dual RNA-seq. *PLoS Negl Trop Dis* 11::e0005357
20. Hajnsdorf E, Kaberdin VR (2018) RNA polyadenylation and its consequences in prokaryotes. *Phil Trans Roy Soc B* 373: 20180166
21. Haugel-Nielsen J, Hajnsdorf E, Regnier P (1996) The rpsO mRNA of *Escherichia coli* is polyadenylated at multiple sites resulting from endonucleolytic processing and exonucleolytic degradation. *EMBO J* 15:3144–3152
22. Heinken A, Thiele I (2015). Systems biology of host-microbe metabolomics. *WIRE Syst Biol Med* 7:195–219
23. Hentschel U, Piel J, Degnan SM, Taylor MW (2012) Genomic insights into the marine sponge microbiome. *Nat Rev Microbiol* 10:641–54
24. Hui MP, Foley PL, Belasco JG (2014) Messenger RNA degradation in bacterial cells. *Annu Rev Genet* 48:537–559
25. Kim IV, Ross EJ, Dietrich S, Döring K, Sánchez Alvarado A, Kuhn CD (2019) Efficient depletion of ribosomal RNA for RNA sequencing in planarians. *BMC Genomics* 29:909
26. Leulier F, MacNeil LT, Lee WJ, Rawls JF, Cani PD, Schwarzer M, Zhao LP, Simpson SJ (2017) Integrative physiology: At the crossroads of nutrition, microbiota, animal physiology, and human health. *Cell Metabol* 25:522–534
27. Leys SP, Larroux C, Gauthier M, Adamska M, Fahey B, Richards GS, Degnan SM, Degnan BM (2008) Isolation of amphimedon developmental material. *CSH Protoc:prot5095*
28. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD (2009) The Sequence Alignment/Map format and SAMtools. *Bioinform* 25:2078–2079
29. Li Z, Jiang J, Yu X, Wu C, Shen D, Feng Y (2017) Poly(A) polymerase I participates in the indole regulatory pathway of *Pantoea agglomerans* YS19. *Microbiol* 163:197–206
30. Maes A, Gracia C, Innocenti N, Zhang K, Aurell E, Hajnsdorf E (2017) Landscape of RNA polyadenylation in *E. coli*. *Nucleic Acids Res* 45:2746–2756
31. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll,AH, Kremer N, Mazmanian SK, Metcalf JL, Nealson K, Pierce NE, Rawls JF, Reid A, Ruby EG, Rumpho M, Sanders JG, Tautz D, Wernegreen JJ (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci USA* 110:3229–3236

32. Moitinho-Silva L, Díez-Vives C, Batani G, Esteves AI, Jahn MT, Thomas T. (2017) Integrated metabolism in sponge-microbe symbiosis revealed by genome-centered metatranscriptomics. *ISME J* 11:1651–1666
33. O'Brien PA, Tan S, Yang C, Frade PR, Andreakis N, Smith HA, Miller DJ, Webster NS, Zhang G, Bourne DG (2020) Diverse coral reef invertebrates exhibit patterns of phyllosymbiosis. *ISME J* 14:2211–2222
34. Perez-Ortin JE, Alepuz P, Chavez S, Choder M (2013) Eukaryotic mRNA decay: methodologies, pathways, and links to other stages of gene expression. *J Mol Biol* 425:3750–3775
35. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Proto* 11:1650–1667
36. Petrova OE, Garcia-Alcalde F, Zampaloni C, Sauer K (2017) Comparative evaluation of rRNA depletion procedures for the improved analysis of bacterial biofilm and mixed pathogen culture transcriptomes. *Sci Rep* 7: 41114
37. Pradeu T (2011) A mixed self: The role of symbiosis in development. *Biol Theory* 6:80–88
38. Robbins SJ, Song W, Engelberts JP, Glasl B, Slaby BM, Boyd J, Marangon E, Botté ES, Laffy P, Thomas T, Webster NS (2021) A genomic view of the microbiome of coral reef demosponges. *ISME J* 15:1641–1654
39. Rosenberg E, Zilber-Rosenberg I (2016) Microbes drive evolution of animals and plants: the hologenome concept. *MBio* 7:01395
40. Rowland I, Gibson G, Heinken A, Scott K, Swann, J, Thiele I, Tuohy K (2017) Gut microbiota functions: metabolism of nutrients and other food components. *Euro J Nutr* 56:1–24
41. Selinger DW, Saxena, RM, Cheung KJ, Church GM, Rosenow C (2003) Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res* 13:216–223
42. Sharova LV, Sharov AA, Nedorezov T, Piao Y, Shaik N, Ko MS. (2009) Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res* 16:45–58
43. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP (2014) Sequencing depth and coverage: key considerations in genomic analyses. *Nature Rev Genet* 15:121–132
44. Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U, Larroux C, Putnam NH, Stanke M, Adamska M, Darling A, Degnan SM, Oakley TH, Plachetzki D, Zhai Y, Adamski M, Calcino A, Cummins SF, Goodstein DM, Harris C, Jackson D, Leys SP, Shu S, Woodcroft BJ, Vervoort M, Kosik KS, Manning G, Degnan BM, Rokhsar DS (2010) The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* 466:20–726
45. Steinert G, Busch K, Bayer K, Kodami S, Arbizu PM, Kelly M, Mills S, Erpenbeck D, Dohrmann M, Wörheide G, Hentschel U, Schupp PJ. (2020) Compositional and Quantitative Insights Into Bacterial and Archaeal Communities of South Pacific Deep-Sea Sponges (Demospongiae and Hexactinellida). *Front Microbiol* 11:716
46. Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A, Heidelberg KB, Egan S, Steinberg PD, Kjelleberg S (2010) Functional genomic signatures of sponge bacteria reveal unique and shared

