2016

Microbial Dynamics within Shed Mucosal Secretions of Hirudo verbana, the European Medicinal Leech

Brittany Maree Ott

Microbial Dynamics within Shed Mucosal Secretions of *Hirudo verbana*, the European Medicinal Leech

by

Brittany Maree Ott

Dissertation submitted to the
Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Biology

Rita V. M. Rio, Ph.D., Chair
Nyles Charon, Ph.D.
Andrew Dacks, Ph.D.
Stephen DiFazio, Ph.D.
Jennifer Hawkins, Ph.D.

Department of Biology

Morgantown, West Virginia
2016

Keywords: symbiosis, transmission, leech, *Aeromonas veronii*, mucus, Illumina sequencing, RNA-seq

ABSTRACT

Microbial Dynamics within Shed Mucosal Secretions of *Hirudo verbana*, the European Medicinal Leech

Brittany Ott

Microbes are widespread throughout our planet, residing in soils, oceans, and even beneath the arctic glaciers. Most interact with each other and other life forms, and there is no known complex organism that lacks an associated microbiome. These microbial symbionts are critical to the survival and proliferation of their host by assisting in nutrient provisioning, physiological development, immunological priming, providing protection from pathogens, contributing to predator evasion, etc. As such, the transmission of these symbionts is a crucial aspect of host biology. Despite a growing appreciation for the prevalence of mixed transmission (incorporating vertical, or from a parental route, and horizontal, or environmental, mechanisms) for establishing microbial symbioses, features that enable these infections are not well understood. My work investigated the mechanistic basis of symbiont acquisition by the European medicinal leech, *Hirudo verbana*. In the leech cocoon, only a portion of the albumenotrophic larvae obtain their beneficial gut symbiont, the Gammaproteobacterium *Aeromonas veronii*. However, by early adulthood, all leeches harbor this bacterium, indicating the complementation of incomplete vertical transmission by an unknown horizontal mechanism. Insight from a number of other organisms, aquatic and terrestrial, suggests that host-secreted mucus may provide a transmission vehicle. Through the use of genetically-tractable *A. veronii*, I demonstrated that cyclical host mucosal secretions are seeded by digestive tract symbionts. Using quantitative PCR, I verified that *A. veronii* are not only viable, but also proliferate within mucosal casts at a frequency synchronous to host shedding. Subsequent experiments using mucus inoculated with a *gfp*-expressing *A. veronii* demonstrates that mucosal contact is sufficient for symbiont transmission to novel hosts. Importantly, behavior assays show that leeches are attracted to these castings, providing an efficient mechanism for transmission of *A. veronii* between conspecifics. Additional behavioral assays show that host symbiont-state does not influence attraction towards mucus and that symbiont content of the mucus does not alter attraction, suggesting that *A. veronii* exploits a preexisting host physiological process.

Lastly, Illumina-based RNA-seq was used to identify how the *A. veronii* transcriptome, with regards to metabolism and information processing, responds to the lifestyle shift from mutualistic within the leech digestive tract to free-living within shed mucus. This dual mode of symbiont transmission may prove evolutionarily advantageous, as it not only ensures the infection of leeches by beneficial symbionts, but also provides accessibility to a higher genetic diversity of symbionts and biome lifestyle options to *A. veronii* beyond that of mutualism. Additionally, examination of the composition of the mucosal microbial community utilizing two separate culture-independent sequencing techniques (i.e. Sanger-sequenced 16S rRNA clone libraries and Illumina deep-sequencing of the V3-V4 hypervariable region of the 16S rRNA gene) revealed that a diverse microbiota resides within the mucus, consisting of both previously-described and potentially novel leech symbionts. Additional transcriptomic analyses indicate that this microbial community actively engages in cell-to-cell communication via quorum sensing and potential DNA transfer. In depth genomic analyses also prove that a novel microbial species, *Pedobacter*, is a major mucosal microbiota member worthy of future study. Understanding the features that enable mixed transmission, particularly those involving mixed species assemblages, may prove instrumental for the development of strategies to promote beneficial symbioses.

Acknowledgements

I would first, and foremost, like to thank my incredible Doctoral advisor, Rita Rio, Ph.D., for all of the support and guidance that you have given me over the years, both professionally and personally. Back in 2009, when I completed the summer REU in your lab, I was still debating whether I wanted to continue into research. However, those ten weeks encouraged me to pursue research as my career. When I returned for graduate school, I continued to learn about what it was to be a scientist. I appreciate your pushing me beyond my limits, because it showed me how strong I can be and how much I can accomplish. I also thank you for supporting the decisions I have made for my future; it means more than you can know.

I want to also thank my committee members, Nyles, Jen, Steve and Andrew. I appreciate all of your questions, advice and suggestions, as they have strengthened my research and pushed it to meet the high standards that you hold.

I would also like to graciously thank all those who helped contribute to my thesis research. I would first like to thank our collaborators at the University of Connecticut in Dr. Joerg Graf's lab. Your assistance by answering my many questions and providing critical bacterial strains has greatly helped progress my research. I could not have done many of my experiments without your help. I would next like to thank Anna Snyder for all of the late nights we spent in the lab and in our office. I will forever appreciate your opinions and advice that helped me through the tough moments of difficult experiments, as well as difficult times in my life. Additionally, this work would not be possible without the assistance of many past and present lab members, so I would like to thank; Vivian and Adam, as well as many undergraduate students who assisted with so many of my experiments and ideas, specifically Michael, Grace, Noelle, James and Allen. The time spent with you in the lab has helped me become a better researcher and mentor, as well as colleague and friend, and has helped with both my research and peace of mind.

While most of my time as a graduate student has been spent in research, I still met with challenges associated with being a graduate student. So I would like to thank Mickey, Diana, Chuck, Judy, Pat and Wendy in the Biology Department for helping make my time here a little bit easier. I would also like to thank the Biology Department graduate students, both past and present, who have helped me make it through each day; I would especially like to thank Micah, Mohna, Dhanu, Kristyn, Jessie Brie, Jen, Tyler and Luke for helping me out with the research-related, graduate student-related, and life-related issues that I have encountered.

I would like to thank Katrina Stewart for helping with all of the issues I encountered as a first-time teaching assistant, as I struggled to understand how to teach students not much younger than myself. I would also like to thank Dana Huebert-Lima for helping me go above and beyond as a teacher, and guiding me to find the right balance between strict and lenient, caring and distant, so I could be the best I could be for my students.

Finally, I would like to thank my friends Lauren, Alan, Carol, Matt, Samantha, Kate and Kyle for always understanding and remaining my friends as I would occasionally drop off the face of the Earth as my work engulfed me. I want to thank my mum and dad for ALWAYS giving me unwavering support through this whole endeavor. And last, but certainly not least, I want to thank Alex, my rock, my love, and my partner. Your continued support and faith in me have helped me through the good times and the bad. You are my other half, and as long as we have each other, we can make it through anything.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: Introduction

Symbiosis

Symbiosis, the long-term physical association between two or more species [1, 2], is ubiquitous throughout the natural world [3], affects every facet of life [4] and has greatly impacted evolutionary history [5]. While these relationships can occur between macroorganisms, such as the grooming symbioses of large mammals in Africa by oxpeckers [6], or similarly that of cleaner wrasses, which ingest the dead skin and ectoparasites of other fish in exchange for protection from predation [7], there are no known animals that have evolved separately from bacterial symbionts [8, 9]. Partnerships involving microbes can be very ancient, or relatively recent in their time of establishment and can range from facultative, where the association is optional, to obligate, where the fitness of at least one partner necessitates the relationship [10]. Older symbioses, in particular, tend to show more intricate interdependencies and cooperative behavior, demonstrating the dynamics of these relations through time (reviewed in [11]).

Diversity in symbiosis

Symbioses display a wide variety of forms including parasitism/pathogenesis, commensalism or mutualism [10]. These association types are routinely characterized based on the net effect of the relationship towards both partners [10]. In parasitism/pathogenesis, one partner benefits at the expense of the other; an example of this includes the colonization of the human upper respiratory tract by the bacterium *Bordetella pertussis*, the causative agent of whooping cough [12]. A variety of virulence factors are utilized to cause severe coughing [13], enhancing *B. pertussis* transmission to naive hosts through infected respiratory droplets [12]. Commensalism can be demonstrated by *Staphylococcus aureus* that resides on the human skin, but neither benefits nor harms humans, if the host epithelium remains intact [14, 15]. Mutualism,

where both partners benefit from the association, is crucial to the fitness of many organisms [16]. For example, legumes would not be able to thrive without the diazotrophs, which are capable of fixing atmospheric nitrogen to a fertilizable form, that grow within their root nodules [17]. In another fascinating example, some oligochaete worms utilize a microbe-microbe interaction between sulfate-reducing δ-proteobacteria and sulfide-oxidizing γ-proteobacteria symbionts to thrive in environments high in toxic sulfide by-products [18]. In return, the microbes receive a stable niche [19], as well as a potential reduction in inter- and intra-species competition relative to that experienced within a free-living environment [20, 21].

Microbial transmission and effects on microbial genome evolution

The transmission of microbial partners is critical for the establishment and persistence of symbioses. However, much remains to be discovered concerning the mechanisms that are used for establishment. Broadly speaking, microbial establishment may occur through vertical transmission (VT), the acquisition of an infection by offspring via the parental route, or horizontal transmission (HT), the transfer of infection from one individual to another irrespective of relation and mediated through the environment (Fig. 1). The frequency of VT is directly correlated with host fecundity [22], whereby transmission can occur by the infection of developing or fertilized eggs (reviewed in [23]), through the nutrient-providing milk glands of the mother [24, 25], or even through parental smearing of the egg exterior with the symbiont [26] (reviewed in [27]). In contrast, HT frequency is positively correlated with host density and requires physical contact with environmental microbes (reviewed in [23]), either by direct contact or through the utilization of a vehicle which bridges partners. Cases of HT are also very diverse. For example, newly emerged *Apis mellifera*, the Western honeybee, workers acquire their gut microbiota from contact with older workers and the hive [28]. In addition, some sessile

plants, such as hornworts, induce motility in soil-inhabiting cyanobacteria by secreting hormogonium-inducing factors (HIFs), driving the microbes to enter stomata-like openings in the plant root (reviewed in [29]).

