# Directed Culturing of Microorganisms Using Metatranscriptomics

**Lindsey Bomar, Michele Maltz,* Sophie Colston, and Joerg Graf**

Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut, USA

* Present address: Yale School of Public Health, Department of Epidemiology and Microbial Diseases, New Haven, Connecticut, USA.

**ABSTRACT** **The vast majority of bacterial species remain uncultured, and this severely limits the investigation of their physiology, metabolic capabilities, and role in the environment. High-throughput sequencing of RNA transcripts (RNA-seq) allows the investigation of the diverse physiologies from uncultured microorganisms in their natural habitat. Here, we report the use of RNA-seq for characterizing the metatranscriptome of the simple gut microbiome from the medicinal leech *Hirudo verbana* and for utilizing this information to design a medium for cultivating members of the microbiome. Expression data suggested that a *Rikenella*-like bacterium, the most abundant but uncultured symbiont, forages on sulfated- and sialated-mucin glycans that are fermented, leading to the secretion of acetate. Histological stains were consistent with the presence of sulfated and sialated mucins along the crop epithelium. The second dominant symbiont, *Aeromonas veronii*, grows in two different microenvironments and is predicted to utilize either acetate or carbohydrates. Based on the metatranscriptome, a medium containing mucin was designed, which enabled the cultivation of the *Rikenella*-like bacterium. Metatranscriptomes shed light on microbial metabolism *in situ* and provide critical clues for directing the culturing of uncultured microorganisms. By choosing a condition under which the desired organism is rapidly proliferating and focusing on highly expressed genes encoding hydrolytic enzymes, binding proteins, and transporters, one can identify an organism's nutritional preferences and design a culture medium.**

**IMPORTANCE** **The number of prokaryotes on the planet has been estimated to exceed $10^{30}$ cells, and the overwhelming majority of them have evaded cultivation, making it difficult to investigate their ecological, medical, and industrial relevance. The application of transcriptomics based on high-throughput sequencing of RNA transcripts (RNA-seq) to microorganisms in their natural environment can provide investigators with insight into their physiologies under optimal growth conditions. We utilized RNA-seq to learn more about the uncultured and cultured symbionts that comprise the relatively simple digestive-tract microbiome of the medicinal leech. The expression data revealed highly expressed hydrolytic enzymes and transporters that provided critical clues for the design of a culture medium enabling the isolation of the previously uncultured *Rikenella*-like symbiont. This directed culturing method will greatly aid efforts aimed at understanding uncultured microorganisms, including beneficial symbionts, pathogens, and ecologically relevant microorganisms, by facilitating genome sequencing, physiological characterization, and genetic manipulation of the previously uncultured microbes.**

**Received** 14 January 2011 **Accepted** 8 March 2011 **Published** 5 April 2011

**Citation** Bomar L, Maltz M, Colston S, Graf J. 2011. Directed culturing of microorganisms by use of metatranscriptomics. mBio 2(2):e00012-11. doi:10.1128/mBio.00012-11.

**Editor** Jo Handelsman, Yale University



Address correspondence to Joerg Graf, joerg.graf@uconn.edu.

Most microorganisms have resisted cultivation in the laboratory despite, or perhaps because of, the tremendous diversity in physiological capabilities that microorganisms possess. It is commonly cited that 99 to 99.9% of microorganisms in an environment have not been grown in the laboratory (1). Even the most abundant and metabolically active members of the microbial community can be challenging to cultivate in the laboratory. Tailored cultivation techniques utilizing antibiotics, dilute nutrient conditions, and cocultivation with other organisms allowed the cultivation of previously "unculturable" bacteria (2–4). Sequencing the genomes of microbial communities, or metagenomics, has provided unprecedented insight into the metabolic potential of microorganisms (5, 6) but has rarely led to the cultivation of an uncultured microorganism (6), and overall metagenomics has not had a dramatic impact on improving cultivation. A better understanding of contemporaneous microbial physiology *in situ* could provide the information required to culture more microorganisms, particularly the organisms that excel in the environment examined. Next-generation sequencing is providing new opportunities for gaining such insight into the metabolic relationships within microbial communities by directly sequencing the metatranscriptomes (7–11). This advance allows a new approach for directing the design of culturing conditions based on the *in situ* physiology.

One environment that houses complex microbial communities is the digestive tract. This microbiome provides the host with essential nutrients, prevents the colonization of pathogens, stimulates the immune system, and aids in digestion (12, 13). The overarching goal of the human microbiome project is to identify the members of this community and elucidate the contribution of the microbiome to human health (14). The presence of hundreds of species and closely related strains represents a challenge for assembling large contigs or reconstructing entire genomes (5). This challenge can be overcome by the use of reference genomes

from cultured organisms (15). While a greater percentage of microbes from the human gut than from other environments has been cultured in the laboratory, many species remain uncultured (16).

The digestive-tract symbiosis of the medicinal leech *Hirudo verbana* is a simple model system where two symbionts, *Aeromonas veronii* and a *Rikenella*-like bacterium, dominate the crop microbiota (17, 18). The crop is the largest compartment of the digestive tract and stores ingested blood for months between feedings (19). Water and salts are removed from the ingested blood meal, creating a viscous intraluminal fluid (ILF) (19). *A. veronii*, a gammaproteobacterium, is readily cultivated and genetically manipulated (20, 21). The more abundant *Rikenella*-like symbionts are predicted to be obligate anaerobes and represent a novel genus related to *Rikenella microfusus*, a member of the *Bacteroidetes*. Different 16S rRNA ribotypes of the *Rikenella*-like bacterium have been found in *H. verbana*, ribotype PW3, and *Hirudo orientalis*, ribotype AL5CE1 (18, 22). This *Rikenella*-like symbiont has remained uncultured despite several isolation attempts. The simplicity of this tripartite association allows us to study the molecular mechanisms underlying host-microbe and microbe-microbe interactions in a naturally occurring digestive-tract symbiosis (20, 23) and develop new approaches. In order to gain a greater understanding of this digestive-tract association, we aimed to characterize the metatranscriptome of the crop symbionts *in vivo* using high-throughput sequencing of RNA transcripts (RNA-seq). Prior to this study, little was known about the physiology of *A. veronii* and the *Rikenella*-like bacterium within the crop. We predict that the *in vivo* physiological conditions would provide important clues for designing a medium, which would allow the cultivation of the *Rikenella*-like bacterium in the laboratory.

