Published in final edited form as:
Biol Bull. 2012 August ; 223(1): 155–166.

Investigation into the Physiologies of *Aeromonas veronii in vitro* and Inside the Digestive Tract of the Medicinal Leech Using RNA-seq

Lindsey Bomar and **Joerg Graf***
Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut 06269

Abstract

Host-associated microbial communities are widespread in nature and vital to the health and fitness of the host. Deciphering the physiology of the microbiome *in vivo* is critical to understanding the molecular basis of the symbiosis. Recently, the development and application of high-throughput sequencing techniques, particularly RNA-seq, for studying microbial communities has enabled researchers to address not only which microbes are present in a given community but also how the community functions. For microbes that can also be cultivated in the laboratory, RNA-seq provides the opportunity to identify genes that are differentially expressed during symbiosis by comparing *in vitro* to *in vivo* transcriptomes. In the current study, we used RNA-seq to identify genes expressed by the digestive-tract microbiome of the medicinal leech, *Hirudo verbana*, and by one of the two dominant symbionts, *Aeromonas veronii*, in a rich medium. We used a comparative approach to identify genes differentially expressed during symbiosis and gain insight into the symbiont's physiology *in vivo*. Notable findings include evidence for the symbionts experiencing environmental stress, performing arginine catabolism, and expressing noncoding RNAs that are implicated in stationary phase survival, a state in which *A. veronii* persists for months within the host.

Introduction

A key component to understanding host-microbe associations is the physiology of the microbial community because it reveals insight into the contribution of the microbiome to the host and the conditions the microbes experience inside the animal (Wier *et al.*, 2010; Petersen *et al.*, 2011). Researchers can reconstruct metabolic pathways and develop testable hypotheses about how the microbial population functions by identifying genes that are expressed during symbiosis. Several techniques can be used to examine gene expression, providing insight into the microbiome's physiology. Reverse-transcription PCR (RT-PCR) or quantitative reverse-transcription PCR (qRT-PCR) are used to investigate differentially expressed genes (Noda *et al.*, 1999). These techniques are useful for examining a few targets; however, they are challenging for hundreds of targets and require *a priori* knowledge of the exact target sequences to which the primers anneal. These caveats make RT-PCR and qRT-PCR more useful in confirmatory experiments than in gene discovery applications. Clone libraries of mRNA-derived cDNA can also be used to gain access to the symbionts' physiologies and do not require sequenced genomes. For example, Tartar *et al.* (2009) used clone libraries to uncover the digestive contributions of the protists from the hindgut microbiome in the termite *Reticulitermes flavipes*. Though valuable, construction of cDNA clone libraries for prokaryotes is more challenging than for eukaryotes due to the lack of a

*To whom correspondence should be addressed. joerg.graf@uconn.edu.

poly-A tail, and cloning can introduce bias into the resulting data sets. Compared to next-generation sequencing approaches, clone library sequencing is low-throughput and not quantitative.

Traditionally, researchers have used microarrays to assess global gene expression in microorganisms, and this technique has been successfully applied to many host-associated microbiomes such as the gut microbiota of gnotobiotic mice and *Vibrio fischeri* in the squid light organ (Sonnenburg *et al.*, 2005; Martens *et al.*, 2008; Wier *et al.*, 2010). However, the application of microarrays is not as reliable for all host-associated microbiomes, particularly for field-caught animals whose symbiont communities can vary in composition. For instance, natural microbial communities can contain allelic variation, and transcripts from these variants may not be detected or may introduce an artificially lower expression level if the oligonucleotide probes do not match the transcript. The pangenome, which is the collective gene pool found within a particular species of prokaryotes, poses another challenge. Genome sequencing projects of strains belonging to the same bacterial species have shown for some species that as little as 40%–50% of a given genome is conserved between all strains examined. For example, a study that compared *Escherichia coli* genomes found that only about 2200 genes out of the approximately 4238 to 5589 genes encoded on each genome were conserved in the 17 genomes examined (Rasko *et al.*, 2008). If different strains of the same species are present and the pangenome is unknown, entire genes or genomic islands can be missed using microarrays. This problem is magnified if not the species but the function of the species in the association or guild is preserved, which is often the case in the digestive-tract microbiome of mammals (Dethlefsen *et al.*, 2007).

The advent of high-throughput sequencing in 2005 provided researchers with new opportunities to explore the metabolic potential of microorganisms by sequencing microbial genomes and metagenomes at a very high coverage (Margulies *et al.*, 2005). The application of high-throughput sequencing to cDNA constructed from RNA transcripts, known as RNA-seq, facilitated investigations into physiologies of naturally occurring microbial communities such as those found in the ocean (Frias-Lopez *et al.*, 2008; Gilbert *et al.*, 2008) and in the soil (Urich *et al.*, 2008). More recently, RNA-seq has been used to characterize the metatranscriptomes of host-associated microbial communities. Examples include the digestive-tract microbiomes of humanized mice using RNA from cecal contents (Turnbaugh *et al.*, 2009), and cecal contents of pigs (Poroyko *et al.*, 2010) and of humans using fecal samples as a proxy (Turnbaugh *et al.*, 2010). RNA-seq has also been used to study *in vivo* gene expression of the intracellular symbionts of bivalves (Stewart *et al.*, 2011). In contrast to microarrays, RNA-seq does not require *a priori* knowledge of community composition or genomes, though this information greatly aids in experimental design and data analysis. A challenge for this approach is the difficulty of removing rRNA prior to cDNA library preparation. rRNA often constitutes well over 90% of total RNA in a cell (Neidhardt, 1996), and the lack of a poly-A tail on mRNA in bacteria presents a challenge to enriching mRNA. However, mRNA enrichment methods that utilize oligonucleotides or enzymes that target rRNA are available (Yoder-Himes *et al.*, 2009; Poretsky *et al.*, 2009). In addition, the high-throughput and reduced sequencing costs of the Illumina platform (Luo *et al.*, 2012) can overcome this difficulty.

We recently used RNA-seq to examine symbiont physiology within the digestive tract of the medicinal leech, *Hirudo verbana* (Bomar *et al.*, 2011). The medicinal leech serves as a model for naturally occurring digestive-tract symbioses (Graf *et al.*, 2006; Nelson and Graf, 2012). In contrast to most mammalian digestive tracts, the crop, the largest compartment of the leech gut, is colonized by two dominant symbionts. One of the symbionts, *Aeromonas veronii*, is a gram-negative, facultative anaerobe that is often cultured from aquatic environments and is also a human pathogen (Janda and Abbott, 2010). Interestingly, our

initial RNA-seq study suggested that the *A. veronii* population in the crop was not homogeneous but exhibited at least two distinct physiologies at the same time: one of free-swimming or pelagic cells and one in mixed-species microcolonies with the second dominant symbiont, a recently cultured *Rikenella*-like bacterium (Worthen *et al.*, 2006; Kikuchi and Graf, 2007; Bomar *et al.*, 2011).

RNA-seq studies can also reveal the presence of previously unrecognized noncoding RNAs that can be highly abundant in bacterial cells (Shi *et al.*, 2009). One such system is the CsrB family that is involved in global gene regulation. The CsrB family contains regulatory RNAs that bind to and sequester CsrA, an RNA binding protein that post-transcriptionally regulates gene expression (Babitzke and Romeo, 2007). Binding of CsrA by CsrB is achieved through a series of GGA repeats (Babitzke and Romeo, 2007). In various bacterial genera, the Csr system regulates carbon metabolism, biofilm formation, secondary metabolite production, quorum sensing, and stationary-phase gene expression (Babitzke and Romeo, 2007).

In the current study, our goal was to gain a better understanding of the differences in *A. veronii*'s physiology *in vitro* compared to *in vivo*. We used RNA-seq to gain insight into *A. veronii*'s physiology during exponential growth in a rich medium and compared it to the physiology inside the leech 42 h after ingestion of a blood meal. The data suggest that CsrB is highly expressed, *Aeromonas* degrades arginine, and bacteria experience extracytoplasmic stress.