VT and HT have been traditionally assumed to be negatively correlated, (i.e. a trade-off exists between VT and HT). For example, if a symbiont attempts to increase its chances of HT by replicating to the point of host death, it will obviously terminate the chances of VT. Lytic (HT) and lysogenic (VT) cycles of many phages demonstrate such a trade-off between transmission routes [30]. However, a trade-off is not a definitive aspect when correlating VT and HT. An example is the association between the protozoan *Ophryocytis elektroscirrha* and the monarch butterfly *Danaus plexippus*, where a higher rate of replication not only increases its chances of HT by making the butterfly more infectious through the increased number of parasite spores shed into the environment for other hosts to ingest, but also simultaneously boosts its population within the eggs of the host [31], thereby enhancing VT probability. In summary, as long as one mechanism does not compromise the other, a trade-off does not necessarily have to exist, and these two transmission routes do not have to be mutually exclusive.

As such, an integration of both VT and HT referred to as mixed mode transmission (MMT; Fig. 1) (reviewed in [32]) may frequently occur. While it was previously believed that MMT was an exceptional case rather than a common mode for transmission, deeper investigation into symbionts labeled as strictly vertically transmitted has shown that these microbes should actually be considered MMT, as many show signs of horizontal transmission [33, 34] (reviewed in [35, 36]). Some examples of this include *Wolbachia*, a reproductive parasite which infects a wide range of invertebrates through the maternal germline ([37]; and reviewed in [36]) and can be transmitted via insect parasitoids [38]; three Enterobacteriaceae symbionts of the pea aphid,

Acyrthosiphon piscum, that are horizontally transmitted from males to females, and can subsequently move among the aphid matriline [39]; and many species of microsporidia, which are obligate, intracellular parasites of many vertebrate and invertebrate taxa, whose spores can be released to the environment (HT) or transferred to the host eggs (VT) [34]. Elucidating transmission routes is critical for understanding the evolution of a relationship, as these mechanisms have distinct and significant impacts on the symbiosis (reviewed in [32]).

In systems that utilize strict vertical transmission, lifestyle efficiency is enhanced through the genome tailoring of bacterial partners (reviewed in [35]). Relaxed selection on genes no longer necessary for maintenance of the symbiosis may occur, leading to their erosion (i.e. becoming a pseudogene) and eventual purging from the genome, resulting in reduced genome size (reviewed in [35]). These genomes may become so streamlined that microbes are tethered to their host in an obligate symbiotic relationship (reviewed in [35]), with little to no opportunity to revert to a free-living lifestyle. In fact, the genomes of bacterial symbionts that have been vertically transmitted over an evolutionarily significant amount of time represent some of the smallest described to date (reviewed in [40]). Accelerated mutation rates, reflected in AT bias in genome contents, also occur due to bottlenecks during transmission and small effective population sizes exacerbating genetic drift [41] (reviewed in [32, 35]). Lastly, due to their clonal presence often within isolated locations of the host body, such as intracellular localization within specific host cell types, there is little chance for inter-strain recombination or the acquisition of novel genetic material and capabilities through horizontal gene transfer (HGT) and phage infection.

However, the acquisition of even a small proportion of the symbiont population through HT enables symbiont swapping, recombination, and HGT, potentially leading to increased

genetic diversity within the host microbiota. Through MMT, symbionts are capable of rapid adaptation by adjusting the rates of either transmission mode (reviewed in [32]), potentially allowing persistence in a wide range of ecological conditions (reviewed in [32]). This rapid adaptation enables the symbiosis to survive in situations that may limit certain forms of transmission (reviewed in [32]), especially critical for obligate relations.

Scientific technology has accelerated our knowledge of symbiosis

Some key questions in symbiosis involve our understanding of how symbioses establish and how partners impact one another. To address questions in symbiosis research, scientists have taken advantage of the latest advancements in technologies. For example, the production of microscopes during the time of Leeuwenhoek not only allowed observation of microbes for the first time, but also microbial symbioses (reviewed in [4]). The development of the sterile technique and culturing methodologies in the 1800s enabled reliable experimental study of cultivable microbes. Additionally, the discovery of DNA, and thus the hereditary code of life, along with the establishment of Crick's Central Dogma in the 1950s lay the foundation for today's burgeoning genomic and transcriptomic fields. Currently, we have sequencing technologies, such as Illumina MiSeq and HiSeq [42], and Roche 454 pyrosequencing [43, 44], in addition to well-established protocols to characterize metagenomic and metatranscriptomic libraries at high speeds, deep coverage and low cost. Combining the power of these technologies, we are now well poised to ask questions which will further our basic understanding of the formation, persistence and impact of symbiosis.

The use of simple symbiotic model systems

Mutualistic bacteria play essential roles towards host digestion [27], reproduction [45], behavior [46, 47], immune system stimulation [48, 49] and aiding in the development of body

features[46, 50, 51]. The Human Genome Project revealed that there are far fewer human genes than expected for our level of metabolic and physiological complexity [52], and this begs the question: what attributes to that complexity? To answer this question, many scientists are examining various features of the human microbiota, the microbial community that lives on or in us. One major feature that deserves recognition is the gut microbiota, which can be seen as its own organ within humans [53]. A tremendous impediment to studying the human gut microbiota is its rich microbial species composition [54], with greater than a thousand species [55] interacting with each other and host cells [56]. Study of simple symbioses, such as between a host and a small and consistent number of microbial partners, can enhance our understanding of microbiota evolution and ecology.

European medicinal leech as a simple symbiosis model system

The European medicinal leech, *Hirudo verbana*, is a widely studied organism due to the simplicity of its physiology [57], nervous system [57], immune system [58] and microbial symbioses [59]. Simple symbiotic communities are harbored within the leech nephridial and digestive tracts. The leech excretory system consists of 17 pairs of 3-lobed nephridia connected to bladders that run the length of the leech [57]. A maximum of six bacterial species are localized within the bladders with these symbionts exhibiting stratified spatial arrangement in their community organization [60], likely impacted by resource availability. While the functions of these symbionts, relative to their leech host, are mainly unknown, it is possible that they contribute to the recycling of carbon and nitrogenous waste [60], potentially enabling long, sustained gaps between blood meals.

The second simple symbiotic community is housed in the digestive tract, which is basically comprised of a crop and the relatively smaller intestinum, functioning in nutrient

storage and digestion/adsorption, respectively [57, 61]. While the intestinum harbors a greater microbial diversity, the majority are believed to be transient [62]. The crop retains a simpler resident microbial community, housing only two predominant and consistent species, *Aeromonas veronii* biovar sobria and *Mucinovorans hirudinis* [63, 64]. These two species have been shown to interact with each other [63, 65] as well as with the leech host. Nutrient production necessitates the three partners working in conjunction (see below) [63], while the microbial symbionts assist with digestion and nutrient provisioning [66] and potentially offer shielding from pathogens via protective colonization [61].

***Aeromonas veronii* biovar sobria: the pioneer symbiont of the leech digestive tract**

The genus *Aeromonas* consists of gram-negative, aquatic γ-proteobacteria that are common pathogens of fish and other amphibious, cold-blooded species [67]. These bacteria are also known to be the cause of disease in immunocompromised humans. Aeromonads are associated with gastroenteritis [67], along with some cases of wound infection following medicinal leech therapy [68]. *A. veronii* is a pioneer symbiont, or an early colonizing microbial species, of the digestive tract of *Hirudo verbana* [62, 65, 66], possibly assisting in blood meal digestion [61], providing essential nutrients deficient in the host's strictly hematophagous diet [61, 63], and influencing the colonization of other members of the microbiota [61]. The *A. veronii* symbionts, localized mainly to the leech crop [62], are known to utilize both Type II [69] and Type III [70] secretion systems when colonizing their leech host, demonstrating the use of traditionally recognized virulence mechanisms to establish a beneficial association. There are approximately twenty-six published species of *Aeromonas* [67, 71–76]. Three of these *Aeromonas* species are associated with distinct *Hirudo* leech host species: *H. medicinalis* harbors

A. hydrophila [77], *H. orientalis* has been shown to house both *A. veronii* and *A. jandaei* [78], and, as mentioned above, *H. verbana* harbors *A. veronii*.

Mucinivorans hirudinis

The second bacterium in the digestive tract of *H. verbana*, as well as that of *H. orientalis* [77, 78] and *H. medicinalis* [77], is *Mucinivorans hirudinis*, previously identified as a *Rikenella*-like bacterium [64] in the phylum Bacteroidetes [62]. Members of the Rikenellaceae family have been isolated from the guts of a diverse range of animals [62, 79] including termites [62], mice [80] and humans [54, 81]. *M. hirudinis* is an obligate anaerobe [62, 63] that utilizes host mucin glycans sloughed from the crop epithelial layer as an energy source [63], fermenting these and secreting short chain fatty acids that can be further utilized by both the leech enterocytes and *A. veronii* as a carbon source [63]. Within the leech digestive tract, *M. hirudinis* typically form sole species microcolonies, or are mixed with *A. veronii* [65], that tend to be associated with the host epithelial tissue [65], but can also be found in the intraluminal fluid (ILF) of the leech [65]. Interestingly, the mixed microcolonies are typically larger and denser than those of a single species [65], indicating synergism between the two bacteria [65].

Leech reproduction and development

With *Hirudo* leeches, the process of reproduction requires both aquatic and terrestrial environments [57]. Although leeches are hermaphroditic, they are unable to self-fertilize [57]. Leeches reproduce by internal fertilization, after which a cocoon is deposited above the shoreline under damp soil or moss [57]. Each cocoon can house up to 30 eggs [82] immersed in a nutritious albumen fluid [57]. As the embryos develop they are albumenotrophic as larvae [57], but just prior to juvenile state maturation, the larval mouth recedes and consumption is terminated [83]. About four weeks post-cocoon deposition, the juveniles will emerge fully

pigmented and functional [84]. At this time, less than 80% of the host population harbors *A. veronii* symbionts [85], with no *M. hirudinis* detected. Interestingly, both symbionts can be detected in 100% of 2 week-old leech individuals [85], indicating the complementation of incomplete VT (likely through the maternal seeding of albumen) by HT following cocoon emergence (i.e. MMT). Despite this, mechanisms involved in HT of the leech digestive tract microbiota remain to be identified.