features of symbiosis. ISME J 4:1557–1567

47. Upadhyay SK, Sharma S, Singh H, Dixit S, Kumar J, Verma,PC, Chandrashekar K (2015) Whitefly genome expression reveals host-symbiont interaction in amino acid biosynthesis. PLoS One 10: e0126751
48. Webster NS, Thomas T (2016) The Sponge Hologenome. MBio 7:e00135
49. Westermann AJ, Förstner KU, Amman F, Barquist L, Chao Y, Schulte LN, Müller L, Reinhardt R, Stadler PF, Vogel J (2016) Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. Nature 529:496–501
50. Westermann AJ, Gorski SA, Vogel J (2012) Dual RNA-seq of pathogen and host. Nature Rev Microbiol 10: 618–630
51. Xiang X (2021) Investigating host-symbiont crosstalk in the coral reef demosponge *Amphimedon queenslandica*. PhD Thesis, University of Queensland, Brisbane, Australia
52. Yang E, van Nimwegen E, Zavolan M, Rajewsky N, Schroeder M, Magnasco M, Darnell JE Jr (2003) Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. Genome Res 13:1863–72
53. Yen JH, Barr AR (1971) New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens* L. Nature 232:657–658

Figures

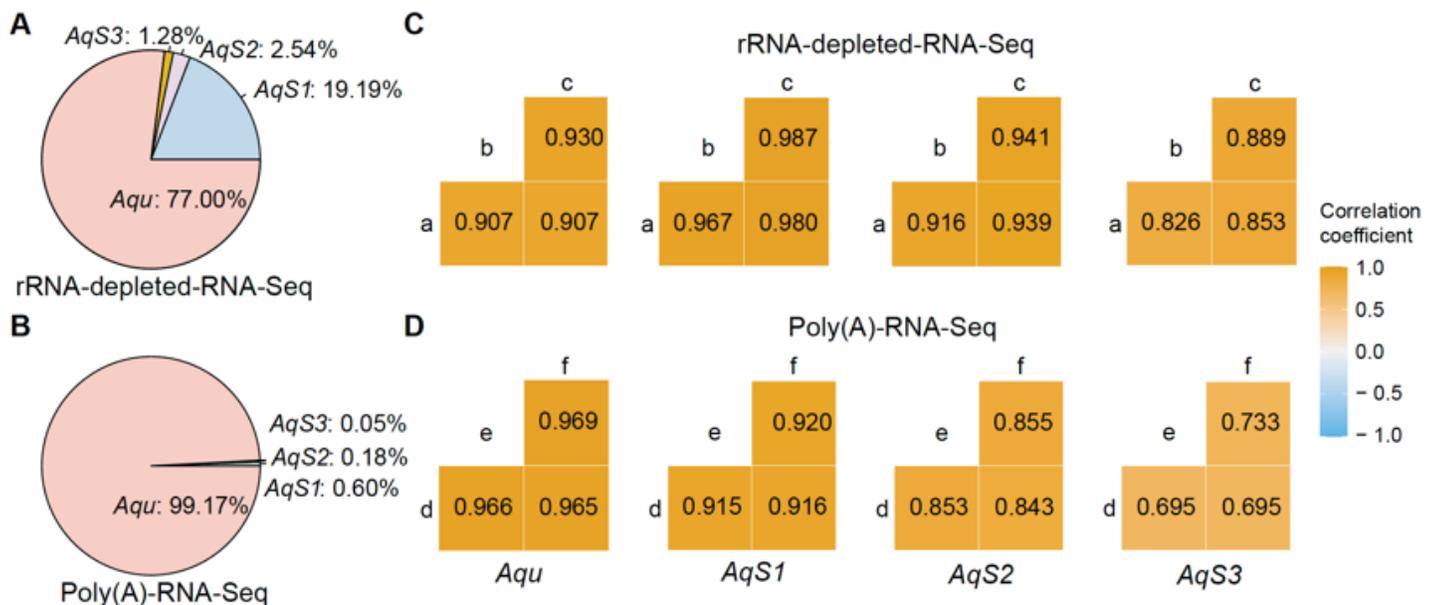


Figure 1

Comparison between two different RNA-Seq methods in capturing transcriptomes of the *A. queenslandica* holobiont. A, B. Taxonomic distribution of reads, expressed as percent of reads aligned to