## RESULTS

**Sequencing the metatranscriptome of the leech gut microbiome.** We investigated the physiology of the bacterial symbionts by sequencing the crop metatranscriptome using Illumina RNA-seq (24). The great sequencing depth can overcome the challenges of abundant rRNA and allele, strain, and species variation in natural microbial communities. For sample collection, we selected 42 h after the leech consumed a blood meal to explore the metatranscriptome. At this time, the leech has absorbed osmolytes and water from the ingested blood meal (19) and *A. veronii* is close to reaching its maximum density while the *Rikenella*-like bacteria are still growing rapidly (25). Four leeches were sacrificed, and the ILF, containing the bacteria residing in the crop, was harvested. Total RNA was extracted from the ILF, and equal amounts were pooled from each leech. mRNA was enriched for using an rRNA pulldown approach. The mRNA-enriched sample was used to construct an Illumina cDNA library that was sequenced on two flow cells for 34 bp. Approximately 15 million cDNA reads were generated and mapped against an ~200× draft genome of Hm21 (see Table S1 in the supplemental material), an *A. veronii* strain isolated from the medicinal leech (17), and an ~20× draft of the *Rikenella*-like genome (Table S2), which is part of a crop metagenome. Approximately 4.5% of mapped reads aligned to protein coding sequences (CDS), suggesting an inefficient removal of rRNA. About 2.5 Mb (74,524 reads) and 16.4 Mb (481,160 reads) of sequence mapped to CDS of *A. veronii* and the *Rikenella*-like bacterium, respectively. That a higher number of reads mapped to the genome of the *Rikenella*-like bacterium is likely due to the facts that the *Rikenella*-like bacterium is more abundant than *A. veronii* in the crop and that it is actively proliferating, while at least a portion of *A. veronii* appears to enter stationary phase.

**The *Rikenella*-like bacterium forages host mucin glycans.** The *Rikenella*-like bacterium's transcriptome indicated that energy is obtained by fermenting sugars to acetate (Table 1; see also Table S2 in the supplemental material), which is secreted into the ILF. In the transcriptome, one operon stood out as being expressed at higher levels than even the ribosomal proteins (Fig. 1). One of the genes in this operon, *endoG*, is predicted to encode an endo-β-*N*-acetylglucosaminidase, a member of carbohydrate active enzyme glycosyl hydrolase family 18. This enzyme is predicted to target the glycosidic bond between adjacent *N*-acetylglucosamine residues that attach glycans to asparagine (26), thus liberating glycans from proteins. The other genes of this operon are predicted to encode a hypothetical protein with a domain of unknown function (family 1735) and SusC-like and SusD-like proteins. In *Bacteroides thetaiotaomicron*, SusC-like and SusD-like proteins are encoded by polysaccharide utilization loci and form an outer membrane complex that binds polysaccharides (27). Bound polysaccharides are hydrolyzed by glycosyl hydrolases, and monosaccharides enter the cell where they are metabolized. The very high expression levels of this polysaccharide utilization operon suggested that the acquisition of protein-bound glycans in the crop is critical for the *Rikenella*-like bacterium.

Genes encoding sulfated-mucin desulfatases (*smdS1* and *smdS2*) and sialidases were also expressed (Table 1; see also Table S2 in the supplemental material). Sulfated-mucin desulfatases remove sulfate from mucin glycans, making the sugar more accessible to bacterial degradation (28). These findings suggest that the polysaccharides being foraged by *Rikenella* were sulfated and sialated mucin glycans. In addition, the expression of sialidase genes also suggests that *Rikenella* could forage on sialated glycoproteins present on the surfaces of the ingested erythrocytes.

Expression of *endoG* and *smdS1* was verified using quantitative reverse transcription-PCR (qRT-PCR). Relative gene expression was measured 8 h, 24 h, 42 h, and 96 h after feeding using RNA pooled from 2 to 4 leeches per time point (Fig. 2). This experiment confirmed the expression of these two genes at 42 h after feeding and indicated that between 8 and 24 h after feeding, *endoG* and *smdS1* expression increased approximately 15-fold and 5-fold, respectively. *smdS1* expression was stable between 24 and 96 h after feeding. *endoG* expression increased an additional ~10-fold at 42 h after feeding but decreased by 96 h after feeding. By 96 h after feeding, the *Rikenella*-like symbiont may utilize an alternate carbon and energy source or the organism's overall metabolism may be diminishing as it approaches stationary phase.

**Histological stains detected sialated and sulfated glycans in leech tissue.** If mucins serve as a nutrient for *Rikenella*, histological stains should reveal the presence of glycans either in intracellular storage vacuoles or extracellularly along the crop epithelium, which would be consistent with the presence of mucins. We investigated the nature and temporal variability of sulfated and sialated glycans in a series of histochemical sequences of host tissue sections obtained 8 h, 24 h, 42 h, 96 h, and 7 days after feeding. The high-iron diamine (HID)/Alcian blue (AB) staining sequence (29) consistently demonstrated the presence of sulfomucins and sialomucins, respectively, on the luminal sides of the crop epithelial cells at all time points (Fig. 3A to E). The stained sections showed sialoglycans, marked by the distinctive blue stain, present in a

**TABLE 1** Expression data provide insight into symbionts' physiologies *in situ*

| Predicted gene product | GenBank accession no.[a] | Expected value[b] | Metabolic function | EV[c] |
|---|---|---|---|---|
| *Rikenella*-like bacterium | | | | |
| Endo-β-*N*-acetylglucosaminidase | ZP_06995334 | $2.06 \times 10^{-73}$ | Glycan foraging | 9,664 |
| SusC-like protein | NP_809953 | 0 | Glycan foraging | 9,014 |
| SusD-like protein | NP_809952 | $3.33 \times 10^{-173}$ | Glycan foraging | 10,871 |
| Sulfated-mucin desulfatase 1 | YP_098213 | $3.16 \times 10^{-89}$ | Mucin foraging | 1,868 |
| Sulfated-mucin desulfatase 2 | ZP_05416513 | $5.92 \times 10^{-78}$ | Mucin foraging | 391 |
| Sialidase 1 | ZP_05255794 | $2.70 \times 10^{-146}$ | Mucin foraging | 451 |
| Sialidase 2 | YP_001301367 | $1.02 \times 10^{-137}$ | Mucin foraging | 1,291 |
| Phosphate acetyltransferase | YP_001304694 | $2.29 \times 10^{-117}$ | Fermentation | 1,047 |
| Acetate kinase | CBK63862 | $4.81 \times 10^{-131}$ | Fermentation | 681 |
| Phosphofructokinase | YP_001740717 | $2.99 \times 10^{-121}$ | Glycolysis | 542 |
| Ribosomal protein S3 | CBK64159 | $1.38 \times 10^{-89}$ | Translation | 2,275 |
| *A. veronii* | | | | |
| Malate synthase | YP_001141353 | 0 | Glyoxylate cycle | 150 |
| Isocitrate lyase | YP_001141352 | 0 | Glyoxylate cycle | 876 |
| Phosphofructokinase | YP_856906 | $1.37 \times 10^{-180}$ | Glycolysis | 138 |
| Fructose-1,6 bisphosphatase | YP_001141703 | 0 | Gluconeogenesis | 350 |
| 2-Oxoglutarate dehydrogenase E2 | YP_856459 | 0 | TCA cycle | 282 |
| CcoN | YP_856818 | 0 | Aerobic respiration | 301 |
| Ribosomal protein S19 | YP_854843 | $7.78 \times 10^{-47}$ | Translation | 6,926 |
| Ribosome modulation factor | YP_856804 | $4.76 \times 10^{-28}$ | Stationary phase | 40,628 |
| RpoD | YP_855378 | 0 | Transcription | 708 |
| RpoS | YP_855373 | $2.64 \times 10^{-167}$ | Transcription | 2,018 |

[a] GenBank accession number for the top BLASTX hit at the NCBI.
[b] Expected value based on querying the nonredundant database at the NCBI using BLASTX.
[c] 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)].

generalized pattern throughout the tissue layer at all time points, while sulfoglycans, stained dark gray or black, were also present but more localized towards the cuticle and crop endothelium. At the endothelium, the signal for sulfoglycans was weakest at 8 h, with the intensity of HID staining increasing at the gut endothelium and brush border at later time points (Fig. 3A to E). An observed thin extension of the epithelium into the lumen also demonstrated similar positive patterns for both the HID/AB staining (Fig. 3F). These data demonstrate the presence of mucin in the crop of the leech and are consistent with the notion of leech mucins serving as nutrients for the *Rikenella*-like bacteria.