Co-evolution between the medicinal leech and its symbionts

Symbioses affect the evolution of all parties involved [5, 86], sometimes even exhibiting deep concordant evolution between two or more partners [5, 86]. Co-evolution between two partners can be observed through phylogenetic tree construction [5, 86], where the trees of the individual species basically mirror one another [5, 86]. However, co-evolution between *A. veronii* and the medicinal leech has not been observed using traditional phylogenetic studies [77], despite the tight *Aeromonas* species-specificity to the medicinal leech species [77]. Moreover, it is difficult to create a consistent phylogeny for *A. veronii* using multiple gene trees [87], as different evolutionary histories are observed in and among gene sets [87]. It is likely that recombination and HGT have blurred the evolutionary history of these genes [87]. This lack of co-evolutionary history between the medicinal leech and its respective *Aeromonas sp.*, in addition to the incomplete vertical transmission of this symbiont [85], strongly suggests *Aeromonas sp*. HT between leech individuals and supports use of the *H. verbana* leech-*A. veronii* symbiosis to enhance our understanding of MMT in regards to the formation and persistence of symbiosis.

Vehicle for horizontal transmission in the leech

As mentioned above, HT requires contact with the microbial symbiont in question, whether directly or through a vehicle connecting hosts in an often unpredictable environment (reviewed in [23]). Mucosal secretions are regularly shed by leeches [57, 88], typically seen as white tendrils attached to plants, or white rings found on the ground or bottom of their holding tank [88][personal observation]. These secretions consist of glucosaminoglycans [88] produced by tubular mucus gland cells, collectively known as globose glands, distributed below the epithelia [57, 88]. Additionally, leech mucus is also believed to house chemo-attractants that facilitate conspecific recognition [57, 88].

Notably, it has been demonstrated that regularly secreted, host-generated mucus facilitates HT-mediated symbiont recruitment in a number of marine animals. In *Euprymna scolopes*, the Hawaiian seploid squid, mucosal secretions are used to help with biofilm formation and aggregation of its symbiont, *Vibrio fischeri*, prior to host colonization [89]. Representing basal metazoan evolution, the Hydra (*Hydra vulgaris*) maintains a mucus layer, termed the glycocalyx, where prospective symbionts are recruited and localized during early embryogenesis [90]. To this end, investigation of the potential of cyclical leech mucosal secretions to serve as a HT mechanism, complementing VT [85], and warrants further investigation.

Conclusions

Microbial symbioses are critical for the survival of the majority of organisms. In the same vein, transmission is vital for the formation and persistence of these symbioses. However, most symbioses are very complex, with many microbial species cohabitating and interacting simultaneously, making it difficult to unravel how particular microbes may be acquired and how the transmission mode utilized impacts the symbiosis. Therefore, naturally simple model systems, such as the medicinal leech digestive tract symbiosis, provide tractable models to

elucidate the basic principles underlying the formation of microbial symbioses. Here, I describe how a host-derived substrate (i.e. cyclical mucosal secretions) can be used as a vehicle for facilitating the HT of a crucial gut symbiont, *A. veronii*, to other leech conspecifics, thereby complementing imperfect VT through the cocoon [85]. These results shed light on features that enable MMT, a transmission mode that incorporates VT and HT, and is likely utilized in many symbioses, including those with higher microbial community complexity (e.g. mammals). Furthermore, I describe how these mucosal secretions serve as an alternative, albeit more stressful, niche for *A. veronii* through the examination of population dynamics and metatranscriptomic analyses post mucus shedding. By occupying this dual niche, *A. veronii* is subject to different selection pressures relative to symbionts that utilize strict VT, and likely encounters reduced genome erosion. Moreover, within host-generated mucus *A. veronii* is not limited to a symbiotic lifestyle and may be dispersed into various lifestyles beyond that of mutualism. Lastly, mucus also provides a residence for bacterial species not previously described within the leech, suggesting the existence of yet to be determined microbiota members within the leech likely contributing to host biology.

Questions addressed in this project include:

1. How may host biological processes be co-opted as a mechanism for HT of a bacterial symbiont?
2. What are the population dynamics and transcriptional profiles of symbiotic bacteria within vehicles that facilitate HT?
3. Are other previously described leech symbionts also located within these secretions, potentially indicating the broad use of this HT mechanism among other leech microbiota members?

4. Are there members of the mucosal microbiota that have not yet been characterized within the host, indicating a potential gap in our knowledge of leech symbionts?

REFERENCES

1. Frank AB. *Ueber die auf Wurzelsymbiose beruhende Ernahrung gewisser Baume dursh unterirdische Pilze.*. Ber. Deut. Bot. Ges., 1885. **3**:p. 128–45.
2. Anton de Bary H. *Die Erscheinung der Symbiose* 1879. Verlag von Karl J. Trubner, Strassburg.
3. McFall-Ngai MJ. *The development of cooperative associations between animals and bacteria: Establishing detente among domains*. Am. Zool., 1998. **38**:p. 593–608.
4. McFall-Ngai M. *Are biologists in "future shock"? Symbiosis integrates biology across domains*. Nat. Rev. Microbiol., 2008. **6**:p. 789–92.
5. Currie CR et al. *Ancient tripartite coevolution in the attine ant-microbe symbiosis*. Science, 2003. **299**:p. 386–8.
6. Dean WRJ, MacDonald IAW. *A review of african birds feeding in association with mammals*. Ostrich, 1981. **52**:p. 135–55.
7. Cote IM. *Evolution and ecology of cleaning symbioses in the sea*. Oceanogr. Mar. Biol., 2000. **38**:p. 311–55.
8. Bosch TC, McFall-Ngai MJ. *Metaorganisms as the new frontier*. Zool., 2011. **114**:p. 185–90.
9. Rosenberg E et al. *The evolution of animals and plants via symbiosis with microorganisms*. Env. Microbiol. Rep., 2010. **2**:p. 500–6.
10. Rio R V et al. *Strategies of the home-team: symbioses exploited for vector-borne disease control*. Trends Microbiol., 2004. **12**:p. 325–36.
11. Moya A et al. *Learning how to live together: genomic insights into prokaryote-animal symbioses*. Nat. Rev. Genet., 2008. **9**:p. 218–29.

12. Friedman RL. *Pertussis: the disease and new diagnostic methods*. Clin. Microbiol. Rev., 1988. **1**:p. 365–76.

13. Tuomanen E, Weiss A. *Characterization of two adhesins of Bordetella pertussis for human ciliated respiratory-epithelial cells*. J. Infect. Dis., 1985. **152**:p. 118–25.

14. Chiller K et al. *Skin microflora and bacterial infections of the skin*. J Investig Dermatology Symp. Proc., 2001. **6**:p. 170–4.

15. Williams RE. *Healthy carriage of Staphylococcus aureus: its prevalence and importance*. Bacteriol. Rev., 1963. **27**:p. 56–71.

16. Palmer TM et al. *Synergy of multiple partners, including freeloaders, increases host fitness in a multispecies mutualism*. Proc. Natl. Acad. Sci. U.S.A., 2010. **107**:p. 17234–9.

17. Beijerinck MW. *Über oligonitrophile Mikroben*. Cent. für Bakteriol. Parasitenkunde, Infekt und Hyg, 1901. **7**:p. 561–82.

18. Dubilier N et al. *Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm*. Nature, 2001. **411**:p. 298–302.

19. Mylona P et al. *Symbiotic Nitrogen Fixation*. Plant Cell, 1995. **7**:p. 869–85.

20. Frank SA. *Mutual policing and repression of competition in the evolution of cooperative groups*. Nature. **377**:p. 520–2.

21. Kiers ET et al. *Host sanctions and the legume-rhizobium mutualism.*. Nature, 2003. **425**:p. 78–81.

22. Lipsitch M et al. *The evolution of virulence in pathogens with vertical and horizontal transmission*. Evolution (N Y), 1996. **50**:p. 1729–41.

23. Bright M, Bulgheresi S. *A complex journey: transmission of microbial symbionts*. Nat. Rev. Microbiol., 2010. **8**:p. 218–30.

24. Giere O, Langheld C. *Structural organisation, transfer and biological fate of endosymbiotic bacteria in gutless oligochaetes.*. Mar. Biol., 1987. **93**:p. 641–50.

25. Martin R et al. *Human milk is a source of lactic acid bacteria for the infant gut*. J. Pediatr., 2003. **143**:p. 754–8.

26. Douglas AE. *Mycetocyte symbiosis in insects*. Biol. Rev. Camb. Philos. Soc., 1989. **64**:p. 409–34.

27. Warnecke F et al. *Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite*. Nature, 2007. **450**:p. 560–5.

28. Martinson VG et al. *Establishment of characteristic gut bacteria during development of the honeybee worker*. Appl. Env. Microbiol., 2012. **78**:p. 2830–40.

29. Adams DG, Duggan PS. *Cyanobacteria-bryophyte symbioses*. J. Exp. Bot., 2008. **59**:p. 1047–58.

30. Bull JJ et al. *Genetic details, optimization and phage life histories*. Trends Ecol. Evol., 2004. **19**:p. 76–82.

31. de Roode JC et al. *Strength in numbers: high parasite burdens increase transmission of a protozoan parasite of monarch butterflies (Danaus plexippus).* Oecologia, 2009. **161**:p. 67–75.

32. Ebert D. *The Epidemiology and Evolution of Symbionts with Mixed-Mode Transmission*. Annu. Rev. Ecol. Evol. Syst., 2013. **44**:p. 623–43.

33. Brandvain Y et al. *Horizontal transmission rapidly erodes disequilibria between organelle and symbiont genomes*. Genetics, 2011. **189**:p. 397–404.

34. Dunn AM, Smith JE. *Microsporidian life cycles and diversity: the relationship between virulence and transmission*. Microbes Infect., 2001. **3**:p. 381–8.

35. Moran NA et al. *Genomics and evolution of heritable bacterial symbionts*. Annu. Rev. Genet., 2008. **42**:p. 165–90.

36. Werren JH et al. *Wolbachia: master manipulators of invertebrate biology*. Nat. Rev. Microbiol., 2008. **6**:p. 741–51.

37. Frydman HM et al. *Somatic stem cell niche tropism in Wolbachia*. Nature, 2006. **441**:p. 509–12.

38. Vavre F et al. *Phylogenetic evidence for horizontal transmission of Wolbachia in host-parasitoid associations*. Mol. Biol. Evol., 1999. **16**:p. 1711–23.

39. Moran NA, Dunbar HE. *Sexual acquisition of beneficial symbionts in aphids*. Proc. Natl. Acad. Sci. U.S.A., 2006. **103**:p. 12803–6.