A. *queenslandica* (*Aqu*) and its proteobacterial symbionts AqS1, AqS2 and AqS3 for rRNA-depleted-RNA-Seq (A) and Poly(A)-RNA-Seq (B) libraries. C, D. Correlation of expressed genes between three biological replicates for rRNA-depleted-RNA-Seq (C, replicates a-c) and Poly(A)-RNA-Seq (D, replicates d-f) libraries.

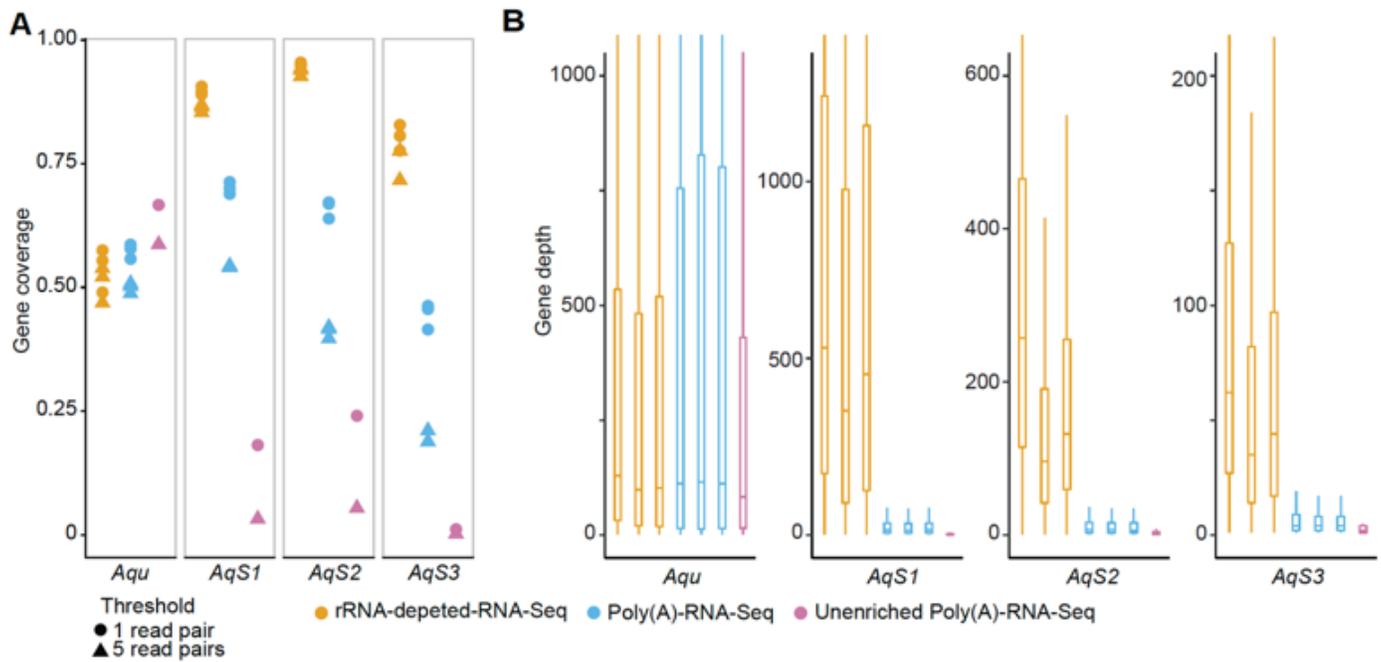


Figure 2

Gene coverage and depth of rRNA-depleted-RNA-Seq, Poly(A)-RNA-Seq and unenriched Poly(A)-RNA-Seq transcriptomes of the *A. queenslandica* holobiont. A. Gene coverage, expressed as percent of genes to which reads aligned to *A. queenslandica* (*Aqu*) and its proteobacterial symbionts AqS1, AqS2 and AqS3. Shown are results for 1 (dots) and 5 (triangles) reads mapped to CDS thresholds for the three biological replicate libraries (rRNA-depleted-RNA-Seq and Poly(A)-RNA-Seq) and single unenriched Poly(A)-RNA-Seq library. B. Gene depth, expressed as boxplot of the number of reads mapped to each expressed gene.