***A. veronii* utilizes multiple carbon sources in the crop.** In *A. veronii*, the expression of the stationary-phase alternative σ-factor *rpoS* was ~2.9-fold higher than that of the housekeeping σ-factor *rpoD*, and one of the most highly expressed genes encodes a ribosome modulation factor that has previously been shown to be induced in stationary phase (Table 1; see also Table S1 in the supplemental material) (30). These data suggest that at least some of the *Aeromonas* cells are entering stationary phase, which is consistent with our previous findings that indicated that *A. veronii* is transitioning into stationary phase by 42 h after feeding (25). Having a proportion of *A. veronii* enter stationary phase likely explains why genes involved in catabolism have low expression values in comparison to those encoding ribosomal proteins (Tables 1 and S1). *A. veronii* appeared to utilize at least two carbon sources 42 h after feeding, fatty acids such as acetate and carbohydrates. Genes encoding enzymes for the glyoxylate shunt of the tricarboxylic acid (TCA) cycle were expressed (Tables 1 and S1). The glyoxlyate shunt replenishes biosynthetic intermediates of the TCA cycle when cells are growing on acetate or other fatty acids (31), suggesting that *A. veronii* was using these fatty acids as a carbon and energy source.

Genes encoding fructose bisphosphatase (*fbp*) and phosphofructokinase (*pfk*), enzymes that perform irreversible steps of gluconeogenesis and glycolysis, respectively, were also expressed (Table 1; see also Table S1 in the supplemental material). The expression of *fbp* indicates that *A. veronii* synthesized glycolytic intermediates, a necessary requirement for growth on fatty acids. Expression of *pfk*, additional glycolytic genes (Table S1), and genes encoding enzymes for aerobic respiration (cytochrome *c* oxidase CcoN) and the oxidative TCA cycle (2-oxoglutarate dehydrogenase) (Tables 1 and S1) suggests that a subpopulation of *A. veronii* catabolized carbohydrates to $CO_2$. A previous study using epifluorescence microscopy revealed that *A. veronii* occupies two habitats inside the ILF; *A. veronii* is both pelagic and sessile in polysaccharide-embedded microcolonies together with *Rikenella*-like bacteria that resemble biofilms (25). One would expect the physiology of *A. veronii* in these environments to differ, and our data are consistent with this hypothesis.

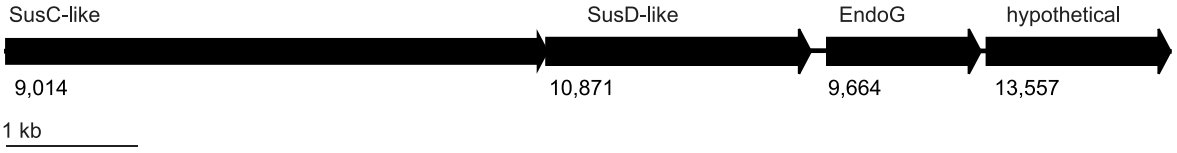


**FIG 1** Glycan utilization operon. The numbers below the arrows represent the expression values.

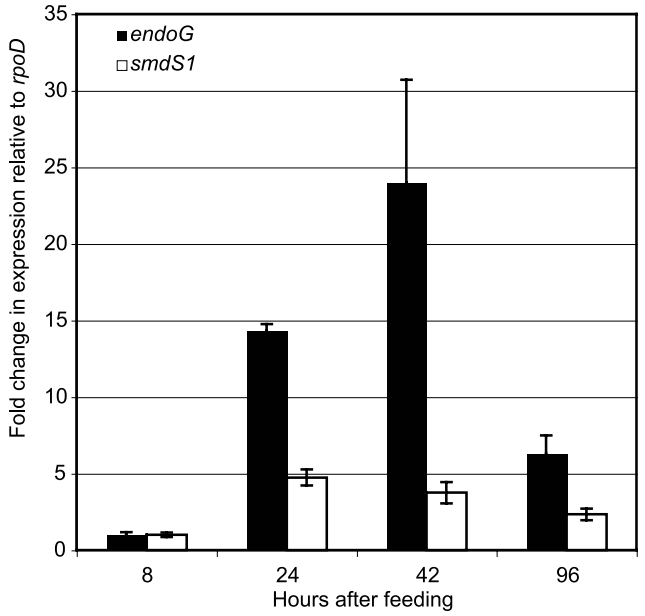


**FIG 2** Expression of *endoG* and *smdS1* inside the leech was confirmed using qRT-PCR. Relative expression of *endoG* and *smdS1* was measured at 8, 24, 42, and 96 h after feeding. The results are expressed as fold changes in expression relative to the level for *rpoD* and normalized to the level obtained 8 h after feeding. Each target was quantified in duplicate. Error bars represent standard deviations.

**Designing a culture medium for the uncultured symbiont by use of the metatranscriptome data.** A powerful aspect of the metatranscriptome is that it reveals the physiology of uncultured microbes during a state of active proliferation. The *Rikenella*-like bacterium had evaded our attempts to culture it in the laboratory using media others had used to grow *Rikenella microfusus* (32). The high expression levels of the mucin and glycan utilization genes strongly suggested that mucins were the main source of energy and carbon for the *Rikenella*-like bacterium inside the crop. We modified Eggerth-Gagnon medium (33) by replacing glucose with mucin from bovine submaxillary glands (EG-BM; 2 mg/ml). Plates were inoculated with serially diluted ILF and incubated in an anaerobic chamber for 2 weeks at 25°C. Two colony morphologies were observed on the EG-BM plates, large, gray, round, convex colonies and small, gray, round, convex colonies with white centers. Colonies were identified by diagnostic PCR (Fig. 4A) using the universal 16S rRNA gene primer 27F (34) and the *Rikenella*-specific 16S rRNA gene primer RikF1 (35), *Rikenella*-specific primers hypoF/hypoR, endoF/endoR, smdsF/smdsR, and rpoDF/rpoDR, and *Aeromonas*-specific primers gyrBF2/gyrBR2 (35) (Table 2). The diagnostic PCR and 16S rRNA gene sequencing identified the large colonies as *A. veronii* and the small colonies as the *Rikenella*-like bacteria. Pure cultures of the *Rikenella*-like bacteria were obtained (Fig. 4B) and subcultured using either bovine submaxillary gland mucin or porcine gastric mucin. The full-length 16S rRNA gene from the first *Rikenella*-like isolate, clone M3, was sequenced. This sequence was 100% identical to the 16S rRNA gene sequence from the metagenome, which was 99.5% identical to the sequence for AL5CE1, suggesting that these *H. verbana* leeches carried a *Rikenella*-like symbiont similar to that carried by *H. orientalis* (Fig. 5).