40. Wernegreen JJ. *Genome evolution in bacterial endosymbionts of insects*. Nat. Rev. Genet., 2002. **3**:p. 850–61.

41. O'Fallon B. *Population structure, levels of selection, and the evolution of intracellular symbionts*. Evolution (N Y), 2008. **62**:p. 361–73.

42. Bennett S. *Solexa Ltd*. Pharmacogenomics, 2004. **5**:p. 433–8.

43. Ronaghi M et al. *A sequencing method based on real-time pyrophosphate*. Science, 1998. **281**:p. 363-5.

44. Ronaghi M et al. *Real-time DNA sequencing using detection of pyrophosphate release*. Anal. Biochem., 1996. **242**:p. 84–9.

45. Nogge G. *Sterility in Tsetse Flies (Glossina morsitans Westwood) Caused by Loss of Symbionts*. Experientia, 1976. **32**:p. 995–6.

46. Heijtz RD et al. *Normal gut microbiota modulates brain development and behavior*. Proc. Natl. Acad. Sci. U.S.A., 2011. **108**:p. 3047–52.

47. Hsiao EY et al. *Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders*. Cell, 2013. **155**:p. 1451–63.

48. Round JL et al. *The Toll-Like Receptor 2 Pathway Establishes Colonization by a Commensal of the Human Microbiota*. Science, 2011. **332**:p. 974–7.

49. O'Hara AM, Shanahan F. *The gut flora as a forgotten organ*. EMBO Rep., 2006. **7**:p. 688–93.

50. Shin SC et al. *Drosophila microbiome modulates host developmental and metabolic homeostasis via insulin signaling*. Science, 2011. **334**:p. 670–4.

51. Koropatnick TA et al. *Microbial factor-mediated development in a host-bacterial mutualism*. Science. **306**:p. 1186–8.

52. Turnbaugh PJ et al. *The human microbiome project*. Nature, 2007. **449**:p. 804–10.

53. Backhed F et al. *Host-bacterial mutualism in the human intestine*. Science, 2005. **307**:p. 1915–20.

54. Eckburg PB et al. *Diversity of the human intestinal microbial flora*. Science, 2005. **308**:p. 1635–8.

55. Qin JJ et al. *A human gut microbial gene catalogue established by metagenomic sequencing*. Nature, 2010. **464**:p. 59–U70.

56. Ley RE et al. *Ecological and evolutionary forces shaping microbial diversity in the human intestine*. Cell, 2006. **124**:p. 837–48.

57. Sawyer RT. *Leech Biology and Behaviour* 1986. Clarendon Press, Oxford, United Kingdom.

58. de Eguileor M et al. *Leech immune responses: contributions and biomedical applications*. A New Model Anal. Antimicrob. Pept. with Biomed. Appl., 2002.

59. 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**:p. 1–7.

60. Kikuchi Y et al. *Stratified bacterial community in the bladder of the medicinal leech, Hirudo verbana*. Env. Microbiol., 2009. **11**:p. 2758–70.

61. Graf J et al. *Leeches and their microbiota: naturally simple symbiosis models*. Trends Microbiol., 2006. **14**:p. 365–71.

62. Worthen PL et al. *Culture-independent characterization of the digestive-tract microbiota of the medicinal leech reveals a tripartite symbiosis*. Appl. Env. Microbiol., 2006. **72**:p. 4775–81.

63. Bomar L et al. *Directed culturing of microorganisms using metatranscriptomics*. mBio, 2011. **2**:p. e00012–11.

64. Nelson M et al. *Mucinivorans hirudinis gen. nov., sp. nov., an anaerobic, mucin-degrading bacterium isolated from the digestive tract of the medicinal leech, Hirudo verbana.* Int. J. Syst. Evol. Microbiol., 2015.

65. 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. Env. Microbiol., 2007. **73**:p. 1984–91.

66. Graf J. *The effect of symbionts on the physiology of Hirudo medicinalis, the medicinal leech*. Invert. Reprod. Dev., 2002. **41**:p. 269–75.

67. Janda JM, Abbott SL. *The genus Aeromonas: taxonomy, pathogenicity, and infection*. Clin. Microbiol. Rev.. **23**:p. 35–73.

68. Sartor C et al. *Nosocomial Infections with Aeromonas hydrophila from Leeches*. Clin.

Infect. Dis., 2002. **35**:p. E1–5.

69. Maltz M, Graf J. *The type II secretion system is essential for erythrocyte lysis and gut colonization by the leech digestive tract symbiont Aeromonas veronii*. Appl. Env. Microbiol., 2011. **77**:p. 597–603.

70. Silver AC et al. *Interaction between innate immune cells and a bacterial type III secretion system in mutualistic and pathogenic associations*. Proc. Natl. Acad. Sci. U.S.A., 2007. **104**:p. 9481–6.

71. Minana-Galbis D et al. *The reference strain Aeromonas hydrophicla CIP 57.50 should be reclassified as Aeromonas salmonicida CIP 57.50*. Int. J. Syst. Evol. Microbiol., 2010. **60**:p. 715–7.

72. Minana-Galbis D et al. *Proposal to assign Aeromonas diversa sp. nov. as a novel species designation for Aeromonas group 501*. Syst. Appl. Microbiol., 2010. **33**:p. 15–9.

73. Alperi A et al. *Aeromonas taiwanensis sp. nov. and Aeromonas sanarellii sp. nov., clinical species from Taiwan*. Int. J. Syst. Evol. Microbiol., 2010. **60**:p. 2048–55.

74. Alperi A et al. *Aeromonas fluvialis sp. nov., isolated from a Spanish river*. Int. J. Syst. Evol. Microbiol., 2010. **60**:p. 72–7.

75. Beaz-Hidalgo R et al. *Aeromonas piscicola sp nov., isolated from diseased fish*. Syst. Appl. Microbiol., 2009. **32**:p. 471–9.

76. Figueras MJ et al. *Aeromonas rivuli sp. nov., isolated from the upstream region of a karst water rivulet*. Int. J. Syst. Evol. Microbiol., 2011. **61**:p. 242–8.

77. Siddall ME et al. *Bacterial symbiont and salivary peptide evolution in the context of leech phylogeny*. Parasitology, 2011. **138**:p. 1815–27.

78. Laufer AS et al. *Characterization of the digestive-tract microbiota of Hirudo orientalis, a*

european medicinal leech. Appl. Env. Microbiol., 2008. **74**:p. 6151–4.

79. Kaneuchi C, Mitsuoka T. *Bacteroides microfusus, a New Species from Intestines of Calves, Chickens, and Japanese Quails*. Int. J. Syst. Bacteriol., 1978. **28**:p. 478–81.

80. Ley RE et al. *Obesity alters gut microbial ecology*. Proc. Natl. Acad. Sci. U.S.A., 2005. **102**:p. 11070–5.

81. Rautio M et al. *Bacteriology of histopathologically defined appendicitis in children*. Pediatr. Infect. Dis. J., 2000. **19**:p. 1078–83.

82. Kutschera U, Roth M. *Cocoon deposition and cluster formation in populations of the leech Hirudo verbana (Hirudinea: Hirudinidae)*. Lauterbornia, 2006. **56**:p. 5–8.

83. Fernandez J, Stent GS. *Embryonic development of the hirudinid leech Hirudo medicinalis: structure, development and segmentation of the germinal plate*. J. Embryol. Exp. Morphol., 1982. **72**:p. 71–96.

84. Reynolds SA et al. *Staging of middle and late embryonic development in the medicinal leech, Hirudo medicinalis*. J. Comp. Neurol., 1998. **402**:p. 155–67.

85. Rio R V et al. *Symbiont succession during embryonic development of the European medicinal leech, Hirudo verbana*. Appl. Env. Microbiol., 2009. **75**:p. 6890–5.

86. Chen X et al. *Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus Glossina and its bacteriome-associated endosymbiont, Wigglesworthia glossinidia*. J. Mol. Evol., 1999. **48**:p. 49–58.

87. Silver AC et al. *Complex evolutionary history of the Aeromonas veronii group revealed by host interaction and DNA sequence data*. PLoS One, 2011. **6**:p. e16751.

88. Michalsen A et al. *Medicinal Leech Therapy* 2007. Thieme Medical Publishers, New York, New York.

89. Nyholm S V et al. *Roles of Vibrio fischeri and nonsymbiotic bacteria in the dynamics of mucus secretion during symbiont colonization of the Euprymna scolopes light organ*. Appl. Env. Microbiol., 2002. **68**:p. 5113–22.

90. Fraune S et al. *In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides*. Proc. Natl. Acad. Sci. U.S.A., 2010. **107**:p. 18067–72.