Materials and Methods

Sample collection for harvesting RNA

Leech crop samples—Two to four *Hirudo verbana* specimens (Leeches USA, Westbury, NY) per time point were fed heparinized sheep blood (Quad Five, Ryegate, MT) and maintained at room temperature for 2 h, 4 h, 8 h, 24 h, 42 h, 96 h, or 14 d. Leeches were dipped in 70% ethanol and sacrificed; 50-μl aliquots of the intraluminal fluid (ILF) were then harvested and flash-frozen in liquid nitrogen. *In vitro* grown *A. veronii* Hm21, an *A. veronii* strain isolated from the medicinal leech, was cultivated in lysogeny broth, LB, at 30 °C at 200 rpm (Sambrook and Russell, 2001). Aliquots of cells representing different phases of growth were collected and flash-frozen in liquid nitrogen. Samples were stored at −80 °C until further use.

In vivo growth curve of native *Aeromonas*

Three to eight leeches per time point were fed heparinized sheep blood and maintained at room temperature for 2 h, 4 h, 8 h, 24 h, 42 h, or 96 h. Leeches were sacrificed as described above, and the ILF was harvested. ILF was serially diluted in 0.85% (wt/vol) NaCl and plated on LB agar. Plates were incubated at 30 °C overnight, and colonies were enumerated to determine the level of native *Aeromonas* colonization in the crop.

RNA extraction

Total nucleic acid was extracted from ILF or *in vitro* grown Hm21 using the MasterPure RNA purification kit (Epicentre, Madison, WI). For the ILF samples, the whole-blood protocol was followed except that samples were treated with 50 μg of proteinase K in 300 μl of tissue and cell lysis solution prior to nucleic acid precipitation. For the *in vitro* samples, the protocol for cell samples was followed. DNA was removed using the Turbo DNA-*free* kit (Ambion, Austin, TX), and DNA contamination was tested for using *Aeromonas*-specific primers RpoBF1/RpoBR1 and the DNAse-treated RNA as template, as described previously (Bomar *et al.*, 2011).

cDNA library preparation and sequencing

Total RNA was quantified with the NanoDrop ND-1000 (Thermo Scientific, Wilmington, DE), and integrity was analyzed using an RNA FlashGel (Lonza, Basel, Switzerland). Ten micrograms of equally pooled, total RNA from the leech and 10 μg of total RNA from the *in vitro* samples were enriched for mRNA using mRNA-only (Epicentre, Madison, WI). Enriched RNA was quantified using RiboGreen (Invitrogen, Carlsbad, CA). One hundred nanograms each of the *in vitro* and *in vivo* enriched RNA were used for library preparations using SuperScript II (Invitrogen, Carlsbad, CA) and an mRNA-seq kit (Illumina, San Diego, CA) following the manufacturer's instructions, except that the poly(T) bead enrichment step was omitted. The cDNA libraries were submitted to the Translational Genomics Core at the University of Connecticut Health Center for cluster generation and 1 × 76 bp sequencing.

Data analysis

cDNA reads were mapped against the draft genome of *A. veronii* strain HM21 using CLC Genomics Workbench (ver. 4.0.2; Aarhus, Denmark). Reads that mapped to more than one region of the draft genome or mapped with more than two mismatches to the draft genome were excluded from the analysis. An expression value (EV) was calculated for each gene using the following formula: EV = [(number of reads mapped to the gene)/(length of the gene in kilobases) × (total number of reads mapped in millions)].

Nucleotide sequence alignment of CsrB sequences

Nucleotide sequences representative of the CsrB family were aligned in Geneious Pro (ver. 5.5.6; Auckland, New Zealand) using MUSCLE (Edgar, 2004).

Colony PCR screen for *csrB* and *csrC*

The primer pairs CsrBF1/CsrBR1 and CsrCF1/CsrCR1 were used for the amplification of *csrB* and *csrC*, respectively. Primer sequences are listed in Table 1 and the following amplification protocol was used: (i) 120 s at 95 °C; (ii) 30 cycles of 30 s at 95 °C, 30 s at 50 °C, 30 s at 72 °C; (iii) 120 s at 72 °C. Reactions contained 1× GoTaq Green MasterMix (Promega, Madison, WI), 1 μmol l^{-1} each of the appropriate forward and reverse primers, and 1 μl of a 1:20 colony dilution in a final reaction volume of 20 μl. The specificity of the primers was confirmed by sequencing the PCR amplicons generated using *A. veronii* wildtype genomic DNA as template, as previously described (Silver *et al.*, 2007).

Quantitative, reverse-transcription PCR

Double-stranded cDNA synthesis was carried out as described above using reagents from the Illumina mRNA-seq kit or New England Biolabs (Beverly, MA). Copy number of target genes was quantified using 1× SsoFast EvaGreen Supermix (Biorad, Hercules, CA), 500 nmol l^{-1} each of the appropriate forward and reverse primers, and 1 μl of a 10-fold dilution of cDNA in a final reaction volume of 12.5 μl. The following primer pairs were used for amplification of qPCR targets: CsrBF2/CsrBR1, CsrCF1/R1, RpoDF1/R1, RpoEF1/R1, PspA2F1/R1, ArcAF1/ArcAR1, and ArcBF1/ArcBR1. All primer sequences are listed in Table 1. The following amplification protocol was used for all primer pairs: (i) 30 s at 95 °C; (ii) 35 cycles of 0.5 s at 95 °C, 15 s at 50 °C. Primer specificity was confirmed by sequencing either amplicons generated in standard PCR using *A. veronii* genomic DNA as template or cloned amplicons. A melting curve analysis was also done where the dwell temperature increased from 50 °C to 95 °C in 0.5 °C increments every 10 s. The $2^{-\Delta\Delta C_T}$ method was used to calculate relative gene expression. Primer efficiency was evaluated using the standard curve method and ranged between 98% and 101%.

RNA-seq data accession numberss

Raw and processed RNA-seq data files were deposited in Gene Expression Omnibus at the NCBI under accession number GSE38068.

Results

Sequencing *Aeromonas veronii*'s transcriptome in culture and inside the leech gut

Genes that were expressed by *A. veronii* during exponential growth in a rich medium and inside the medicinal leech digestive tract were identified using Illumina RNA-seq. For *in vitro* growth of *A. veronii* we cultivated a medicinal leech isolate, Hm21 (Graf, 1999) under typical laboratory conditions in a rich medium (200 rpm at 30 °C in LB). Under these conditions, *A. veronii* doubles approximately every 30 min (data not shown). *A. veronii* was sampled in mid-log phase. For *in vivo* growth we sampled the leech intraluminal fluid 42 h after the animal had ingested a blood meal. At this time, *A. veronii* is still actively proliferating, although a portion of the population is entering stationary phase (Kikuchi and Graf, 2007; Bomar *et al.*, 2011). By this time the host has modified the blood meal by the removal of water and osmolytes (Sawyer, 1986), and hemocyte-like cells are patrolling the crop, phagocytosing sensitive bacteria (Silver *et al.*, 2007). We have previously examined only the leech crop microbiome 42 h after feeding using RNA-seq (Bomar *et al.*, 2011), using a different method to remove ribosomal RNA before library preparation. In the previous study, we utilized an oligonucleotide-based approach that was not very efficient (Table 2), and in the present study we utilized an enzymatic mRNA-enrichment method in an effort to increase our coverage of non-ribosomal transcripts.

The cDNA libraries were sequenced using an Illumina platform, generating 5,500,677 and 11,107,024 reads for the *in vitro* and *in vivo* libraries, respectively (Table 2). Approximately 201,558 of the *in vitro* library reads and 52,514 of the *in vivo* library reads mapped to non-ribosomal regions of Hm21's genome. For the *in vivo* library, the lower percentage of the reads mapping to the genome of Hm21 is largely due to reads mapping to the genome of the other dominant symbiont, the *Rikenella*-like bacterium. Overall, 95% and 98% of the reads were ribosomal for the *in vitro* and *in vivo* libraries, respectively, indicating that the mRNA-enrichment method we used was not very efficient. However, the sequencing depth of Illumina enabled us to overcome this issue because we were still able to detect cDNA reads that originated from protein-coding transcripts.