We were interested in isolating additional clones of the

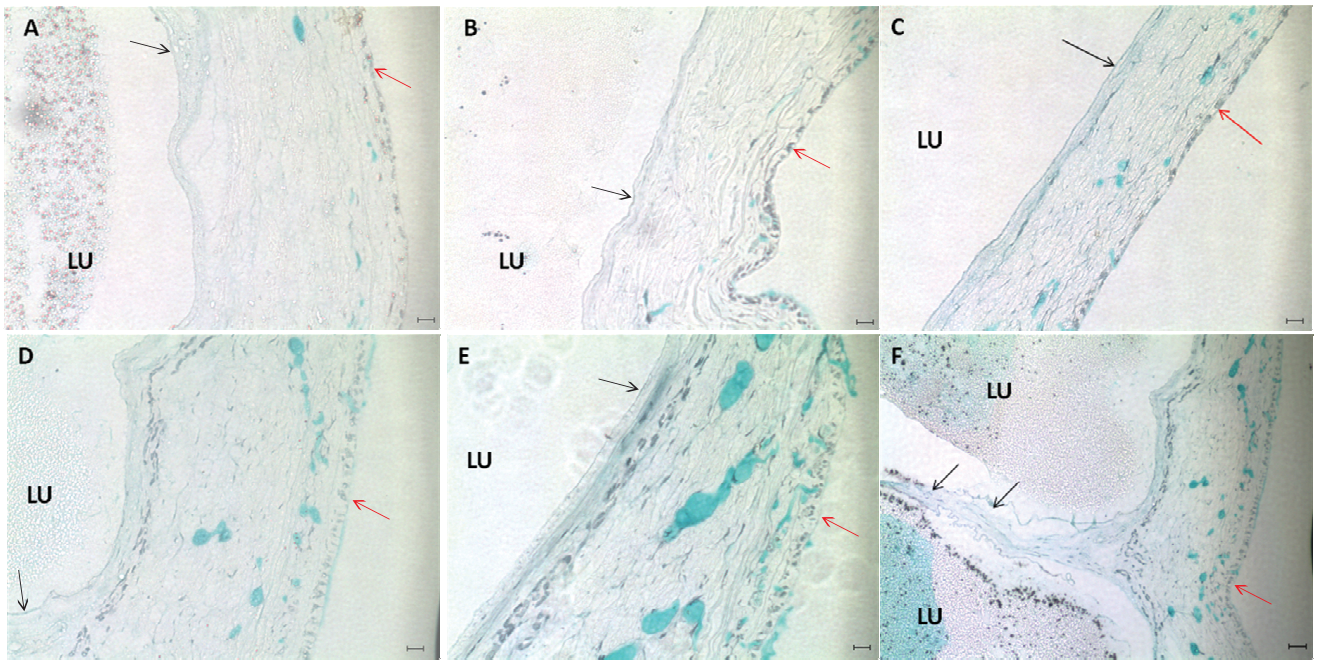


**FIG 3** Histochemical staining of the leech crop epithelium tissue detected sulfated and sialated carbohydrates. Results for HID/AB staining of leech tissue obtained at the below-indicated sampling time points after feeding are shown. Staining patterns consistent with the presence of both sulfoglycans (black-gray) and sialoglycans (blue) were present at all sampling time points, 8 h (A), 24 h (B), 42 h (C), 96 h (D), and 7 days (E). Sulfoglycans (black-gray) appear in the epithelial tissue with an increased signal seen at the lumen interface (black arrows) over time. Sulfoglycans can also be found in areas near the cuticle (red arrow). Sialoglycans (blue) are also present. LU, lumen of the crop. The scale bar represents 50 $\mu$m. (F) HID-AB staining of the epithelial flap (indicated by black arrows) extending into the lumen. LU, lumen. The scale bar represents 100 $\mu$m.

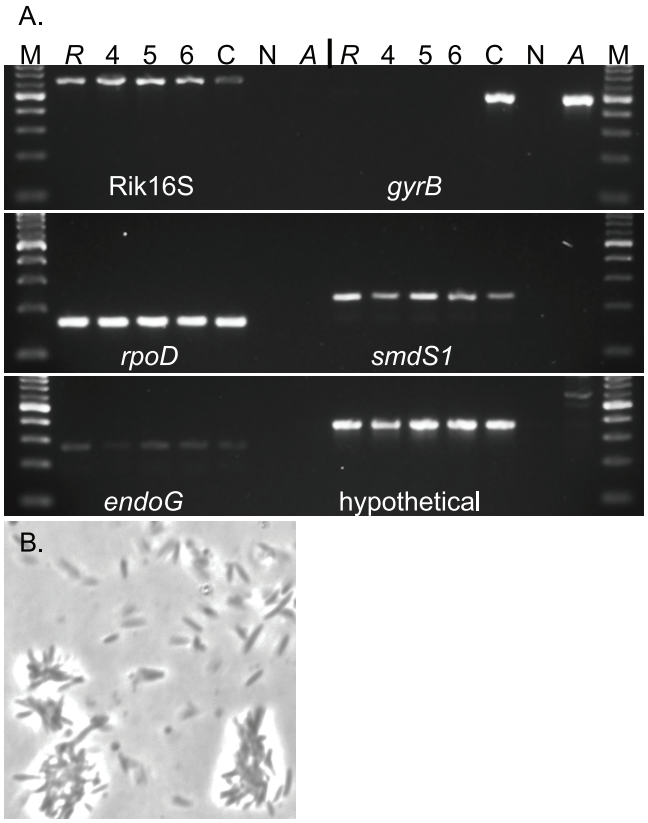


**FIG 4** Evidence supporting the cultivation of *Rikenella*-like bacteria. (A) Diagnostic PCR identification of *Rikenella*-like bacteria. Lanes: *R*, *Rikenella* clone M3 (genomic DNA template); 4, *Rikenella* clone M4 (colony template); 5, *Rikenella* clone M5 (colony template); 6, *Rikenella* clone M6 (colony template); C, crop contents (genomic DNA template); N, no template control; A, *A. veronii* (genomic DNA template); M, DNA marker (100, 200, 300, 400, 500, 600, 700, and 800 bp, from bottom to top). (B) Phase contrast image of *Rikenella*-like bacteria. The scale bar represents 10 $\mu$m.