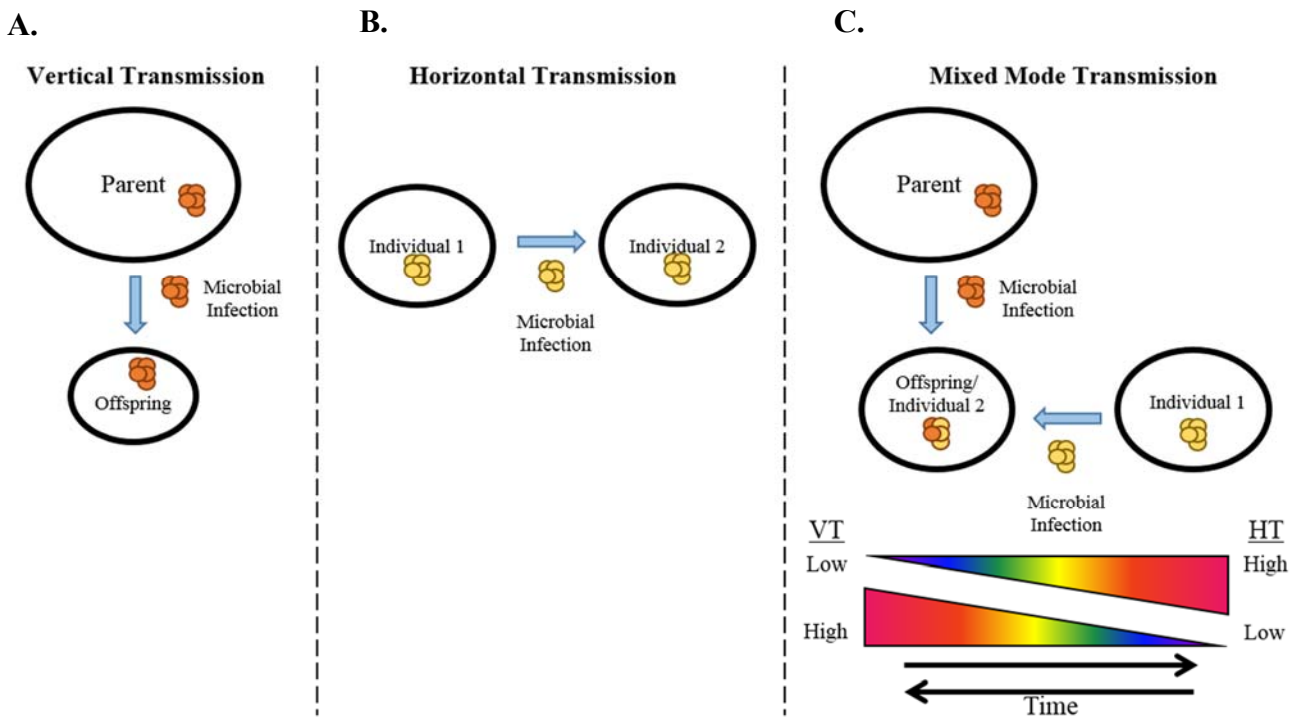

Figure 1-1. The three general mechanisms of microbial symbiont transmission. A. Vertical transmission (VT), the transmission of a microbial infection from parent to offspring. B. Horizontal transmission (HT), the transmission of a microbial infection from one individual to another, irrespective of relation and usually originating in the environment. C. Mixed mode transmission (MMT), which integrates A. and B., generally results in a more genetically diverse microbial infection with a higher fidelity than what may arise through HT solely. MMT can incorporate varying proportions of VT and HT within a single symbiotic association (i.e. host and transmitted microbe) through evolutionary time, as shown by the gradient triangles.

CHAPTER 2: Hitchhiking of host biology by beneficial symbionts enhances transmission[1]

ABSTRACT

Transmission plays a key role in the evolution of symbiosis. Mixed mode transmission combines horizontal and vertical mechanisms for symbiont acquisition. However, features that enable mixed transmission are poorly understood. Here, we determine the mechanistic basis for the recruitment of the beneficial bacterium, *Aeromonas veronii* by the leech, *Hirudo verbana*. We demonstrate that host mucosal secretions complement imperfect symbiont vertical transmission. First, we show that the *A. veronii* population within secretions originates from the host digestive tract and proliferates synchronously with shedding frequency, demonstrating the coupling of partner biology. Furthermore, leeches are attracted to these castings with oral contact proving sufficient for symbiont transmission. Leech attraction to mucus is not affected by the symbiont state of either the host or mucus, suggesting that *A. veronii* exploits preexisting host behavior and physiological traits. A dual transmission mode, integrating multiple layers of host contributions, may prove evolutionarily advantageous for a wide range of symbioses. Using such a strategy, host infection is ensured, while also providing access to a higher genetic diversity of symbionts. Countless host-associated microbes exhibit mixed mode transmission, supporting the use of the leech symbiosis as a model for enhancing our understanding of the specificity, establishment and persistence of microbiotas.

INTRODUCTION

Transmission is instrumental for the colonization and persistence of microbial symbioses [1]. Stable beneficial relations predominantly rely on vertical transmission, often using

[1] Reprinted from Ott BM, Cruciger M, Dacks AM, Rio RVM. Hitchhiking of host biology by beneficial symbionts enhances transmission. Sci. Rep., 2014. **4**, 5825; DOI:10.1038/srep05825.

mechanisms that tap into unique aspects of host reproductive biology. With time, microbial genomes can evolve to complement host biology by eliminating loci involved with extraneous physiological processes. As a trade-off, this genome reduction tends to decrease free-living capabilities and the ability to respond to environmental change (reviewed in [2]). In contrast, horizontal transfer of mutualists [3] requires partner contact and/or infection through the environment. The stochastic nature of partner encounters suggests that a horizontal strategy is consequently less evolutionarily stable [4], as the horizontal transmission of a genetically diverse symbiont population tends to eliminate congruence with partner phylogenetic trees [5] (reviewed in [4]). A mixed strategy, blending transmission modes, may also occur. While mixed strategies have been previously considered to be transitory in nature, with fitness trade-offs[6] selecting for the evolutionarily preferred mode [7], it is possible that the mixed strategy can be the more advantageous scenario particularly if the transmission modes are not exclusive of one another [4, 8]. However, the adaptive features that may enable the persistence of an evolutionary favorable mixed strategy are not well understood.

The medicinal leech, *Hirudo verbana,* uses a mixed strategy for the transmission of a relatively simple microbiota within its digestive tract [9], yet the mechanistic basis remains unknown. This microbiota is dominated by the Gammaproteobacteria *Aeromonas veronii* and the Bacteroidia *Rikenella*-like residents [10, 11], which have been proposed to play a wide range of host fitness enhancing roles (reviewed in [12]). *A. veronii* exhibits diverse lifestyles, as a free-living waterborne microbe and as both a mutualist and a pathogen depending on the host background [13-16], while members of the Rikenellaceae are found in the guts of a diverse spectrum of animals [17-20], indicating a high ecological versatility. *A. veronii* is a pioneer symbiont (i.e. early establishing member of the microbiota) in the crop of *H. verbana* [9], likely

influencing the recruitment of later colonizing symbionts, including *Rikenella*-like bacteria. The *A. veronii* symbiont can be vertically transmitted during development within the leech cocoon [9], yet, only a proportion of juveniles harbor *A. veronii* upon cocoon emergence, and none of the juveniles harbor the *Rikenella*-like symbiont. Despite this, two-weeks post cocoon emergence, all juvenile leeches harbor both symbiont species [9], suggesting that some form of horizontal transmission complements imperfect maternal transmission.

Although distinct *Aeromonas spp.* infect the digestive tracts of medicinal leech species [10, 11, 21], displaying tight host-symbiont species specificity, signatures of co-evolution between *A. veronii* and its *H. verbana* host are lacking [22]. This lack of congruence in the phylogenies is in stark contrast to bacterial partners that are vertically transmitted with high fidelity such as the ancient intracellular endosymbionts of some insects (reviewed in [2, 23]). High rates of gene flow within *A. veronii* populations have been described [24], and the presence of several mobile genetic elements, specifically prophages and insertion elements, have been identified within the genome of a leech associated *A. veronii* isolate, HM21 [25], arguing against strict maternal transmission. Such features are lacking in many purely heritable symbionts due to reductive genome evolution within the host (reviewed in [2]). The presence of these genomic features, coupled with incomplete vertical transmission and the absence of deep host-symbiont co-evolution, suggests that symbiont acquisition may also occur from an unknown extrinsic source following cocoon eclosion.

In this study, we identify the mechanistic basis for the mixed transmission strategy of a beneficial symbiont. Specifically, we demonstrate how periodic shedding of leech mucus serves as a naturally occurring substrate for *A. veronii* horizontal transfer to conspecific leech individuals. Leech mucosal secretions consist of glucosaminoglycans produced by globose

glands irregularly distributed throughout the epidermis [26, 27] and have been suggested to influence an array of biological functions such as protection from desiccation, respiration and conspecific recognition [26, 27]. These are shed by the host at regular intervals, always from the anterior to posterior end, thus providing consistent contact with the digestive tract via the fecal matter. This made mucus a potential candidate to harbor gut symbionts. In support, mucosal secretions provide a substrate for symbiont aggregation and recruitment for many other aquatic animals [28-30]. As such, we examined the potential for mucosal castings to harbor and transmit viable digestive tract symbionts, thus serving as a mechanistic basis for the horizontal component of a mixed transmission strategy.

We report that mucosal secretions harbor proliferating populations of *A. veronii* symbionts. Moreover, leeches are attracted to these secretions, which prove sufficient for horizontal transmission of *A. veronii*. Although symbiont-state did significantly influence leech host exploratory behavior, it did not affect attraction to mucosal castings, suggesting that mutualists are exploiting host behavior and physiological processes that likely evolved for another function prior to the symbiosis. Consistent with this hypothesis, symbiont-infection of mucus did not affect attraction, suggesting that intrinsic properties of the mucus could attract conspecifics, serving to drive host-microbe specificity. These findings suggest that a mixed transmission strategy, exploiting unique facets of host biology, may prove fitness savvy by circumventing trade-offs between transmission modes while also increasing the likelihood of partner encounter for these relatively non-social animals. Remarkably, this strategy may also provide an accessible pool of symbionts with greater genetic flexibility to accommodate environmental changes, while providing alternative lifestyles to *A. veronii* beyond that of mutualism.

MATERIALS AND METHODS

(a) Leech husbandry. Wild type (WT) *H. verbana* leeches were obtained from Leeches USA (Westbury, NY, USA) and maintained in sterilized, Leech Strength Instant Ocean H_2O (0.004% IO, LSIO) in the Department of Biology at West Virginia University at either 15°C or 23°C. Leeches were maintained on defibrinated bovine blood (Hemostat, CA). Aposymbiotic (APO) leeches, in this study referred to leeches lacking their *A. veronii* gut symbiont, were generated through the administration of at least one blood meal containing an antibiotic cocktail of 15 µg/mL of kanamycin (Km) and 10 µg/mL of rifaximin (Rif).

(b) Mucosal secretion sampling. Tanks were cleaned to remove all mucus and fresh LSIO added (denoted as Day 0). Mucosal samples gathered the following day were considered 1 d old or were aged at the temperatures at which they were extracted for 3, 5 or 8 d within LSIO. All samples were snap frozen at -80°C until further processing. To determine host mucosal secretion rate, individual leeches were maintained in LSIO, with daily examination for the shed castings.

(c) Symbiont population dynamics. The population dynamics of *A. veronii* symbionts within host mucus at 1, 3, 5 and 8 d following secretion were determined by utilizing real-time quantitative PCR (q-PCR). DNA isolation from mucus was performed following a Holmes-Bonner protocol [59]. Analyses were performed in an iCycler iQ Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA) using Bio-Rad SsoFast EvaGreen Supermix, 10 mM of primers (either QT-AvgyrB For and QT-AvgyrB Rev q-PCR, for the amplification of *A. veronii* (Table 1)), and 2 µL of DNA template as described [9]. Quantification of the amplicons relative to standard curves was performed using Bio-Rad CFX Manager software v 2.0. The respective DNA concentrations (ng/ul) of samples were used for the normalization of copy numbers. All assays were performed in triplicate and replicates were averaged for each sample.