Carbon storage regulator (Csr) system

A powerful aspect of having both *in vitro* and *in vivo* data sets for *A. veronii* is the ability to identify genes that are differentially expressed between the two conditions. The most highly expressed gene *in vivo, csrB*, is expressed 65-fold lower *in vitro* and was originally annotated as a hypothetical protein in Hm21's draft genome (Table 3). Prior to discovering the identity of this gene, we queried the nucleotide sequence against the non-redundant GenBank database using BLASTN (Altschul *et al.*, 1990), and the closest match corresponded to a genomic region in *A. hydrophila* PPD134/91 (83% identity, 5×10^{-114} expected value, accession number AY378289). This region was identified during a genomic subtraction screen for virulence loci in *A. hydrophila* strains that used a non-virulent *A. hydrophila* strain to remove DNA of high similarity from the fish pathogen (Yu *et al.*, 2005). Our gene of interest, which is 469 bp in length, has 203 bp that overlap with an open reading frame (ORF) predicted to encode a 115 amino acid protein of unknown function. Because of its high expression and its implication in virulence, we investigated the highly expressed gene further.

Non-coding RNAs (ncRNAs), which are RNAs that do not encode for proteins, are often short in length and can be highly abundant (Shi *et al.*, 2009), similar to the highly expressed gene that we observed. Therefore, we queried the nucleotide sequence of the gene against the RFam database (Gardner *et al.*, 2011). On the basis of the results, this gene is predicted to be a CsrB family ncRNA (1.8×10^{-13} expected value), and thus we refer to this ncRNA as CsrB. In *A. veronii*'s CsrB we found 40 GGA repeats, 31 of which are preceded by an A residue or AN residues. This sequence structure is similar to that found in other CsrB family members. Bacterial genomes can possess functionally redundant CsrBs (Lenz *et al.*, 2005), so we queried Hm21's genome with CsrB's nucleotide sequence using BLASTN. This analysis revealed a 20-nucleotide stretch that was 100% identical to the query sequence. Further examination of the sequence flanking the alignment revealed a series of 16 GGA repeats, 13 of which are preceded by an A residue or AN residues. Based on the alignment with CsrB and other CsrB family members from the Rfam database and from *A. salmonicida* (Fig. 1), we annotated this newly identified, second ncRNA of the draft genome as CsrC. Similar to CsrB, CsrC expression is 50-fold higher during symbiosis than *in vitro* (Table 3). A BLASTN search of Hm21's genome using the nucleotide sequence of CsrC did not reveal additional CsrB family members, nor did querying the Rfam database (version 10.1) with the Hm21 genome. As this is only a draft genome, we cannot exclude the possibility that another CsrB homolog is present in the genome.

Because a genomic region containing *csrB* was detected during a screen for virulence loci in *A. hydrophila*, we were interested in the prevalence of *csrB* and *csrC* in other aeromonads. We screened 17 strains of *A. veronii*, including clinical and leech isolates, and 9 additional *Aeromonas* species for *csrB* and *csrC* using colony PCR (Table 4). The screen revealed the presence of PCR products for both genes in all the *A. veronii* isolates tested, including those that cannot colonize the medicinal leech to normal levels, suggesting that these genes do not account for the observed competitive dominance of leech isolates over strains isolated from other sources (Silver *et al.*, 2011). In our screen for *csrC*, amplicons of the expected size were detected in all of the nine *Aeromonas* species tested. *csrB* amplicons of the expected size were also detected for all nine *Aeromonas* species tested; however, amplification of *A. hydrophila* and *A. caviae* template produced an additional amplicon of slightly larger size. The PCR primers used for this screen were optimized using Hm21 DNA as template, and the additional amplicons could be due to nonspecific amplification of *A. hydrophila* and *A. caviae* template DNA. Overall, these particular ncRNAs appear to be widely conserved in the *Aeromonas* genus.

We validated the RNA-seq expression levels using qRT-PCR to quantify *csrB* and *csrC* expression over time *in vitro* and *in vivo*. We examined five time points *in vitro* that ranged from early exponential growth (OD_{600nm} 0.158) to late stationary phase (OD_{600nm} 5.34) (Fig. 2B). From early exponential growth to late stationary phase, *csrB* and *csrC* expression increased approximately 29-fold and 52-fold, respectively (Fig. 2A).

The expression of these genes was also measured *in vivo*. Because *csrB* and *csrC* expression has been reported to change depending on an organism's growth phase, it was important to determine the growth rates of the native *Aeromonas* community at the time points we sampled: 2 h, 4 h, 8 h, 24 h, 42 h, 96 h, and 14 d after feeding. Figure 2C shows a growth curve for native *Aeromonas* colonizing the leech gut at some of the corresponding sampling times. The data suggest that *A. veronii* is rapidly proliferating in the crop, perhaps doubling as fast as every 60 min at the earliest time points and slowing down to a doubling time of more than 250 min between 8 and 24 h after feeding. There are two important caveats to these measurements: the starting number of *A. veronii* in each animal is unknown, and the animal removes water from the crop to concentrate the blood, which accounts for about a 50% weight loss during the first 48 h after consuming a blood meal. These data are

consistent with a doubling time of 70 min during the first 12 h that we determined previously with an *A. veronii* strain that was introduced with a blood meal (Graf, 1999). The expression of *csrB* and *csrC* is below the limit of detection 2 h after feeding, but expression is detected at 4 h after feeding (Fig. 2A). Surprisingly, *csrB* and *csrC* expression is high 8 h after feeding (Fig. 2A), a time when the bacteria are still increasing in cell number (Fig. 2C). The observed expression levels are comparable to what was measured in late stationary phase under laboratory conditions when *A. veronii* is no longer proliferating. From 8 h onward, the expression of these two genes remains high.

Aeromonas veronii utilizes arginine as a nutrient source within the crop

An important question in all host-microbe associations is "what is the nutrient source for the microbiome?" In our previous report, we provided evidence for *A. veronii*'s *in vivo* catabolism of acetate and fatty acids *via* the glyoxylate shunt of the citric acid cycle and for the utilization of carbohydrates *via* the Embden-Meyerhoff Parnas pathway (Bomar *et al.*, 2011). Here, we present evidence that *A. veronii* also utilizes arginine *in vivo*. This hypothesis is supported by expression of *arcA, arcB, arcC*, and *arcD*, genes predicted to encode an arginine deiminase, an ornithine carbamoyltransferase, a carbamate kinase, and an arginine/ornithine antiporter, respectively (Table 3). qRT-PCR was used to verify expression of select arginine catabolism-related genes in mid-log phase (OD_{600nm} = 0.301) and in the crop at 42 h after feeding (Fig 3). As expected, the expression of *arcA* and *arcB* was elevated in the crop relative to *in vitro*, suggesting that arginine is utilized as a nutrient source *in vivo*.

A. veronii experiences extracytoplasmic stress inside the leech digestive tract

Another interesting feature revealed by the transcriptome data is that several stress-response-related genes are highly expressed during symbiosis and are expressed at a lower level or not detected during growth *in vitro* (Table 3). Most notable is the elevated expression of the alternative sigma factor, *rpoE, in vivo*. Bacteria use alternative sigma factors to direct the RNA polymerase to bind different sets of promoters, allowing them to adapt to changing environmental conditions. In bacteria, *rpoE* expression is induced when cells experience extracytoplasmic stress. This stress can be the result of many factors including changes in osmolarity, high temperatures, or exposure to antimicrobial compounds or reactive oxygen species (Rowley *et al.*, 2006). *In vivo, rpoE* expression is 7.2-fold higher than *in vitro*. In addition, *in vivo rpoE* expression is 6.3-fold greater than the expression of the housekeeping gene sigma factor *rpoD in vivo*. For comparison, *in vitro rpoE* expression is about 1.3-fold lower than *in vitro rpoD* expression. Phage shock genes (*pspA, pspB*, and *pspC*) are also highly expressed in the crop relative to *in vitro* (Table 3). Similar to *rpoE*, the *psp* genes have been implicated in the extracytoplasmic stress response. Depending on the microorganism and the environment, *pspA* gene expression can be induced by a number of stimuli including osmotic stress, high temperatures, loss of proton motive force, exposure to organic solvents, stationary phase growth, and filamentous phage infection (Rowley *et al.*, 2006). Again, we used qRT-PCR to verify the expression of select stress-response-related genes, *pspA2* and *rpoE* (Fig. 3). *pspA2* and *rpoE* were expressed about 7087-fold and 62-fold higher *in vivo*, validating our transcriptome findings that these genes are upregulated in the leech gut.