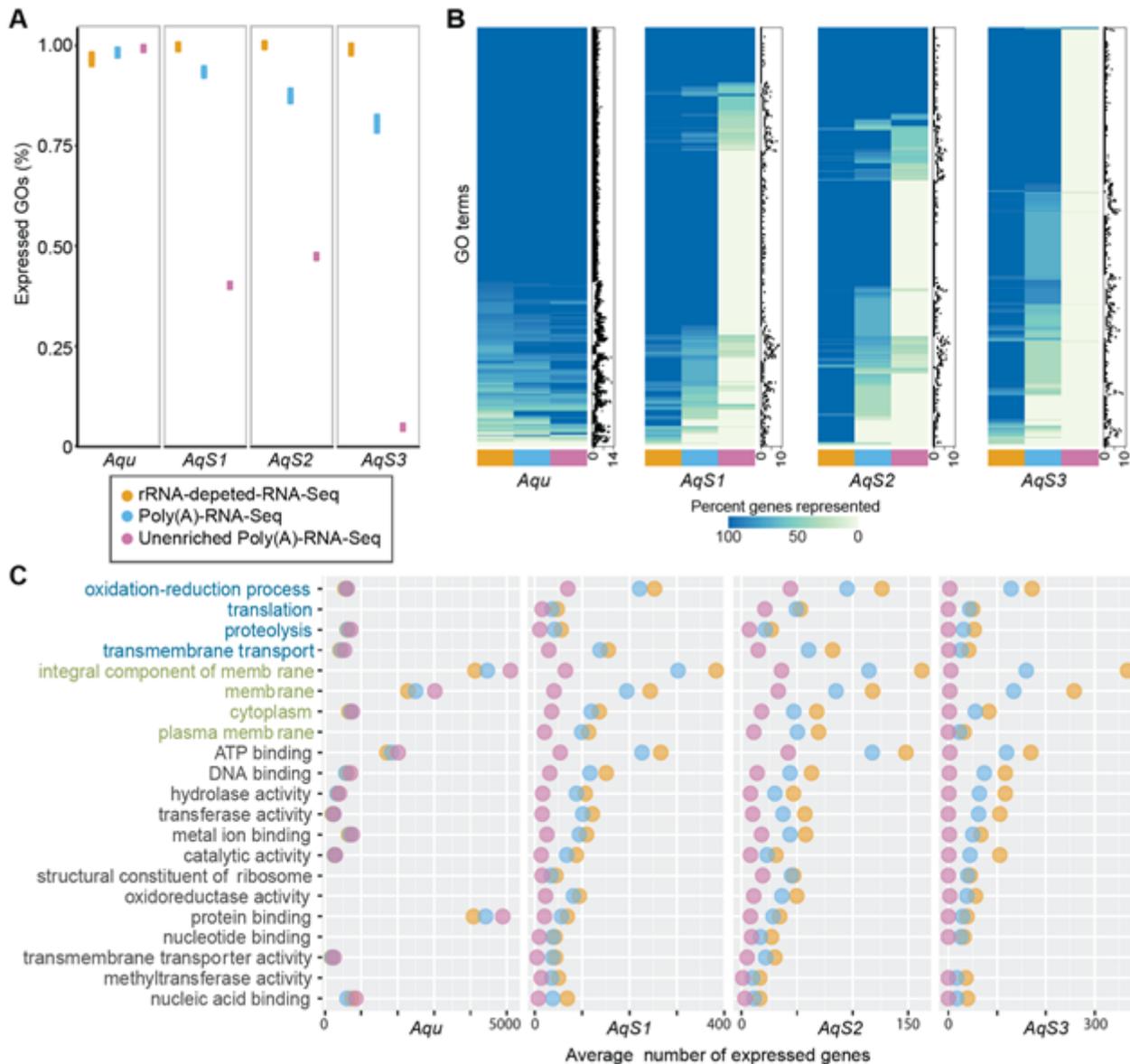


Figure 3

Gene ontology analyses of transcriptomes generated by the RNA-Seq data sets of the partners in the *A. queenslandica* holobiont. A. Percent of GO terms identified from expressed genes present in rRNA-depleted-RNA-Seq, Poly(A)-RNA-Seq and unenriched Poly(A)-RNA-Seq datasets with at least 1 read pair per CDS compared to genome GO terms. B. Heatmaps presenting the average percent of genes represented by each GO term; adjacent dot plots show total gene number attributed to each GO term with the x-axis on a log₂ scale. C. Average number of expressed genes for the 30 GO terms that are most differentially represented between the different RNA-Seq datasets.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTableS1.xlsx](#)