(d) Behavioral testing. Behavioral assays, comparing baseline locomotion between WT and APO leeches, were performed by placing leech individuals within tanks (44.5 cm X 36.2 cm) filled with 10 L of sterile LSIO. Fresh LSIO and a new sterilized tank were provided for each leech during behavioral testing. Behavior was recorded for 10 min with a Casio Exilim 60 FPS Digital Camera (Ex-F1) and video footage processed using VirtualDub v.1.9.11 (Avery Lee, Informer Technologies). Leech movement was tracked using CTrax v.0.3.14[60] and analyzed using custom MatLab v.8.1.0.604 scripts (MathWork, Natick, MA, USA). The tanks were divided into 10x10 grids *in silico* and heat maps depicting the proportion of time spent in each grid unit of the experimental tank were generated. Experiments were performed in the same two-hour time frame every day to eliminate any temporal bias and all assays were performed and analyzed with the observer blind to symbiont status of the individual leeches and the presence/absence of symbionts within mucus castings.

Three behavioral assays were performed. The goal of the first assay was to determine if symbiont status (i.e. WT or APO) affected baseline locomotion of leeches. The average speed and relative proportion of time spent by leeches in the empty behavioral arena was calculated over the ten-minute testing period. The leeches were then rested for one week. The goal of the second behavioral assay was to determine if leeches were attracted to mucosal secretions and if there were any differences in level of attraction based on symbiont-state. Mesh sacs (#9013, Seattle Fabrics Inc., Seattle, WA), either empty or containing WT mucus, were suspended from opposite sides of the behavioral arena and the behavior of WT or APO leeches observed over a 10 minute period. These sacs served the purpose of containment, to control the location of mucus in our experiments, and to facilitate visualization as the mucus is translucent. It should be noted that this mesh is porous, enabling leeches to still make physical contact. The sides containing the

empty or mucus mesh balls were switched between individual trials to eliminate any potential side bias. The objective of the third behavioral assay was to determine if the presence of symbionts in the mucosal castings affected attraction to the mucus. WT leeches were placed in the behavioral arena with mesh sacs containing mucus from either WT or APO leeches. The sides containing the WT or APO mucus mesh balls were switched between leech individuals to eliminate any potential side bias.

(e) Transmission assays. To assess whether symbionts introduced into the digestive tract of *H. verbana* were capable of establishing within mucosal secretions, APO and WT leeches were fed 5 mL of defibrinated bovine blood supplemented with either a low (250 CFU/mL) or high (20,000 CFU/mL) inoculum of a constitutively *gfp*-expressing *A. veronii* mutant, HM21S::Tn7gfp (kindly provided by the Graf lab, Univ. of Connecticut). To assess both *A. veronii* transmission and stability within mucus through time, the first mucus and a mucus shed approximately sixty days following feeding were obtained and examined for the presence of HM21S::Tn7gfp by homogenizing and spreading the samples on selective nutrient agar plates (LB + Km (100 µg/mL) + Streptomycin (Str; 100 µg/mL)) incubated overnight at 30°C. Plates were then screened for *gfp*-expressing colonies using a Nikon Eclipse Ti[1] inverted fluorescence microscope and images were taken using NIS Elements Imaging Analysis Software. Single colonies were confirmed to be *A. veronii* through PCR using *gyrB* primers (Table 1) and sequencing of amplicons.

A second form of transmission assay was used to assess whether mucus can serve as natural mechanism for symbiont acquisition. To accomplish this, APO leeches were placed in jars containing mucosal castings in autoclaved mesh bags (see above) for seven days. The mucosal samples used in these studies were isolated from APO leeches, verified free of *A.*

veronii through PCR, and were subsequently inoculated with either high or low HM21S::Tn7gfp concentrations (see above). Following the exposure period, leeches were rinsed in sterile LSIO to ensure removal of any residual mucus. Leeches were subsequently placed in autoclaved jars with sterile LSIO and their first mucosal casts were examined for the presence of HM21S::Tn7gfp as described above.

(f) Statistical analyses. The data were analyzed using JMP 10 (SAS Institute Inc., Cary, NC, USA) software. Statistically significant differences in symbiont densities within secreted mucus through time were determined by performing F-tests, followed by the appropriate *t*-test (Mann-Whitney or Student's). To compare frequency of *A. veronii* colonies expressing *gfp* within mucus following oral administration, Pearson's chi-squared tests were used. ANOVAs were performed to determine if leeches demonstrated side preference in experimental arenas during behavioral assays. Mean APO and WT leech speed (mm traveled/sec) was determined and compared using a two-tailed Student's *t*-test. Lastly, the exploratory behavior of APO and WT individuals was measured by calculating total time spent in the inner 6x6 matrix (denoting the center of the arena) and compared between the two groups with a Student's *t*-test.

RESULTS

Periodically shed mucosal secretions harbor proliferating microbial symbionts.

To determine if leech mucosal secretions (Fig. 1A) house viable *A. veronii*, we assayed the population density of *A. veronii* using quantitative PCR of the single copy *gyrB* gene locus from mucus samples at two biologically relevant temperatures representing a pond's edge in *H. verbana*'s natural geographical distribution [31], where breeding between leeches typically occurs in summer months (~23°C, [26]), and a pond's base (~15°C). At both of these temperatures, *A. veronii* population densities peaked on day 3 and decreased thereafter (Fig. 1B).

Although *A. veronii* harbored within mucus at the two temperatures demonstrated similar growth profiles (Fig. 1B), significant differences in overall density were observed (p=0.0086, one-way ANOVA; n=77) with mucus shed at 23°C harboring an overall higher symbiont load over the examination period. With time, *A. veronii* abundance decreased, indicating either an exhaustion of mucosal resources and/or a dispersal of bacteria from the mucus. In addition, symbiont proliferation peaked approximately when mucus was again secreted by the leech host (Fig. S1), thereby ensuring a continuous supply of seeded mucosal secretions within the environment. Thus, leech gut symbionts are not only viable, but also proliferate in mucus following its secretion in concert with host shedding patterns.

Digestive tract A. veronii seed mucosal castings and are sufficient for symbiont transmission.

Mucus is shed from the leech posterior (Fig. 1A), enabling the gut symbionts to potentially seed the secretion through the excretory route. This observation implicates the digestive tract as the origin of the mucosal *A. veronii* population. To verify this hypothesis, the castings of leeches that had been cleared of their *A. veronii* gut symbiont via antibiotic treatment (aposymbiotic or APO leeches) were collected and verified to lack *A. veronii* (Fig. 2A). This observation suggests that digestive tract *A. veronii* are the likely source of the mucosal symbiont population. To further confirm that gut symbionts are transferred to mucus, both APO and WT leeches were fed a *gfp*-expressing *A. veronii* strain at high (2 x 10^4 CFU/mL) and low (2.5 x 10^2 CFU/mL) inoculums, which flank the natural *A. veronii* population levels found within the mucus (Fig.1B). Additionally, to determine if gut symbionts persistently seed the mucosal castings of the host, the first casting and a casting 60 days post-feeding (representing ~ 30^{th} mucosal cast since feeding) were collected and spread on nutrient agar plates and subsequently visualized under fluorescent microscopy (Fig. 2B shows a colony obtained from a first mucosal

casting), with their identity confirmed using *A. veronii* specific *gyrB*-primers (Fig. S2A). APO and wildtype (WT) leeches secreted their mucosal castings at similar temporal intervals (data not shown). Only oral administration of *gfp*-expressing *A. veronii* at the high inoculum resulted in detectable concentrations within subsequently shed mucus (Fig. 2C; CFU/ml grouped across first and 30th castings for APO and WT, p=0.282, Mann Whitney *U*-test, n=24). With this high oral administration, the percentage of leeches with *A. veronii* symbionts present in the first and ~30th casting did not differ (Fig. S2B; p=0.3865, Pearson's chi-squared test, n=24). Furthermore, *A. veronii* concentrations did not differ between first and ~30th castings (p=0.1642, Student's *t*-test, n=11, data not shown) and there was no effect of leech symbiont state (Fig. S2C; p=0.7595, Pearson's chi-squared test, n=24). These results demonstrate that digestive tract *A. veronii* establish very rapidly and persistently infect host secretions through time, regardless of leech symbiont state, although the specific route used for colonization of mucus remains unknown.

Although the gut microbiota is capable of seeding the mucosal castings, transmission of these microbes is unlikely to occur through an infected blood meal, as the inoculum loads needed to ensure oral transmission are too high to be encountered. Furthermore, the levels of *Aeromonas* in the aquatic environments inhabited by leeches are >1 CFU/ml (colony forming unit per milliliter of water) [32]. Therefore, it would be highly unlikely for free-living *Aeromonas* to serve as an inoculum source, as a concentration of 10^2 CFU/ml was insufficient for transmission (Fig. 2C). To determine if mucus is capable of horizontally transferring the *A. veronii* symbiont, APO mucosal secretions were inoculated with high and low densities (as above) of the *gfp*-expressing *A. veronii*. These secretions were placed in mesh sacs within individual jars housing an APO leech for seven days. Following this incubation period, leeches were surface-rinsed and moved into new autoclaved jars with sterile LSIO and their first mucosal casts were examined

for the presence of *gfp*-expressing *A. veronii*. Oral contact with shed mucus proved sufficient to transmit the *A. veronii* symbiont at both inoculum levels (Fig. 2D), with no significant difference in the resulting CFU/mL (p=0.92, Student's t-test, n=6) or frequency of individuals containing symbionts within their castings (i.e. 2 of 3 individuals at both concentrations). In multiple invertebrate systems, particularly with regard to vector-transmitted organisms, the natural vector provides a higher efficiency of microbial transmission relative to artificial introduction of the microbe [33-40]. Consistent with this, mucosal exposure at both inoculum levels proved more effective than blood meal *A. veronii* administration at seeding subsequent mucosal casts, supporting mucosal transfer as a natural and efficient method of microbial transmission between leeches.

Leeches display attraction to mucus shed by conspecifics irrespective of host symbiont state.