A comparison of *in vivo* data sets

We compared the current *in vivo* data set to a previously published data set from the same time point, in which an alternative method for mRNA enrichment was used (Bomar *et al.*, 2011). For the comparison, we calculated gene expression values either including or excluding the highly expressed ncRNAs CsrB and CsrC (Table 5). The reason for this was that between 24.5% and 38.5% of reads mapped to the *csrB* and *csrC* genes, and a small

difference in the number of reads mapping to those genes between the two samples would disproportionally affect the expression values of all other genes in the data set because the expression value is normalized to the total number of reads that mapped to the reference genome (Table 5). In the data set generated using the oligonucleotide method for rRNA removal, *csrB* and *csrC* accounted for a higher proportion of the reads; therefore the gene expression values in this data set are typically lower, potentially skewing the data set comparisons. In both analyses the expression values for most of the genes differ in expression by less than 2-fold (Table 5). The discrepancies that do exist between the two data sets could be due to animal-animal variation, differences in library coverage, or the different methods used to remove ribosomal RNA. Overall, the comparison of the two data sets suggests that the gene expression values obtained are consistent and independent of the rRNA removal method.

Discussion

In this study, we used high-throughput sequencing of RNA transcripts to characterize the transcriptomes of *Aeromonas veronii* during exponential growth in a rich medium and inside the digestive tract of the medicinal leech. Our goal was to gain further insight into the physiology of *A. veronii* within the host by identifying genes that are upregulated during symbiosis. The availability of both *in vitro* and *in vivo* data sets helped us determine the microbe's physiological state inside the host.

Utilizing both data sets, we identified two previously unannotated ncRNAs, *csrB* and *csrC*, that are highly expressed during symbiosis and appear to be conserved among aeromonads. We verified the expression of these genes *in vitro* and *in vivo* over time and found that their expression increases as cells approach stationary phase. This is consistent with previous reports that demonstrate that *csrB* and *csrC* expression is induced as cells approach stationary phase (Babitze and Romeo, 2007). Because of the large number of genes that have been reported to be regulated by the Csr system, it is difficult to predict the functional role of these ncRNAs in *A. veronii* during gut colonization. However, the Csr system has been reported to play a role in host colonization in a range of microorganisms (Lucchetti-Miganeh *et al.*, 2008), and an interesting future direction will be to investigate the role of that system in *A. veronii* colonization of the medicinal leech gut by constructing mutants.

The transcriptome data also provided new insight into the microorganism's nutrient sources within the crop. Data indicated that arginine is being utilized *in vivo*. We verified the expression of select genes involved in the catabolism of arginine and found that expression of *arcA* and *arcB* was elevated in the crop. Surprisingly, we detected *in vitro* expression of *arcA* and *arcB*, though the levels were close to our limit of detection. This was in contrast to the transcriptome data in which no expression was detected, suggesting that qRT-PCR with gene-specific primers is more sensitive than our transcriptome data. This difference could be due to the different priming method (random hexamers) used in the cDNA synthesis reactions for the transcriptome libraries or to an insufficient sequencing depth. The degradation of arginine through the combined activities of ArcA, ArcB, and ArcC is known as the arginine deiminase pathway and results in the production of ATP, ammonia, carbon dioxide, and ornithine, which is excreted through ArcD down a concentration gradient (Cunin *et al.*, 1986). Arginine is a poor nutrient source, and its catabolism results in one net mole of ATP per mole of arginine catabolized. The utilization of this nutrient source suggests that the *A. veronii* population has depleted more energetically favorable nutrients and could explain why the population begins to level off (Kikuchi and Graf 2007; Bomar *et al.*, 2011). In addition, degradation of arginine is induced during anaerobiosis (Cunin *et al.*, 1986), suggesting that there are anoxic conditions in the crop. The release of ammonia could also raise the pH of the intraluminal fluid, counteracting the acidification during

fermentation. Several microbes have been found to use arginine as a growth substrate in the laboratory, and there is evidence for microbes utilizing arginine during host association (D'Hooghe *et al.*, 1997; Fulde *et al.*, 2011). Given the nutrient complexity of the blood meal and the provision of mucin by the leech epithelium, it is not surprising that *A. veronii* is able to use a variety of nutrient sources in the leech gut for growth or maintenance. One would expect the symbiont to have redundant mechanisms for nutrient utilization. A temporal analysis of nutrient utilization inside the animal would provide interesting insights into the nutritional preferences *in vivo*.

A striking feature of the transcriptome data was the upregulation of stress-response-related genes *in vivo*, particularly those indicative of extracytoplasmic stress, such as *rpoE* and the *psp* genes. We verified the expression of *pspA2* and *rpoE* in both *in vitro* and *in vivo* samples and found that expression of both genes was elevated in the crop. In contrast to the transcriptome data, these data showed *in vitro* expression of *pspA2*, suggesting that qRT-PCR with gene-specific primers can detect expression of genes missed in the RNA-seq data. The importance of *rpoE* for host colonization has been reported in other animal models. For example, *Xenorhabdus nematophila rpoE* mutants have a decreased ability to colonize the intestinal vesicle of the entomopathogenic nematode (Heungens *et al.*, 2002). *rpoE* is also required for pathogenic microbes such as *Vibrio cholerae* (Kovacikova and Skorupski, 2002) and *Salmonella typhimurium* to colonize mice (Humphreys *et al.*, 1999). The extracytoplasmic stress experienced by *A. veronii* could be due to a number of factors including osmotic stress. The osmotic pressure of mammalian blood is approximately 300 mosm/l (Sawyer, 1986). Already during feeding, the leech removes water and osmolytes from the blood meal. By 2 d after feeding, the ingested blood meal is isosmotic with the hemolymph, which is about 200 mosm/l (Sawyer, 1986; Zebe *et al.*, 1986). This modification would expose the resident crop bacteria to changes in osmolarity, which likely causes osmotic stress for the bacteria. Another possible cause for the extracytoplasmic stress could be host antimicrobial peptides or antimicrobial proteins such as bactericidal permeability-increasing proteins, which often function by disrupting the bacterial cell membrane. The role of *rpoE* in resistance to antimicrobial peptides, such as polymixin B, has been demonstrated in *S. typhimurium in vitro* (Humphreys *et al.*, 1999). Antimicrobial peptides have been detected in the medicinal leech nervous system (Schikorski *et al.*, 2008) and the intestinal epithelium of a distantly related leech species, *Theromyzon tessulatum* (Tasiemski *et al.*, 2004). It seems likely that there are also antimicrobials in the medicinal leech gut, and that these could induce extracytoplasmic stress in *A. veronii*. A temporal transcriptome study of host and symbiont genes would provide important insight into these interactions.

Direct RNA sequencing techniques, such as RNA-seq, have been slow to be used in bacteria as compared to in eukaryotes because of the lack of a poly-A tail, which greatly hinders the enrichment of mRNA because it accounts for only a small percentage of the total RNA pool, with rRNA making up over 90% of the pool (Neidhardt, 1996). While there are enzymatic- and oligonucleotide-based approaches that remove rRNA, they are not always very efficient, and this can depend on the species being used and also the source of the sample. Some reports observed varying enrichment efficiencies between replicate samples (Yoder-Himes *et al.*, 2009; Stewart *et al.*, 2010). However, the increased throughput of next-generation sequencing platforms such as Illumina, which currently generates hundreds of a millions of reads in a single run, enables researchers to sequence through the ribosomal RNA and detect the more informative mRNA transcripts.

In addition, Illumina RNA-seq makes it feasible for researchers to study microbiomes that are in close association with a host. While there have been only a few published reports of RNA-seq being used to investigate host-associated microbiomes (Turnbaugh *et al.*, 2009,

2010; Stewart *et al.*, 2011; Bomar *et al.*, 2011), we anticipate that the recent advancements in RNA-seq will facilitate a greater use of this technique for symbioses. For instance, the high throughput enables the detection of transcripts from the microbiome despite contaminating host RNA and the presence of a mixed-species community. In addition, the ability to multiplex samples, where multiple samples are sequenced on the same flow cell lane, makes RNA-seq a cost-effective tool for transcriptomics. The application of this technology to *in vivo* and *in vitro* microbiomes, when possible, allows researchers to identify the physiology of a microorganism that is unique to its symbiotic state and can aid in determining the contribution of the microbiome to its host.