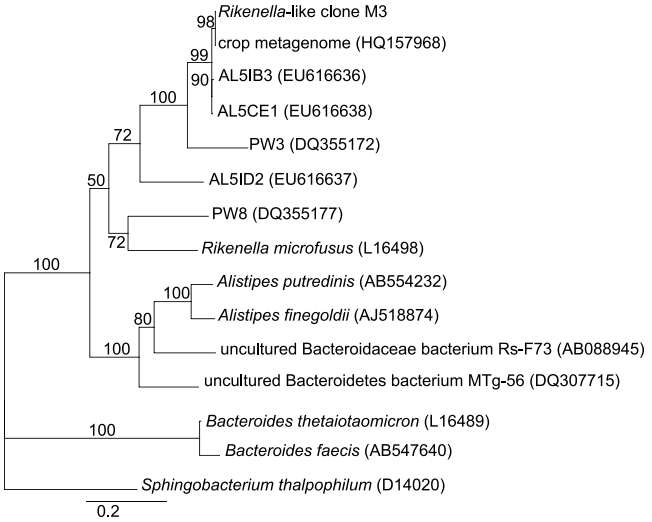


**FIG 5** Maximum-likelihood tree constructed from the alignment of partial 16S rRNA gene sequences. Partial 16S rRNA gene sequences were aligned using MUSCLE, and a PhyML tree was constructed from the alignment. Bootstrap support values from 1,000 resamplings and GenBank accession numbers are shown.

*Rikenella*-like bacterium. Therefore, porcine mucin-containing plates (EG-PM; 2 mg/ml) were inoculated with either ILF or intestinum contents from a second *Hirudo verbana* leech. The intestinum harbors a more complex microbiome and is where the blood meal is thought to be digested (18). Three additional clones of the *Rikenella*-like bacterium, M4, M5, and M6, were recovered from the crop. A fourth *Rikenella*-like bacterium, clone M7, and a *Desulfovibrio* sp. similar to AL5A3 (22) were isolated from the intestinum. Clones M4, M5, and M6 were subjected to the same diagnostic PCR as clone M3, and the results were identical (Fig. 4A), confirming that we had cultured additional *Rikenella*-like bacteria. Clones M4, M5, M6, and M7 were further characterized by sequencing 255 to 500 bp of the 16S rRNA gene. These sequences were also 100% identical to the 16S rRNA gene sequence from the crop metagenome. The diagnostic PCR (Fig. 4A) and the near-full-length 16S rRNA gene sequences support the notion that we obtained pure cultures of the *Rikenella*-like bacterium and now have both dominant symbionts in culture.

**TABLE 2** Primers and PCR amplification parameters used in this study

| Primer | Sequence (5′–3′) | Amplification conditions[a] |
|---|---|---|
| rpoBF | TTATCGTCTCCCAGCTGCACCG | 35 cycles of 30 s at 95°C, 30 s at 60°C, 30 s at 72°C |
| rpoBR | TGCTGGCAGTTTGCGACGAC | |
| smdSF | GTAACCACCCATCTCAAAAA | 35 cycles of 30 s at 95°C, 30 s at 60°C, 30 s at 72°C |
| smdsR | GTGAAGTCGGCAAGCATA | |
| rpoDF2 | CTGCCGAGAAATTTGATGAGACACG | 35 cycles of 30 s at 95°C, 30 s at 60°C, 30 s at 72°C |
| rpoDR2 | CTGTTCAAATCGTGCAAAAGCCTTG | |
| endoGF | TCTTTACCACCTCTTCCC | 35 cycles of 30 s at 95°C, 30 s at 60°C, 30 s at 72°C |
| endoGR | ATAAATTTCAGTCGCCCC | |
| 27F | AGAGTTTGATCMTGGCTCAG | 35 cycles of 30 s at 95°C, 30 s at 55°C, 60 s at 72°C |
| RikF | GCCCAGTATTAATCGCCATGT | |
| hypoF | TGATATTGTCTCGCTGGTGG | 35 cycles of 30 s at 95°C, 30 s at 57°C, 30 s at 72°C |
| hypoR | ATCATTTTGCGGGTTTGGT | |
| gyrBF2 | GCAGATTGGCGACAGCAC | 35 cycles of 30 s at 95°C, 30 s at 60°C, 30 s at 72°C |
| gyrBR2 | GCACCTTGACGGAGATAACG | |
| 27F | AGAGTTTGATCMTGGCTCAG | 32 cycles of 30 s at 95°C, 30 s at 50°C, 90 s at 72°C |
| 1492R | TACGGYTACCTTGTTACGACTT | |

[a] Each amplification reaction was preceded by an initial denaturing step at 95°C for 2 to 5 minutes, followed by a final extension step at 72°C for 2 to 5 minutes.

## DISCUSSION

The initial goal of this study was to investigate the physiology of the relatively simple digestive-tract microbiome of the medicinal leech *Hirudo verbana*. We selected a time point when the most abundant, yet uncultured, *Rikenella*-like symbiont was rapidly proliferating and the second most abundant and cultured symbiont, *A. veronii*, was close to entering stationary phase. Our RNA-seq study suggested that the *Rikenella*-like symbiont foraged host mucin glycans. Initially, this finding seemed surprising considering the recent influx of a nutrient-rich blood meal. However, previous reports have demonstrated that bacteria associated with the gnotobiotic mouse gut, like *Bacteroides thetaiotaomicron*, preferentially utilized host mucin glycans when dietary polysaccharides were scarce (36). *Rikenella* may have preferentially utilized host glycans over nutrients in the blood meal or utilized host glycans along with those present on the surface of the ingested erythrocytes. Interestingly, a recent report demonstrated that *B. thetaiotaomicron*'s ability to utilize host glycans is important for its maintenance in the mouse gut (37). Similarly, *Rikenella*'s capacity to utilize leech-derived mucins may be important for its ability to persist within the leech gut for up to 6 months between feedings.

The most highly expressed *Rikenella* operon (Fig. 1), which encodes a hypothetical protein, an endo-$\beta$-*N*-acetylglucosaminidase, and SusC-like and SusD-like proteins, resembles polysaccharide utilization loci found in other *Bacteroidetes*. Based on a query of the nonredundant database at NCBI using BLASTX (38), the proteins comprising this operon have the greatest amino acid sequence identity, between 54% and 65%, to BT1037-40 in *B. thetaiotaomicron*. In a recent report, expression of *BT1037-40* was shown to be upregulated in the gut of a gnotobiotic mouse fed a simple sugar diet when compared to growth on glucose minimal medium, suggesting that the gene products are involved in foraging host mucin glycans (37). Our data indicated that the related proteins in *Rikenella* may also be involved in utilizing host-derived mucin glycans, suggesting that these genes share a function.

Our data also indicate that *Rikenella* fermented the mucin glycans to acetate, a carbon source that the transcriptome data implicated to be utilized by *A. veronii*, thereby linking the physiologies of the two dominant symbionts. This potential syntrophic relationship between *Rikenella* and *A. veronii* could explain why we previously observed that bacteria found in mixed-species microcolonies proliferated more rapidly than those found in single-species microcolonies (Fig. 6) (25). *Rikenella* bacteria also colonize the gut epithelium where we detected mucins. It is possible that the acetate released by those bacteria may also serve as an energy source for the leech enterocytes (39).