After shedding, mucosal secretions capable of transmitting the *A. veronii* symbiont may float through the water or become caught on debris. Therefore, for mucus to be an effective substrate for transmission, leeches must seek out these castings to enable the transfer of microbes. Here, we sought to determine if leeches are attracted to mucus. Furthermore, because symbiont status affects host behavior in other systems [41-43], we sought to determine if symbiont-state altered any mucus-oriented behavior. We began by examining the influence of symbiont-state on the baseline behavior of leeches to determine what effects, if any, the antibiotic treatment had on locomotion. APO and WT leeches were placed into an experimental arena (Fig. 3A) and their movements within the arena tracked for ten minutes (Fig. 3B). Heat maps indicating the proportion of time leech subjects spent within an overlaid 10 x 10 grid (Fig. 3C) were generated. No significant difference was found between WT and APO leeches in average speed (Fig. S3) (p=0.36, Student's t-test, n=33), indicating that antibiotics do not affect

the ability of the leech to move. In general, leeches were positively thigmotactic (Fig. 3C), preferring contact with the corners and edges of the tank. Interestingly, the two experimental groups did differ in the proportion of time spent away from the edge of the arena (Fig. 3D). APO leeches spent significantly more time in the center of the arena (Fig. 3E; p=0.017, Student's t-test, n=33). A willingness to enter the center of an arena is classically associated with "risky" exploratory behavior in vertebrates and invertebrates [43-47], suggesting that leeches lacking *A. veronii* in their digestive tract produce more exploratory behavior relative to WT leeches.

We hypothesized that the increase in exploratory behavior would raise the likelihood that leeches would encounter the mucosal castings of conspecifics. We therefore examined the behavior of WT and APO leeches in an arena containing two mesh bags: one containing a WT mucus sample and a second empty bag serving as a control (Fig. 4A). The movement of individual leeches within the arena was tracked (Fig. 4B), and the average time distribution of individual leech movement within the experimental arena was determined (Fig. 4C). To investigate leech preference for mucus, we determined the amount of time spent in each of four 2X1 sub-regions of 10x10 grid (Fig. 4C; overlaid *in silico* as above). These regions contained the mucus sac, the empty sac, or one of two control regions on the top and bottom sides of the arena (black boxes; Fig. 4C). Leeches in general exhibited a strong preference for the region containing the mucus filled sac relative to the empty sac (Figs. 4D and 4E) (p=0.0024, one-way ANOVA, n=37), indicating that leeches are attracted to the mucus of conspecifics. Leeches make frequent, repeated oral contact with the mucus in both our colony tanks and during experiments when the mucus is in the mesh sacs, which is consistent with previous reports that suggest that leeches use mucus for conspecific recognition [27]. However, WT and APO leeches did not differ in the amount of time spent in proximity to the mucus (Fig S4A; p=0.609, Student's t-test, n=36), or in

their latency to first investigate the mucus (Fig S4B; p=0.78, Student's t-test, n=23). Therefore, we conclude that shed mucus attracts leeches regardless of host symbiont state, providing an efficient mechanism for transmission of *A. veronii* between leeches.

A. veronii infection of mucus does not impact host attraction.

The lack of influence of host symbiont state on attraction to mucus suggested that *A. veronii* is exploiting a preexisting host physiological process most likely derived for another function separate from symbiosis. Alternatively, it is possible that the attraction of leeches to the mucosal castings of conspecifics depends upon the presence of symbionts in the mucus. To determine if the presence of symbionts in the mucus affects host attraction, WT leeches were placed in the behavior arena with two mesh bags; one containing mucus from an APO leech and one containing mucus from a WT leech (Fig. 5A). No significant difference was observed in the time spent in the vicinity of WT versus APO mucus samples (Fig. 5B; p=0.46, one-way ANOVA, n=18), indicating that an inherent attraction to mucosal secretions occurs between conspecific leeches (Fig. 4C), regardless of *A. veronii* infection. This result is consistent with the suggestion that *A. veronii* likely exploited this preexisting host biological trait.

DISCUSSION

We have shown that leech mucus can provide a hospitable environment for *A. veronii* and can serve as a mechanism for horizontal transmission through its ability to attract conspecific leeches. Seed populations of *A. veronii* in the mucus originate from the digestive tract and mucosal administration of *A. veronii* is sufficient for transmission of the symbiont to novel leech hosts. Although oral administration at high inoculum in blood also enabled transmission, the higher efficiency of symbiont transfer via mucus and the improbability that a leech would encounter prey with such high *A. veronii* blood infections, implicate mucus as the natural

transmission vehicle. Furthermore, to serve as an effective transmission substrate, leeches would have to actively seek out the mucus and consistent with this we found that leeches are attracted to the mucosal secretions of conspecifics.

Since different *Hirudo spp*. overlap geographically [31], but still maintain a high degree of *Aeromonas spp*. selectivity [22], we would predict attraction to the mucus of conspecifics over mucus of other leech species. A similar scenario arises with the strict feeding preferences of different species of *Anopheles* mosquitos that enable the tight coupling with species of the *Plasmodium* protozoa [48-50], suggesting that host ecological preferences can have a significant impact on microbial symbioses. A similar host preference for the mucus of conspecifics could drive *Aeromonas* symbiont selectivity. Alternatively, leech-specific differences in mucosal content may support one species of *Aeromonas* over another.

As in mice [43], leeches lacking gut symbionts displayed more exploratory behavior, which we hypothesized would increase their likelihood of encountering the mucus of conspecifics. However, the symbiont status of both the leech and the mucus itself do not have an impact on the time leeches spend near the mucus, indicating that leech attraction to mucus likely preceded the use of mucus for symbiont transmission. However, through this hitchhiking mechanism, the leech is also likely ensured access to an initial and more genetically diverse supply of this necessary symbiont.

The success of the mixed transmission strategy employed by *H. verbana* and *A. veronii* relies on host contribution, specifically an extrinsic substrate synthesized by the host, which is capable of harboring gut symbionts at concentrations that do not occur while *A. veronii* is free-living. Host-generated mucus has been shown to facilitate symbiont recruitment in a number of aquatic animals. For instance, hydra (*Hydra vulgaris*) maintains a mucus layer, called the

glycocalyx, where prospective symbionts are recruited and localized during early embryogenesis [51]. In *Euprymna scolopes*, the Hawaiian bobtail squid, mucus enables biofilm formation and the aggregation of its light organ symbiont, *Vibrio fischeri*, prior to the initiation of the symbiosis from a complex milieu of microbes [29]. Horizontal transmission in other model systems requires either direct physical contact, such as sexual transmission, or a vehicle to connect partners (reviewed in [1]).

Horizontal transmission involves distinct tradeoffs relative to a free-living lifestyle. For example with horizontal transmission, energy is typically required for processes such as chemotaxis, motility and adherence to ensure appropriate host encounter and colonization. These costs result in transcriptional profile differences, and potentially investment in structural features (i.e. flagella, pili, etc.) that facilitate horizontal transmission. Consequently, we suggest for a mixed strategy to persist, a host-derived substrate must exist which concentrates symbionts and perhaps even attracts other potential hosts, as in the case of leech mucosal castings, so as to maximize the likelihood of transmission and, theoretically, lessen the cost associated with horizontal transmission.

In recent years, mixed transmission strategies have been recognized as more prevalent than originally anticipated (reviewed in [4]). Many symbionts previously thought to be exclusively vertically transmitted may also be transferred horizontally, albeit at low rates [5, 52-56]. In fact, symbionts that are exclusively vertically transmitted are rare and currently limited to the obligate mutualists of certain invertebrate taxa (reviewed in [1, 2]). A variety of circumstances could factor against exclusive vertical transmission. Population bottlenecks during vertical transmission [57, 58] and a lack of available horizontal gene transfer opportunities [4] result in reduced genetic variation within the symbiont population (reviewed in [2]), limiting any

potential adaptation to a changing environment. Importantly, environmental changes to which the host cannot adapt will also result in an evolutionary dead end for its obligate symbiont. Therefore a mixed transmission strategy, which has traditionally been considered transitory in nature [7], can provide persistent benefits to both host and symbiont, and thereby be positively correlated and selected for. While vertical transmission guarantees that host offspring harbor symbionts, additional complementation via horizontal transmission serves to increase genetic variability of its microbiota, which can enhance the ability of the host to cope with a dynamic environment. However, the degree of genetic variability provided to leech symbionts with mixed relative to vertical modes of transmission is unknown. Furthermore, the microbe also benefits due to lifestyle opportunities other than mutualism.

It is clear that *H. verbana* uses a mixed transmission mode for its pioneer symbiont, *A. veronii*. It remains to be seen however, the degree to which a mixed mode of transmission represents a snapshot in the transition from horizontal to vertical transmission, or if the system represents a stable strategy. Additionally, it remains to be seen how conspecific attraction to mucus may contribute to the high specificity of *Aeromonas* and leech associations, as *Hirudo* spp. do geographically overlap [31]. Survival in these secretions could be a mechanism to select for the appropriate symbionts. Similar investigations into the transmission of additional leech symbionts using this mechanism would also be critical in understanding the importance of mucus towards the continuance and stability of the host microbiota. Future work should also explore the role of mixed mode of transmission in the succession of various members of the microbiota within hosts, as many members of the microbiota are thought to be transmitted as species assemblages [4]. It is possible that a mixed mode of transmission benefits both host and

symbiont sufficiently to negate antagonistic selective pressures between vertical and horizontal transmission, particularly if there is considerable host contributions with both modes.

REFERENCES

1. Bright, M. and S. Bulgheresi, *A complex journey: transmission of microbial symbionts.* Nature Rev. Microbiol., 2010. **8**(3): p. 218-230.
2. Moran, N.A., J.P. McCutcheon, and A. Nakabachi, *Genomics and evolution of heritable bacterial symbionts.* Annu. Rev. Genet., 2008. **42**: p. 165-190.
3. Sachs, J.L., R.G. Skophammer, and J.U. Regus, *Evolutionary transitions in bacterial symbiosis.* Proc. Natl. Acad. Sci. U.S.A., 2011. **108 Suppl 2**: p. 10800-7.
4. Ebert, D., *The Epidemiology and Evolution of Symbionts with Mixed-Mode Transmission.* Annu. Rev. Ecol. Evol. Syst., 2013. **44**: p. 623-643.
5. Brandvain, Y., C. Goodnight, and M.J. Wade, *Horizontal transmission rapidly erodes disequilibria between organelle and symbiont genomes.* Genetics, 2011. **189**(1): p. 397-404.
6. Sachs, J.L., J.E. Russell, and A.C. Hollowell, *Evolutionary instability of symbiotic function in Bradyrhizobium japonicum.* PLoS One, 2011. **6**(11): p. e26370.
7. Ewald, P.W., *Transmission modes and evolution of the parasitism-mutualism continuum.* Ann. N.Y. Acad. Sci., 1987. **503**: p. 295-306.
8. van den Bosch, F., et al., *Evolutionary bi-stability in pathogen transmission mode.* Proc. R. Soc. B., 2010. **277**(1688): p. 1735-42.
9. Rio, R.V., et al., *Symbiont succession during embryonic development of the European medicinal leech, Hirudo verbana.* Appl. Environ. Microbiol., 2009. **75**(21): p. 6890-5.
10. Worthen, P.L., C.J. Gode, and J. Graf, *Culture-independent characterization of the digestive-tract microbiota of the medicinal leech reveals a tripartite symbiosis.* Appl. Environ. Microbiol., 2006. **72**(7): p. 4775-81.