Acknowledgments

We thank M. Nelson, A. Tomasi, and E. McClure for helpful comments on the manuscript. This research was supported by NSF Career Award MCB 0448052, NIH RO1 GM095390, and UCONN Research Foundation award J980 to J.G.

Literature Cited

Altschul S, Gish W, Miller W, Myers E, Lipman D. Basic local alignment search tool. J. Mol. Biol. 1990; 215:403–410. [PubMed: 2231712]

Babitzke P, Romeo T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. Curr. Opin. Microbiol. 2007; 10:156–163. [PubMed: 17383221]

Bomar L, Maltz M, Colston S, Graf J. Directed culturing of microorganisms using metatranscriptomics. mBio. 2011; 2:e00012–e00011. [PubMed: 21467263]

Cunin R, Glansdorff N, Piérard A, Stalon V. Biosynthesis and metabolism of arginine in bacteria. Microbiol. Mol. Biol. Rev. 1986; 350:314–352.

Dethlefsen L, McFall-Ngai M, Relman DA. An ecological and evolutionary perspective on human-microbe mutualism and disease. Nature. 2007; 449:811–818. [PubMed: 17943117]

D'Hooghe I, VanderWauven C, Michiels J, Tricot C, de Wilde P, Vanderleyden J, Stalon V. The arginine deiminase pathway in *Rhizobium etli:* DNA sequence analysis and functional study of the *arcABC* genes. J. Bacteriol. 1997; 179:7403–7409. [PubMed: 9393705]

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32:1792–1797. [PubMed: 15034147]

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW, DeLong EF. Microbial community gene expression in ocean surface waters. Proc. Natl. Acad. Sci. USA. 2008; 105:3805–3810. [PubMed: 18316740]

Fulde M, Willenborg J, de Greeff A, Benga L, Smith HE, Valentin-Weigand P, Goethe R. ArgR is an essential local transcriptional regulator of the arcABC operon in *Streptococcus suis* and is crucial for biological fitness in an acidic environment. Microbiology. 2011; 157:572–582. [PubMed: 20947575]

Gardner PP, Daub J, Tate J, Moore BL, Osuch IH, Griffiths-Jones S, Finn RD, Nawrocki EP, Kolbe DL, Eddy SR, et al. Rfam: Wikipedia, clans and the "decimal" release. Nucleic Acids Res. 2011; 39(Suppl 1):D141–D145. [PubMed: 21062808]

Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P, Joint I. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PloS One. 2008; 3:e3042. [PubMed: 18725995]

Graf J. Symbiosis of *Aeromonas veronii* biovar sobria and *Hirudo medicinalis*, the medicinal leech: a novel model for digestive-tract associations. Infect. Immun. 1999; 67:1–7. [PubMed: 9864188]

Graf J, Kikuchi Y, Rio RVM. Leeches and their microbiota: naturally simple symbiosis models. Trends Microbiol. 2006; 14:365–371. [PubMed: 16843660]

Heungens K, Cowles CE, Goodrich-Blair H. Identification of *Xenorhabdus nematophila* genes required for mutualistic colonization of *Steinernema carpocapsae* nematodes. Mol. Microbiol. 2002; 45:1337–1353. [PubMed: 12207701]

Humphreys S, Stevenson A, Bacon A, Weinhardt AB, Roberts M. The alternative sigma factor, σE, is critically important for the virulence of *Salmonella typhimurium*. Infect. Immun. 1999; 67:1560–1568. [PubMed: 10084987]

Janda JM, Abbott SL. The genus *Aeromonas*: taxonomy, pathogenicity, and infection. Clin. Microbiol. Rev. 2010; 23:35–73. [PubMed: 20065325]

Kikuchi Y, Graf J. Spatial and temporal population dynamics of a naturally occurring two-species microbial community inside the digestive tract of the medicinal leech. Appl. Environ. Microbiol. 2007; 73:1984–1991. [PubMed: 17277211]

Kovacikova G, Skorupski K. The alternative sigma factor sigma(E) plays an important role in intestinal survival and virulence in *Vibrio cholerae*. Infect. Immun. 2002; 70:5355–5362. [PubMed: 12228259]

Lenz DH, Miller MB, Zhu J, Kulkarni RV, Bassler BL. CsrA and three redundant small RNAs regulate quorum sensing in *Vibrio cholerae*. Mol. Microbiol. 2005; 58:1186–1202. [PubMed: 16262799]

Lucchetti-Miganeh C, Burrowes E, Baysse C, Ermel G. The post-transcriptional regulator CsrA plays a central role in the adaptation of bacterial pathogens to different stages of infection in animal hosts. Microbiology. 2008; 154:16–29. [PubMed: 18174122]

Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs. Roche 454 Sequencing Technologies on the same microbial community DNA sample. PloS One. 2012; 7(2):e30087. [PubMed: 22347999]

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y, Chen Z, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437:376–380. [PubMed: 16056220]

Martens EC, Chiang HC, Gordon JI. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe. 2008; 4:447–457. [PubMed: 18996345]

Neidhardt, FC., editor. *Escherichia coli* and *Salmonella*. Washington, DC: ASM Press; 1996.

Nelson MC, Graf J. Bacterial symbioses of the medicinal leech *Hirudo verbana*. Gut Microbes. 2012; 3(4)

Noda S, Ohkuma M, Usami R, Horikoshi K, Kudo T. Culture-independent characterization of a gene responsible for nitrogen fixation in the symbiotic microbial community in the gut of the termite *Neotermes koshunensis*. Appl. Environ. Microbiol. 1999; 65:4935–4942. [PubMed: 10543806]

Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, Hourdez S, Girguis PR, Wankel SD, Barbe V, et al. Hydrogen is an energy source for hydrothermal vent symbioses. Nature. 2011; 476:176–180. [PubMed: 21833083]

Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP, Moran MA. Comparative day/night metatranscriptomic analysis of microbial communities in the north pacific subtropical gyre. Environ. Microbiol. 2009; 11:1358–1375. [PubMed: 19207571]

Poroyko V, White JR, Wang M, Donovan S, Alverdy J, Liu DC, Morowitz MJ. Gut microbial gene expression in mother-fed and formula-fed piglets. PloS One. 2010; 5:e12459. [PubMed: 20805981]

Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, et al. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *Escherichia coli* commensal and pathogenic isolates. J. Bacteriol. 2008; 190:6881–6893. [PubMed: 18676672]

Rowley G, Spector M, Kormanec J, Roberts M. Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens. Nat. Rev. Microbiol. 2006; 4:383–394. [PubMed: 16715050]

Sambrook, J.; Russell, DW. Molecular cloning: A laboratory manual. Cold Spring Harbor New York: Cold Spring Harbor Press; 2001.

Sawyer, R. Leech Biology and Behavior. Oxford: Oxford University Press; 1986.