The histological detection of both sialated and sulfated glycans is consistent with the idea that mucin is present along the crop epithelium and provided support for the notion that leech-produced mucin sugars serve as a nutrient source for the *Rikenella*-like bacteria. The RNA-seq data revealed that the *Rikenella*-like bacterium expressed genes encoding both sulfated-mucin desulfatases and sialidases, which should enable *Rikenella* to access the sulfated and sialated mucin glycans comprising the host mucus layer. The cultivation data provide direct evidence for *Rikenella*'s ability to utilize mucins that differed in composition. Gastric porcine mucin contains *N*-acetylgalactosamine, *N*-acetylglucosamine, galactose, and fucose (40), and bovine submaxillary mucin contains sialic acid, galactosamine, and glucosamine (41). It is conceivable that mucin sloughs off the epithelium with attached *Rikenella*-like bacteria and forms the microcolonies that subsequently become colonized by *A. veronii* (25) (Fig. 6). Lectin staining revealed the presence of *N*-acetylglucosamine in the matrix surrounding these microcolonies (25). The growth and dispersion dynamics of these microcolonies resembled that of biofilms, raising the interesting possibility that host mucin can serve as a nucleating center for bacterial biofilms inside the gut environment. Similar processes can occur in other digestive tracts, as a previous study reported that *B. thetaiotaomicron* can be found associated with shed mucus (36).

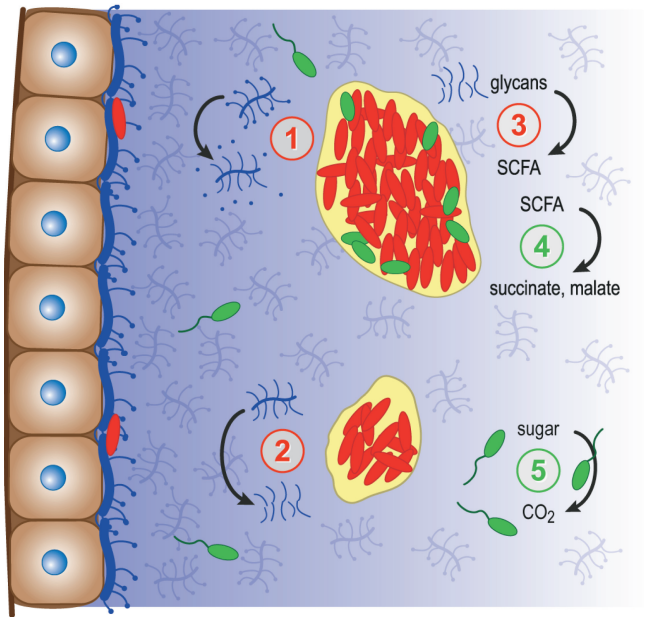


**FIG 6** Schematic displaying the proposed physiological processes carried out by the *Rikenella*-like bacteria (red) and *A. veronii* (green) in the crop. The circled numbers correspond to the steps as follows: 1, *Rikenella*-like bacteria remove sulfate from mucin glycans that are shed from or adjacent to the host tissue and are more concentrated near the endothelium (SmdS1 and SmdS2); 2, glycans are liberated from the protein (EndoG and SusC- and SusD-like proteins); 3, liberated glycans are fermented to short chain fatty acids (SCFAs) and secreted (phosphofructokinase, phosphate acetyltransferase, and acetate kinase); 4, *A. veronii* converts SCFAs to malate and succinate with the glyoxylate shunt of the TCA cycle (isocitrate lyase and malate synthase); 5, pelagic *A. veronii* catabolizes sugars to $CO_2$ (phosphofructokinase, 2-oxoglutarate dehydrogenase, and cytochrome *c* oxidase).

The discovery of mucin serving as *Rikenella*'s primary nutrient enabled the design of a medium that allowed the cultivation of *Rikenella* under conditions that favored its growth. The RNA-seq-based approach to direct culturing can be applied to an array of systems involving uncultured microbes ranging from abiotic environments, such as soil and water, to host-associated ones, including humans. The importance of the microbiome is being increasingly recognized as being critical for the health and well-being of humans, and one major effort of the human microbiome project is to sequence 900 genomes of cultured organisms (12). However, the paucity of cultured isolates hinders the learning of their physiology and the development of better experimental models. One challenge in cultivating a specific organism from a community is that other members of the microbiome can outcompete or hinder its growth. Metatranscriptomes can provide

valuable insight into the preferred nutrients that favor the isolation of an organism under laboratory conditions because metatranscriptomes reveal which genes are highly expressed in a specific *in situ* condition. Selecting time points or conditions when the target organism is growing rapidly, the metatranscriptome can reveal its nutritional preferences. Focusing on highly expressed genes encoding hydrolytic enzymes, binding proteins or transporters can reveal preferred nutrients. This knowledge can be used to design a culture medium containing the nutrients utilized *in situ* by the target microorganism. Many microbes are recalcitrant to cultivation, and we anticipate that the application of metatranscriptomics to growing bacteria in diverse habitats will enable the design of media allowing their cultivation. This will ultimately advance our understanding of beneficial symbionts, pathogens, and free-living microbes by enabling genome sequencing, physiological studies, and genetic approaches. Considering that more than 99% of microbes in most environments have yet to be cultivated, a metatranscriptomics approach to direct the culturing of uncultivated microorganisms is widely applicable.

## MATERIALS AND METHODS

**Leech feedings for the metatranscriptome.** Four *H. verbana* specimens (Leeches USA, Westbury, NY) were fed heparinized sheep blood (Quad Five, Ryegate, MT). After 42 h of incubation at room temperature, leeches were dipped in 70% ethanol and sacrificed. Fifty-microliter aliquots of the ILF were harvested and flash frozen in liquid nitrogen.

**Total RNA extraction and mRNA enrichment.** Total nucleic acid was extracted from 50-$\mu$l aliquots of the ILF using a MasterPure RNA purification kit (Epicentre Biotechnologies, Madison, WI) in accordance with the protocol for whole-blood samples, except samples were treated with 50 $\mu$g of proteinase K in 300 $\mu$l of tissue and cell lysis solution prior to total nucleic acid precipitation. DNA was removed using a Turbo DNA-*free* kit (Ambion, Austin, TX), using the rigorous protocol. DNA contamination was tested for with PCR using *Aeromonas*-specific primers RpoBF1/RpoBR1 (Table 2). PCR mixtures contained 1× GoTaq Green master mix (Promega, Madison, WI), 1 $\mu$M each primer, and the RNA template. Total RNA was quantified with the NanoDrop ND-1000 (Thermo Scientific, Wilmington, DE), and integrity was analyzed using an RNA FlashGel (Lonza, Basel, Switzerland). Equal amounts of total RNA were pooled from each leech. Ten micrograms of total RNA was enriched for mRNA using Microb*Express* (Ambion, Austin, TX). Enriched RNA was quantified using RiboGreen (Invitrogen, Carlsbad, CA).