Schikorski D, Cuvillier-Hot V, Leippe M, Boidin-Wichlacz C, Slomianny C, Macagno E, Salzet M, Tasiemski A. Microbial challenge promotes the regenerative process of the injured central nervous system of the medicinal leech by inducing the synthesis of antimicrobial peptides in neurons and microglia. J. Immunol. 2008; 181:1083–1095. [PubMed: 18606660]

Shi Y, Tyson GW, DeLong EF. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. Nature. 2009; 459:266–269. [PubMed: 19444216]

Silver AC, Kikuchi Y, Fadl AA, Sha J, Chopra AK, Graf J. Interaction between innate immune cells and a bacterial type III secretion system in mutualistic and pathogenic associations. Proc. Natl. Acad. Sci. USA. 2007; 104:9481–9486. [PubMed: 17517651]

Silver AC, Williams D, Faucher J, Horneman AJ, Gogarten JP, Graf J. Complex evolutionary history of the *Aeromonas veronii* group revealed by host interaction and DNA sequence data. PloS One. 2011; 6:e16751. [PubMed: 21359176]

Sonnenburg JL, Xu J, Leip DD, Chen C, Westover BP, Weatherford J, Buhler JD, Gordon JI. Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. Science. 2005; 307:1955–1959. [PubMed: 15790854]

Stewart FJ, Ottesen EA, DeLong EF. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. ISME J. 2010; 4:896–907. [PubMed: 20220791]

Stewart FJ, Dmytrenko O, DeLong EF, Cavanaugh CM. Metatranscriptomic analysis of sulfur oxidation genes in the endosymbiont of *Solemya velum*. Front. Microbiol. 2011; 2:134. [PubMed: 21738524]

Tartar A, Wheeler MM, Zhou X, Coy MR, Boucias DG, Scharf ME. Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. Biotechnol. Biofuels. 2009; 2:25. [PubMed: 19832970]

Tasiemski A, Vandenbulcke F, Mitta G, Lemoine J, Lefebvre C, Sautiere P, Salzet M. Molecular characterization of two novel antibacterial peptides inducible upon bacterial challenge in an annelid: the leech *Theromyzon tessulatum*. J. Biol. Chem. 2004; 279:30973–30982. [PubMed: 15102860]

Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci. Transl. Med. 2009; 1 6ra14.

Turnbaugh PJ, Quince C, Faith JJ, McHardy AC, Yatsunenko T, Niazi F, Affourtit J, Egholm M, Henrissat B, Knight R, et al. Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. Proc. Natl. Acad. Sci. USA. 2010; 107:7503–7508. [PubMed: 20363958]

Urich T, Lanzen A, Qi J, Huson DH, Schleper C, Schuster SC. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PloS One. 2008; 3:e2527. [PubMed: 18575584]

Wier AM, Nyholm SV, Mandel MJ, Prisca Massengo-Tiassé R, Schaefer AL, Koroleva I, Splinter-Bondurant S, Brown B, Manzella L, Snir E, et al. Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical and metabolic change in a beneficial symbiosis. Proc. Natl. Acad. Sci. USA. 2010; 107:2259–2264. [PubMed: 20133870]

Worthen PL, Gode CJ, Graf J. Culture-independent characterization of the digestive-tract microbiota of the medicinal leech reveals a tripartite symbiosis. Appl. Environ. Microbiol. 2006; 72:4775–4781. [PubMed: 16820471]

Yoder-Himes DR, Chain PSG, Zhu Y, Wurtzel O, Rubin EM, Tiedje JM, Sorek R. Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. Proc. Natl. Acad. Sci. USA. 2009; 106:3976–3981. [PubMed: 19234113]

Yu HB, Zhang YL, Lau YL, Yao F, Vilches S, Merino S, Tomas JM, Howard SP, Leung KY. Identification and characterization of putative virulence genes and gene clusters in *Aeromonas hydrophila* PPD134/91. Appl. Environ. Microbiol. 2005; 71:4469–4477. [PubMed: 16085838]

Zebe E, Roters F-J, Kaiping B. Metabolic changes in the medicinal leech *Hirudo medicinalis* following feeding. Comp. Biochem. Physiol. 1986; 84A:49–55.

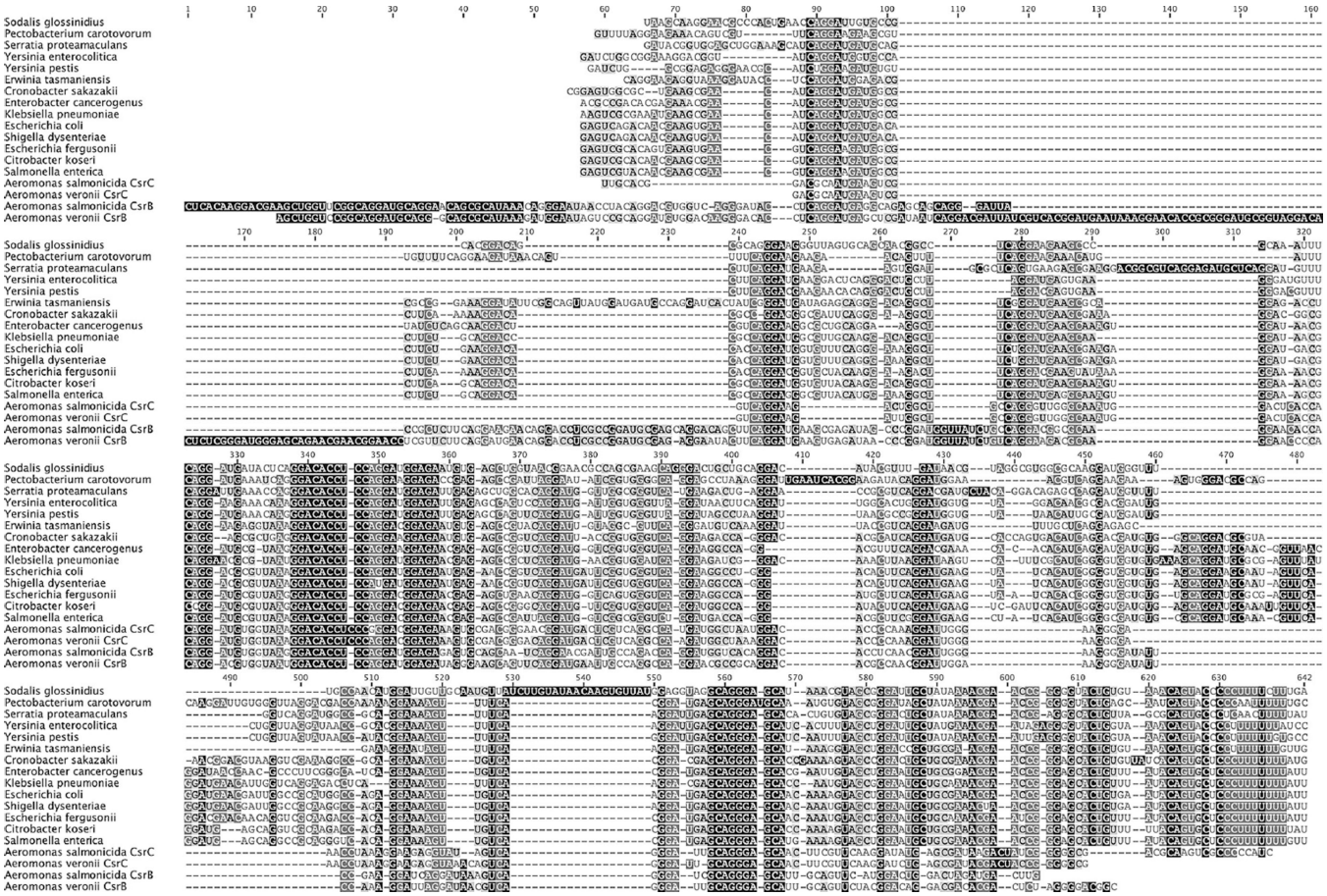

Figure 1.

Nucleotide sequence alignment of CsrB family members from the Rfam database, *Aeromonas salmonicida* and *A. veronii* Hm21. Note the presence of GGA repeats. Nucleotides are shaded in a range of black and gray according to similarity. Black indicates that all sequences have identical nucleotides in the position indicated. Gaps in the alignment are depicted as dashes. The following accession numbers and nucleotide positions for the sequences used in the alignment are as follows: *Yersinia enterocolitica* (AM286415/3595708-3595388), *Yersinia pestis* (AAYV01000020/39993-39672), *Pectobacterium carotovorum* (ABVY01000012/18391-18747), *Sodalis glossinidius* (AP008232/3356002-3355655), *Serratia proteamaculans* (CP000826/4199485-4199142), *Escherichia fergusonii* (CU928158/283863-284220), *Cronobacter sakazakii* (CP000783/487241-487598), *Enterobacter cancerogenus* (ABWM01000067/87302-86944), *Salmonella enterica* (CP000880/4597967-4598323), *Klebsiella pneumoniae* (CP000964/999311-999670), *Escherichia coli* (CP001164/3751312-3750953), *Shigella dysenteriae* (CP000034/2780478-2780119), *Citrobacter koseri* (CP000822/3859356-3858998), and *Erwinia tasmaniensis* (CU468135/3038137-3037797), *Aeromonas salmonicida* CsrB (NC_009348.1/4285409-4285802), *Aeromonas salmonicida* CsrC (NC_009348.1/232825-233093).

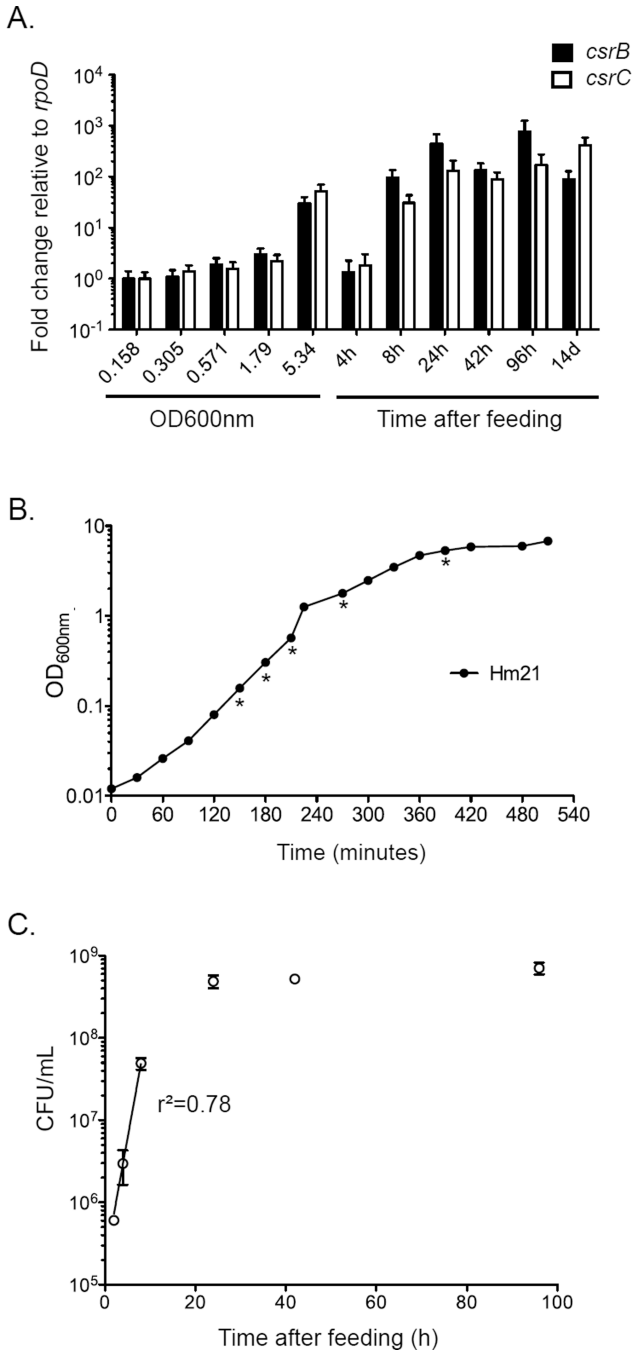

Figure 2.

csrB and *csrC* expression *in vitro* and *in vivo*. (A) Expression of *csrB* and *csrC, in vitro* and *in vivo*, was confirmed using quantitative reverse-transcription PCR (qRT-PCR). Results are displayed as fold change in expression relative to *rpoD* expression and normalized to expression in early exponential growth (OD_{600nm} = 0.158). Each target was quantified in duplicate and the experiment was performed in triplicate. The error bars represent standard deviation. (B) Growth curve for Hm21 in lysogeny broth. An asterisk indicates a sampling time analyzed in qRT-PCR. (C) Representative growth curve for native *Aeromonas* in the leech crop, as determined by plating dilutions of the intraluminal fluid. The CFU/ml of *Aeromonas* in the crop was averaged from multiple animals and is represented by an open

circle. A regression analysis was performed for the time points between 2 h and 8 h. The equation of the line shown is $y = 0.306x + 5.24$ and was used to calculate the growth rate for this time period. Standard error is represented by the vertical bars.

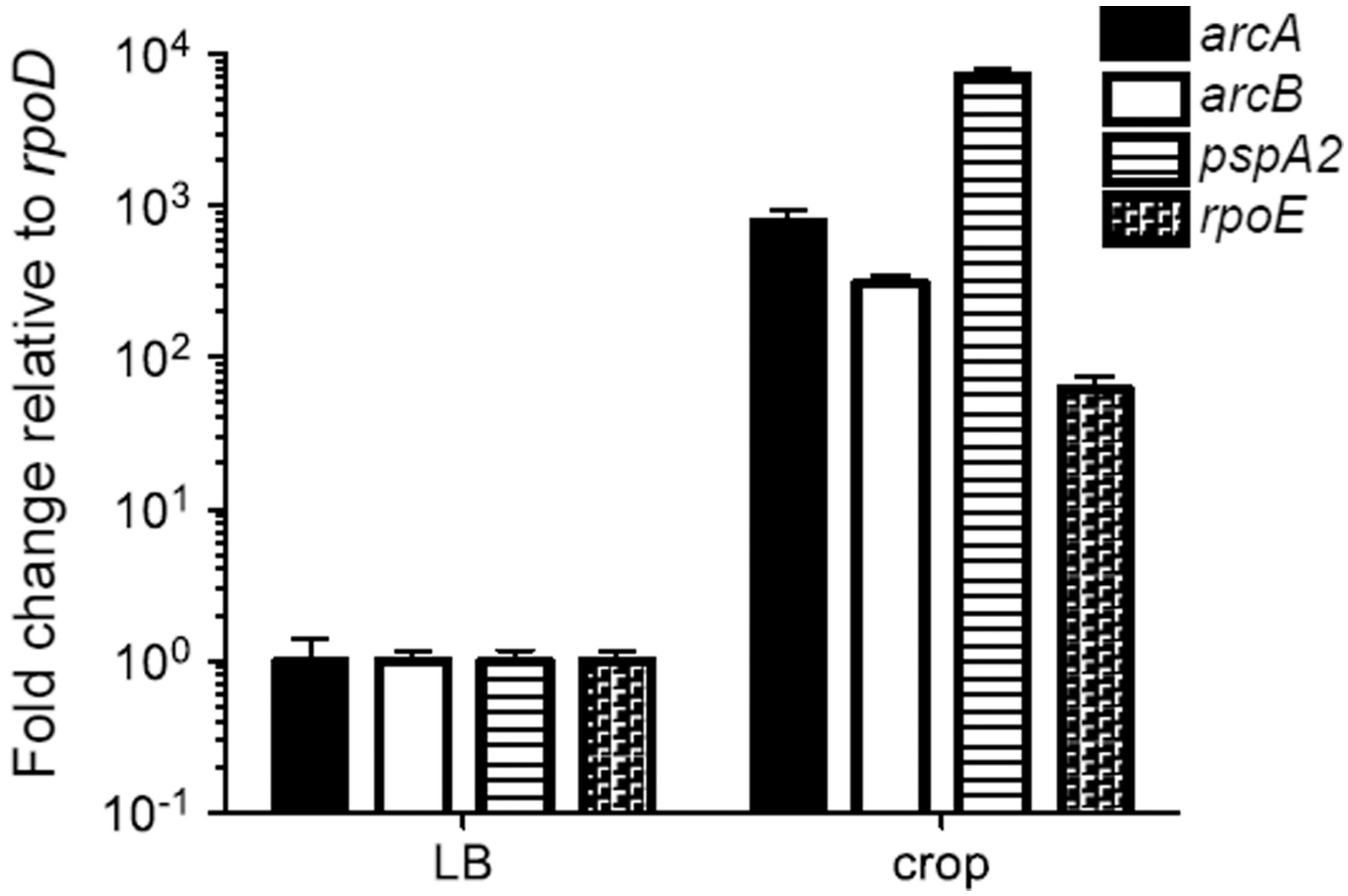

Figure 3.
Quantitative reverse-transcription PCR validation for arginine catabolism and stress-response related genes in lysogeny broth (LB) (OD_{600nm} = 0.301) and in the crop (42 h). Results are displayed as fold change in expression relative to *rpoD* expression and normalized to expression in LB. Each target was quantified in triplicate and the assay was performed twice. The error bars represent standard deviation.