**cDNA library preparation and Illumina sequencing.** One hundred nanograms of mRNA-enriched RNA was used for library preparation using SuperScript II (Invitrogen, Carlsbad, CA) and an mRNA-seq kit (Illumina, San Diego, CA) according to the manufacturer's instructions, except that the poly(T) bead enrichment step was omitted. The cDNA library was sent to the Science Core Facilities at Yale University for cluster generation and 1 × 34 bp sequencing on two lanes of a flow cell, generating 15,108,840 reads.

**Metatranscriptome analysis.** cDNA reads were mapped against the draft genomes of *A. veronii* strain HM21 and the *Rikenella*-like symbiont using CLC Genomics Workbench (version 4.0.2; Aarhus, Denmark). Reads which mapped more than once or with more than two mismatches were excluded from the analysis. Approximately 81% of the total reads mapped to *A. veronii* or *Rikenella*.

**Quantitative RT-PCR time course.** Leech feedings, total RNA extraction, DNase treatment, and quality assessment were carried out as described above. Equal amounts of total RNA were pooled from 2 or 3 leeches per time point. For 42 h, the total RNA used was from the same RNA pool used for the metatranscriptome. Double-stranded cDNA (ds cDNA) was synthesized from approximately 175 ng to 760 ng of total RNA using SuperScript II and Illumina mRNA-seq kit reagents in accordance with the manufacturer's instructions. cDNA copy number was quantified using 1× SsoFast EvaGreen Supermix (Bio-Rad, Hercules, CA), 500 nM each of the appropriate reverse and forward primers, and 0.5 $\mu$l of ds cDNA in a 12.5-$\mu$l total reaction volume on an iQ5 cycler (Bio-Rad, Hercules, CA). The following primer pairs were used: smdSF1/smdSR1, rpoDF2/rpoDR2, and endoF1/endoR1. The primer sequences are listed in Table 2, and the amplification conditions were as follows: (i) 2 min at 95°C and (ii) 35 cycles of 5 s at 98°C and 15 s at 55°C. Primer specificity was confirmed by sequencing and a melting curve analysis where the dwell temperature increased from 65°C to 95°C in 5°C increments every 10 s. The threshold cycle ($2^{-\Delta\Delta CT}$) method was used to calculate relative gene expression. *rpoD* was the reference gene, and 8 h was the control condition. Primer efficiency was evaluated using the standard curve method and ranged between 101 and 105%.

**Leech tissue sectioning and slide preparation.** Leeches were sacrificed 8 h, 24 h, 42 h, 96 h, and 7 days after feeding, with two replicates per time point. Leeches were processed by adapting a previously described method (25). Briefly, each leech was stretched and relaxed in 70% ethanol and prefixed whole in 15 ml Carnoy's solution (ethanol-chloroform-acetic acid; 6:3:1) overnight at 4°C. A 5-mm-thick cross-section was cut from approximately the same segment within the crop region of each leech in 80% ethanol and further fixed in 1.6 ml of Carnoy's solution overnight at 4°C. The tissue was bleached in an 80% ethanol solution containing 7% $H_2O_2$ and then dehydrated in absolute ethanol at 4°C. Nonbleached sections were done as controls to ensure that the results were not affected. The tissue was then subjected to a xylene-ethanol series at room temperature and embedded in paraffin. Sections 4 $\mu$m thick were cut on a rotary microtome and mounted on silane-coated glass slides. Slides were stored at 4°C until use.

**HID/AB staining and imaging.** Sections were cleared three times through an absolute xylene series followed by a graded ethanol series (100%, 90%, and 70%) before rehydration in Nanopure water. Sections were stained sequentially with high-iron diamine (HID) and Alcian blue (AB) at pH 2.5 to reveal the presence of sulfoglycans and sialoglycans, respectively (29, 42, 43). For HID and AB staining, sections were first incubated in a solution containing 0.24% dimethyl *meta*-phenylenediamine dihydrochloride (Acros Organics, Geel, Belgium), 0.04% dimethyl *para*-phenylenediamine dihydrochloride (Acros Organics, Geel, Belgium), and 1.7% $FeCl_3$ for 18 h at 25°C with agitation during the first hour. Sections were rinsed in running tap water, counterstained with a 1% AB solution for 10 min, and rinsed again with running tap water. Sections were then dehydrated in ethanol, cleared in xylene, and mounted with Permount (Fisher Scientific, Morris Plains, NJ). Slides were viewed under a Nikon Eclipse TE2000-S microscope, and images were taken with a Spot RT-KE camera. The brightness of each image was adjusted using Adobe Photoshop CS3 Extended (version 10; San Jose, CA).

**Culturing the *Rikenella*-like bacterium.** A *H. verbana* specimen was dipped in 70% ethanol outside an anaerobic chamber. Dissection was performed at room temperature in an anaerobic chamber under a $N_2$-$H_2$ (97:3) atmosphere. The ILF was serially diluted in 0.85% NaCl and plated on EG-BM agar plates, which is EG medium modified by replacing glucose with 2 mg/ml of bovine submaxillary gland mucin (MP Biomedicals, Solon, OH). Plates were incubated for 2 weeks at room temperature in the anaerobic chamber. The *Rikenella*-like clones could be subcultured on EG-PM, which contains 2 mg/ml of porcine gastric mucin (type III; Sigma-Aldrich).

**Diagnostic PCR.** The primer sequences and amplification conditions are listed in Table 2. The PCR mixtures contained 1× GoTaq Green master mix, 1 $\mu$M each the appropriate forward and reverse primers, and the template in a final volume of 20 $\mu$l. The template was 1 $\mu$l of a 1:20 dilution of a colony or 0.5 $\mu$l of genomic DNA. PCR products were run on a 2% Metaphor agarose gel (Lonza, Basel, Switzerland) and imaged with a Fluorochem FC2 camera (Cell Biosciences, Santa Clara, CA). The brightness and contrast of each image were adjusted using Adobe Photoshop CS3 Extended. The PCR products generated using 27F and 1492R were sequenced as previously described (23).

**Phylogenetic analyses.** 16S rRNA gene sequences were aligned in Geneious Pro (version 5.0.4; Auckland, New Zealand), using MUSCLE. After visual inspection of the alignment, a maximum-likelihood tree was constructed using PhyML (44) under the HKY85 substitution rate model (45). One thousand bootstrap resamplings were performed.

**Nucleotide sequence accession numbers.** Raw and processed RNA-seq data files were deposited in Gene Expression Omnibus at the NCBI under accession number GSE23786. The 16S rRNA gene sequences generated in this study were deposited in GenBank under accession numbers HQ157960 to HQ157968.

## ACKNOWLEDGMENTS

We thank K. Guillemin and Z. Stephens for sharing unpublished data, P. LaPierre and H. W. Kim for help with bioinformatics, K. Noll for the use of the anaerobic chamber, Virge Kask for the artwork, and D. Benson and S. Nyholm 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.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00012-11/-/DCSupplemental.

Table S1, XLS file, 0.423 MB.

Table S2, XLS file, 0.254 MB.

## REFERENCES

1. **Amann RI, Ludwig W, Schleifer KH.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. **59**:143–169.
2. **Stingl U, Tripp HJ, Giovannoni SJ.** 2007. Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. ISME J. **1**:361–371.
3. **D'Onofrio A, et al.** 2010. Siderophores from neighboring organisms promote the growth of uncultured bacteria. Chem. Biol. **17**:254–264.
4. **Kaeberlein T, Lewis K, Epstein SS.** 2002. Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. Science **296**:1127–1129.
5. **Qin J, et al.** 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature **464**:59–65.
6. **Tyson GW, et al.** 2005. Genome-directed isolation of the key nitrogen fixer *Leptospirillum ferrodiazotrophum* sp. nov. from an acidophilic microbial community. Appl. Environ. Microbiol. **71**:6319–6324.
7. **Frias-Lopez J, et al.** 2008. Microbial community surface gene expression in ocean surface waters. Proc. Natl. Acad. Sci. U. S. A. **105**:3805–3810.
8. **Gilbert J, et al.** 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS One **3**:e3042.
9. **Urich T, et al.** 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the metatranscriptome. PLoS One **3**:e2527.
10. **Turnbaugh PJ, et al.** 2010. Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. Proc. Natl. Acad. Sci. U. S. A. **107**:7503–7508.
11. **Turnbaugh PJ, et al.** 2009. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci. Transl. Med. **1**:6ra14.
12. **Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI.** 2005. Host-bacterial mutualism in the human intestine. Science **307**:1915–1920.
13. **Comstock LE.** 2009. Importance of glycans to the host-*Bacteroides* mutualism in the mammalian intestine. Cell Host Microbe **5**:522–526.
14. **Turnbaugh PJ, et al.** 2007. The human microbiome project. Nature **449**:804–810.
15. **Nelson KE, et al.** 2010. A catalog of reference genomes from the human microbiome. Science **328**:994–999.
16. **Eckburg PB, et al.** 2005. Diversity of human intestinal microbial flora. Science **308**:1635–1638.
17. **Graf J.** 1999. Symbiosis of *Aeromonas veronii* biovar *sobria* and *Hirudo medicinalis*, the medicinal leech: a novel model for digestive tract associations. Infect. Immun. **67**:1–7.
18. **Worthen PL, Gode CJ, Graf J.** 2006. Culture-independent characterization of the digestive-tract microbiota of the medicinal leech reveals a tripartite symbiosis. Appl. Environ. Microbiol. **72**:4775–4781.
19. **Sawyer RT.** 1986. Leech biology and behavior. Oxford University Press, Oxford.
20. **Graf J, Kikuchi Y, Rio RV.** 2006. Leeches and their microbiota: naturally simple symbiosis models. Trends Microbiol. **14**:365–371.
21. **Silver AC, Rabinowitz NM, Küffer S, Graf J.** 2007. Identification of *Aeromonas veronii* genes required for colonization of the medicinal leech, *Hirudo verbana*. J. Bacteriol. **189**:6763–6772.
22. **Laufer AS, Siddall ME, Graf J.** 2008. Characterization of the digestive-tract microbiota of *Hirudo orientalis*, a European medicinal leech. Appl. Environ. Microbiol. **74**:6151–6154.
23. **Silver AC, et al.** 2007. Interaction between innate immune cells and a bacterial type III secretion system in mutualistic and pathogenic associations. Proc. Natl. Acad. Sci. U. S. A. **104**:9481–9486.
24. **Yoder-Himes DR, et al.** 2009. Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. Proc. Natl. Acad. Sci. U. S. A. **106**:3976–3981.
25. **Kikuchi Y, Graf J.** 2007. Spatial and temporal population dynamics of a naturally occurring two-species microbial community inside the digestive tract of the medicinal leech. Appl. Environ. Microbiol. **73**:1984–1991.
26. **Elder J, Alexander S.** 1982. Endo-$\beta$-*N*-acetylglucosaminidase F: endoglycosidase from *Flavobacterium meningosepticum* that cleaves both high-mannose and complex glycoproteins. Proc. Natl. Acad. Sci. U. S. A. **79**:4450–4454.
27. **Xu J, Gordon JI.** 2003. Honor thy symbionts. Proc. Natl. Acad. Sci. U. S. A. **100**:10452–10459.
28. **Tsai H, Hart C, Rhodes J.** 1991. Production of mucin degrading sulphatase and glycosidases by *Bacteroides thetaiotaomicron*. Lett. Appl. Microbiol. **13**:97–101.
29. **Spicer SS.** 1965. Diamine methods for differentiating mucosubstances histochemically. J. Histochem. Cytochem. **13**:211–234.
30. **Yamagishi M, et al.** 1993. Regulation of the *Escherichia coli rmf* gene encoding the ribosome modulation factor: growth phase- and growth rate-dependent control. EMBO J. **12**:625–630.
31. **Cronan J, Jr, LaPorte D.** 1996. Tricarboxylic acid cycle and glyoxylate bypass. ASM Press, Washington, DC.
32. **Collins MD, Shah HN, Mitsuoka T.** 1985. Reclassification of *Bacteroides microfusus* (Kaneuchi and Mitsuoka) in a new genus *Rikenella*, as *Rikenella microfusus* comb. nov. Appl. Microbiol. **6**:79–81.
33. **Eggerth AH, Gagnon BH.** 1933. The *Bacteroides* of human feces. J. Bacteriol. **25**:389–413.
34. **Lane D.** 1990. 16S/23S rRNA sequencing. John Wiley & Sons, Chichester, United Kingdom.
35. **Rio RV, Maltz M, McCormick B, Reiss A, Graf J.** 2009. Symbiont succession during embryonic development of the European medicinal leech, *Hirudo verbana*. Appl. Environ. Microbiol. **75**:6890–6895.
36. **Sonnenburg JL, et al.** 2005. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science **307**:1955–1959.
37. **Martens EC, Chiang HC, Gordon JI.** 2008. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe **4**:447–457.
38. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.
39. **Bergman EN.** 1990. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. Physiol. Rev. **70**:567–590.
40. **Ohara S, Ishihara K, Hotta K.** 1993. Regional differences in pig gastric mucins. Comp. Biochem. Physiol. B **106**:153–158.
41. **Nisizawa K, Pigman W.** 1959. The composition and properties of the mucin clot from cattle submaxillary glands. Arch. Oral Biol. **1**:161–170.
42. **Nyholm SV, Deplancke B, Gaskins HR, Apicella MA, McFall-Ngai MJ.** 2002. Roles of *Vibrio fischeri* and nonsymbiotic bacteria in the dynamics of mucus secretion during symbiont colonization of the *Euprymna scolopes* light organ. Appl. Environ. Microbiol. **68**:5113–5122.
43. **Walsh MD, Jass JR.** 2000. Histologically based methods for detection of mucin. Methods Mol. Biol. **125**:29–44.
44. **Guindon S, Gascuel O.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52**:696–704.
45. **Hasegawa M, Kishino H, Yano T.** 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.