Table 1

Primers used in this study

Primer	Sequence (5′ to 3′)
CsrBF1	GACGCAAGGAACACCACGGG
CsrBF2	GAATAAAGGAACACCGCA
CsrBR1	AATATCCCCTTTCCAATCC
CsrCF1	GGTCAGGAAGATTGGCTG
CsrCR1	TAGGTTTCCCCTTCCCAATC
RpoDF1	CACGCATATCGGTTCCGAGCTC
RpoDR1	TCGCGAGCAACTTCCGGATC
RpoEF1	TAAAAGCCTACCGTGCAT
RpoER1	CTTTTTATCGATTGCTTCCC
PspA2F1	ATTAATGCCAACCTGCAC
PspA2R1	CATTTTTTGACCTCCGCC
ArcAF1	TTTCCTGCTGACCAATCT
ArcAR1	TGGTGGAGTTGTCGTAGT
ArcBF1	TATCGAATACCGCGGCTA
ArcBR1	CACACCCTCTTTCACATC

Table 2

Summary of RNA-seq datasets and mRNA enrichment efficiency

Sample	Enrichment method	No. reads	Total no. reads mapped	No. reads mapped to 23S (%)	No. reads mapped to 16S (%)	No. reads mapped to 5S (%)	No. reads mapped to non-rRNA	Reference
Crop	Oligonucleotide	15,108,840	6,934,692	4,641,909 (67)	2,045,615 (30)	25,310 (0.4)	246,388[a]	Bomar *et al.*, 2011
Crop	Enzymatic	11,107,024	3,449,889	2,354,304 (68)	1,043,099 (30)	780 (0.02)	52,514	This study
Lysogeny broth	Enzymatic	5,500,677	3,970,130	3,054,529 (77)	711,176 (18)	2,597 (0.07)	201,558	This study

[a]This value is different than the value previously reported (Bomar *et al.*, 2011): in the previous study we reported the number of reads that mapped to coding regions, but here we report the number of reads that map to noncoding regions as well.

Table 3

In vitro and *in vivo* expression values for select genes

Predicted product	Accession no.[a]	Expected value[b]	Expression Value (EV)		Fold change[d]
			Culture[c]	Leech	
Housekeeping[e]					
RpoD	YP_855378	0	728	628	−1.2
RpoB	YP_004394379.1	0	3472	595	−5.8
Regulatory					
CsrA	YP_004394252	6.83×10^{-37}	471	334	−1.4
CsrB[f]	—	—	9875	6.5×10^{5}	65.8
CsrC[f]	—	—	7947	4.0×10^{5}	50.1
Nutrition					
ArcA	YP_858517	0	0	2478	∞
ArcB	YP_858515	0	0	1160	∞
ArcC	YP_004394442	0	7	1632	233.1
ArcD	YP_001140165	0	5	1031	206.2
Stress-related					
RpoE	YP_004391428	8.00×10^{-134}	546	3954	7.2
PspA1	YP_004392261	5.88×10^{-151}	0	3303	∞
PspA2	YP_004391182	5.64×10^{-137}	0	4421	∞
PspB	YP_004392262	4.90×10^{-35}	27	1596	59.1
PspC	YP_004392263	4.00×10^{-91}	16	1634	102.1

[a] GenBank accession number for the top informative BLASTX hit at the NCBI.

[b] Expected value based on querying the nonredundant database at the NCBI using BLASTX.

[c] Expression value calculated as (number of reads mapped to the gene)/(length of the gene in kilobases) × (total number of reads mapped in millions).

[d] The change in gene expression from *in vitro* to *in vivo*. Calculated as (Leech EV/Culture EV). If the value is less than one, the negative inverse of the value is reported.

[e] Housekeeping genes were included in the table for comparison.

[f] Dash (—) in a column indicates that the gene of interest encodes an ncRNA and that BLASTX information is not relevant.

Table 4

Aeromonas strains used for *csrB* and *csrC* screen and PCR results

Species	Strain	Isolation source	PCR result (*csrB*/*csrC*)
Aeromonas veronii	AER397	Blood	+/+
A. veronii	AMC34	Human feces	+/+
A. veronii	AMC35	Wound	+/+
A. veronii	CDC0437-84	Fish	+/+
A. veronii	AER28	Feces	+/+
A. veronii	AER39	Blood	+/+
A. veronii	AMC22	Feces	+/+
A. veronii	AMC23	Finger wound	+/+
A. veronii	AMC24	Feces	+/+
A. veronii	AMC25	Duck	+/+
A. veronii	AMC26	Foot wound	+/+
A. veronii	HV221	*Hirudo verbana*	+/+
A. veronii	HV231	*H. verbana*	+/+
A. veronii	HV241	*H. verbana*	+/+
A. veronii	HM21	*H. verbana*	+/+
A. veronii	Ho635	*H. orientalis*	+/+
A. veronii	LMG13695	Feces	+/+
A. allosaccharophila	LMG140549T	Eel	+/+
A. hydrophila	ATCC7966T	Canned milk	#[a]/+
A. bestiarum	ATCC14715	Silver salmon	+/+
A. salmonicida	CDC434-84	Fresh water	+/+
A. caviae	ATCC15468	Guinea pig	#[a]/+
A. sobria	CIP7433	Fish	+/+
A. jandaei	ATCC49568T	Feces	+/+
A. encheleia	LMG16330T	Eel	+/+
A. trota	ATCC49657T	Feces	+/+

[a]Indicates that an additional PCR amplicon was generated using this template.

Table 5

Comparison of gene expression values from *in vivo* datasets generated using different mRNA enrichment methods

Predicted product	Analysis including ncRNAs			Analysis excluding ncRNAs		
	EV[a] Enzyme	EV Oligo	Fold change[b]	EV Enzyme	EV Oligo	Fold change
Housekeeping						
RpoD	628	311	−1.9	1,056	708	−1.5
RpoB	595	363	−1.6	1,000	826	−1.2
Rps19	6,214	3,045	−2.0	10,449	6,926	−1.5
Regulatory						
CsrA*	334	874	2.6	561	1,988	3.5
CsrB	6.5×10^5	8.1×10^5	1.3	–	–	–
CsrC	4.0×10^5	7.0×10^5	1.8	–	–	–
Nutrient utilization						
ArcA	2,478	1,391	−1.8	4,167	3,165	−1.3
ArcB	1,160	998	−1.2	1,951	2,270	1.2
ArcC	1,632	535	−3.1	2,744	1,216	−2.3
ArcD	1,031	373	−2.8	1733	849	−2.0
AceA	828	385	−2.1	1,393	876	−1.6
AceB*	137	66	−2.1	231	150	−1.5
Pfk*	195	61	−3.2	327	138	−2.4
Fbp*	152	154	1.0	256	350	1.4
SucB*	211	123	−1.7	356	282	−1.2
CcoN*	332	132	−2.5	558	301	−1.9
Stress-response						
RpoE	3,954	2,939	−1.3	6,648	6,686	1.0
PspA1	3,303	2,542	−1.3	5,555	5,781	1.0
PspA2	4,421	3,959	−1.1	7,435	9,006	1.2
PspB	1,596	1,319	−1.2	2,684	3,001	1.1
PspC	1,634	1,238	−1.3	2,748	2,816	1.0
RpoS	769	887	1.2	1,293	2,018	1.6

Predicted product	Analysis including ncRNAs			Analysis excluding ncRNAs		
	EV[a] Enzyme	EV Oligo	Fold change[b]	EV Enzyme	EV Oligo	Fold change
Stationary phase						
Rmf	13,834	17,861	1.3	23,262	40,628	1.7

[a] EV, expression value calculated as [(number of reads mapped to the gene)/(length of the gene in kilobases) × (total number of reads mapped in millions)].

[b] The change in gene expression between the enzymatically treated RNA-seq data set and the oligonucleotide-based treatment of the previously published RNA-seq data set. Calculated as (Crop EV Oligo/Crop EV Enzyme). If the value is less than one, the negative inverse of the value is reported.

* Indicates there is less than 1× coverage of this gene in one or both of the RNA-seq data sets.

Biol Bull. Author manuscript; available in PMC 2013 August 04.