11. Laufer, A.S., M.E. Siddall, and J. Graf, *Characterization of the digestive-tract microbiota of Hirudo orientalis, a european medicinal leech.* Appl. Environ. Microbiol., 2008. **74**(19): p. 6151-4.

12. Graf, J., Y. Kikuchi, and R.V. Rio, *Leeches and their microbiota: naturally simple symbiosis models.* Trends Microbiol., 2006. **14**(8): p. 365-371.

13. 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): p. 1-7.

14. Janda, J.M. and S.L. Abbott, *Evolving concepts regarding the genus Aeromonas: An expanding panorama of species, disease presentations, and unanswered questions.* Clin. Infect. Dis., 1998. **27**(2): p. 332-344.

15. Bates, J.M., et al., *Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation.* Dev. Biol., 2006. **297**(2): p. 374-386.

16. Hu, M., et al., *Identity and virulence properties of Aeromonas isolates from diseased fish, healthy controls and water environment in China.* Lett. Appl. Microbiol., 2012. **55**(3): p. 224-233.

17. Kaneuchi, C. and T. Mitsuoka, *Bacteroides microfusus, a new species from the intestine of calves, chickens and Japanese quails.* Int. J. Syst. Bacteriol., 1978. **28**: p. 478-481.

18. Ley, R.E., et al., *Obesity alters gut microbial ecology.* Proc. Natl. Acad. Sci. U.S.A., 2005. **102**(31): p. 11070-5.

19. Eckburg, P.B., et al., *Diversity of the human intestinal microbial flora.* Science, 2005. **308**(5728): p. 1635-8.

20. Rautio, M., et al., *Bacteriology of histopathologically defined appendicitis in children.* Pediatr. Infect. Dis. J., 2000. **19**(11): p. 1078-83.

21. Siddall, M.E., et al., *Novel role for Aeromonas jandaei as a digestive tract symbiont of the North American medicinal leech.* Appl. Environ. Microbiol., 2007. **73**(2): p. 655-658.

22. Siddall, M.E., et al., *Bacterial symbiont and salivary peptide evolution in the context of leech phylogeny.* Parasitology, 2011. **138**(13): p. 1815-27.

23. Wernegreen, J.J., *Genome evolution in the bacterial endosymbionts of insects.* Nat. Rev. Genet., 2002. **3**: p. 850-861.

24. Silver, A.C., et al., *Complex evolutionary history of the Aeromonas veronii group revealed by host interaction and DNA sequence data.* PLoS One, 2011. **6**(2): p. e16751.

25. Bomar, L., et al., *Draft Genome Sequence of Aeromonas veronii Hm21, a Symbiotic Isolate from the Medicinal Leech Digestive Tract.* Genome Announc., 2013. **1**(5).

26. Sawyer, R.T., *Leech biology and behavior*. 1986, Oxford, United Kingdom: Clarendon Press.

27. Michalsen, A., M. Roth, and G. Dobos, *Medicinal leech therapy*. 2007, New York, NY: Thieme Medical Publishers.

28. Lema, K.A., B.L. Willis, and D.G. Bourne, *Corals form characteristic associations with symbiotic nitrogen-fixing bacteria.* Appl. Environ. Microbiol., 2012. **78**(9): p. 3136-44.

29. Nyholm, S.V., et al., *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., 2002. **68**(10): p. 5113-22.

30. Rohwer, F., et al., *Diversity and distribution of coral-associated bacteria.* Mar. Ecol. Prog. Ser., 2002. **243**: p. 1-10.

31. Utevsky, S., et al., *Distribution and status of medicinal leeches (genus Hirudo) in the Western Palaearctic: anthropogenic, ecological, or historical effects?* Aquat. Conserv., 2010. **20**(2): p. 198-210.

32. Holmes, P., L.M. Niccolls, and D.P. Sartory, *The ecology of mesophilic Aeromonas in the aquatic environment.*, in *The Genus Aeromonas*, B. Austin, et al., Editors. 1996, John Wiley & Sons Ltd.: West Sussex, England. p. 127-50.

33. Belkaid, Y., et al., *Development of a natural model of cutaneous leishmaniasis: powerful effects of vector saliva and saliva preexposure on the long-term outcome of Leishmania major infection in the mouse ear dermis.* J. Exp. Med., 1998. **188**(10): p. 1941-53.

34. Caljon, G., et al., *Tsetse fly saliva accelerates the onset of Trypanosoma brucei infection in a mouse model associated with a reduced host inflammatory response.* Infect. Immun., 2006. **74**(11): p. 6324-30.

35. Marchal, C., et al., *Antialarmin effect of tick saliva during the transmission of Lyme disease.* Infect. Immun., 2011. **79**(2): p. 774-85.

36. Norsworthy, N.B., et al., *Sand fly saliva enhances Leishmania amazonensis infection by modulating interleukin-10 production.* Infect. Immun., 2004. **72**(3): p. 1240-7.

37. Samuelson, J., et al., *A mouse model of Leishmania braziliensis braziliensis infection produced by coinjection with sand fly saliva.* J. Exp. Med., 1991. **173**(1): p. 49-54.

38. Spence, P.J., et al., *Vector transmission regulates immune control of Plasmodium virulence.* Nature, 2013. **498**(7453): p. 228-31.

39. Theodos, C.M., J.M. Ribeiro, and R.G. Titus, *Analysis of enhancing effect of sand fly saliva on Leishmania infection in mice.* Infect. Immun., 1991. **59**(5): p. 1592-8.

40. Titus, R.G. and J.M. Ribeiro, *Salivary gland lysates from the sand fly Lutzomyia longipalpis enhance Leishmania infectivity.* Science, 1988. **239**(4845): p. 1306-1308.

41. Cryan, J.F. and S.M. O'Mahony, *The microbiome-gut-brain axis: from bowel to behavior.* Neurogastroenterol. Motil., 2011. **23**(3): p. 187-92.

42. Desbonnet, L., et al., *The probiotic Bifidobacteria infantis: An assessment of potential antidepressant properties in the rat.* J. Psychiatr. Res., 2008. **43**(2): p. 164-74.

43. Heijtz, R.D., et al., *Normal gut microbiota modulates brain development and behavior.* Proc. Natl. Acad. Sci. U.S.A., 2011. **108**(7): p. 3047-3052.

44. Bilbo, S.D., et al., *Neonatal infection induces memory impairments following an immune challenge in adulthood.* Behav. Neurosci., 2005. **119**(1): p. 293-301.

45. Crawley, J.N., *Behavioral phenotyping strategies for mutant mice.* Neuron, 2008. **57**(6): p. 809-18.

46. Goehler, L.E., et al., *Campylobacter jejuni infection increases anxiety-like behavior in the holeboard: possible anatomical substrates for viscerosensory modulation of exploratory behavior.* Brain Behav. Immun., 2008. **22**(3): p. 354-66.

47. Sullivan, R., et al., *The International Society for Developmental Psychobiology annual meeting symposium: Impact of early life experiences on brain and behavioral development.* Dev. Psychobiol., 2006. **48**(7): p. 583-602.

48. van den Broek, I.V.F. and C.J. den Otter, *Olfactory sensitivities of mosquitoes with different host preferences (Anopheles gambiae s.s., An. arabiensis, An. quadriannulatus, An. m. atroparvus) to synthetic host odours.* J. Insect Physiol., 1999. **45**(11): p. 1001-1010.

49. Wekesa, J.W., R.S. Copeland, and R.W. Mwangi, *Effect of Plasmodium falciparum on blood feeding behavior of naturally infected Anopheles mosquitoes in western Kenya.* Am. J. Trop. Med. Hyg., 1992. **47**(4): p. 484-8.

50. White, G.B., *Anopheles gambiae complex and disease transmission in Africa.* Trans. R. Soc. Trop. Med. Hyg., 1974. **68**(4): p. 278-301.

51. Fraune, S., et al., *In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides.* Proc. Natl. Acad. Sci. U.S.A., 2010. **107**(42): p. 18067-72.

52. Dunn, A.M. and J.E. Smith, *Microsporidian life cycles and diversity: the relationship between virulence and transmission.* Microbes Infect., 2001. **3**(5): p. 381-8.

53. Huigens, M.E., et al., *Infectious parthenogenesis.* Nature, 2000. **405**(6783): p. 178-9.

54. Moran, N.A. and H.E. Dunbar, *Sexual acquisition of beneficial symbionts in aphids.* Proc. Natl. Acad. Sci. U.S.A., 2006. **103**(34): p. 12803-6.

55. Snyder, A.K., et al., *The phylogeny of Sodalis-like symbionts as reconstructed using surface-encoding loci.* FEMS Microbiol. Lett., 2011. **317**(2): p. 143-51.

56. Werren, J.H. and J. van den Assem, *Experimental analysis of a paternally inherited extrachromosomal factor.* Genetics, 1986. **114**(1): p. 217-33.

57. Mira, A. and N.A. Moran, *Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria.* Microb. Ecol., 2002. **44**(2): p. 137-43.

58. Rio, R.V., et al., *Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora.* Proc. Biol. Sci., 2006. **273**(1588): p. 805-14.

59. Holmes, D.S. and J. Bonner, *Preparation, molecular weight, base composition, and secondary structure of giant nuclear ribonucleic acid.* Biochemistry, 1973. **12**(12): p. 2330-8.

60. Branson, K., et al., *High-throughput ethomics in large groups of Drosophila.* Nat. Methods, 2009. **6**(6): p. 451-7.

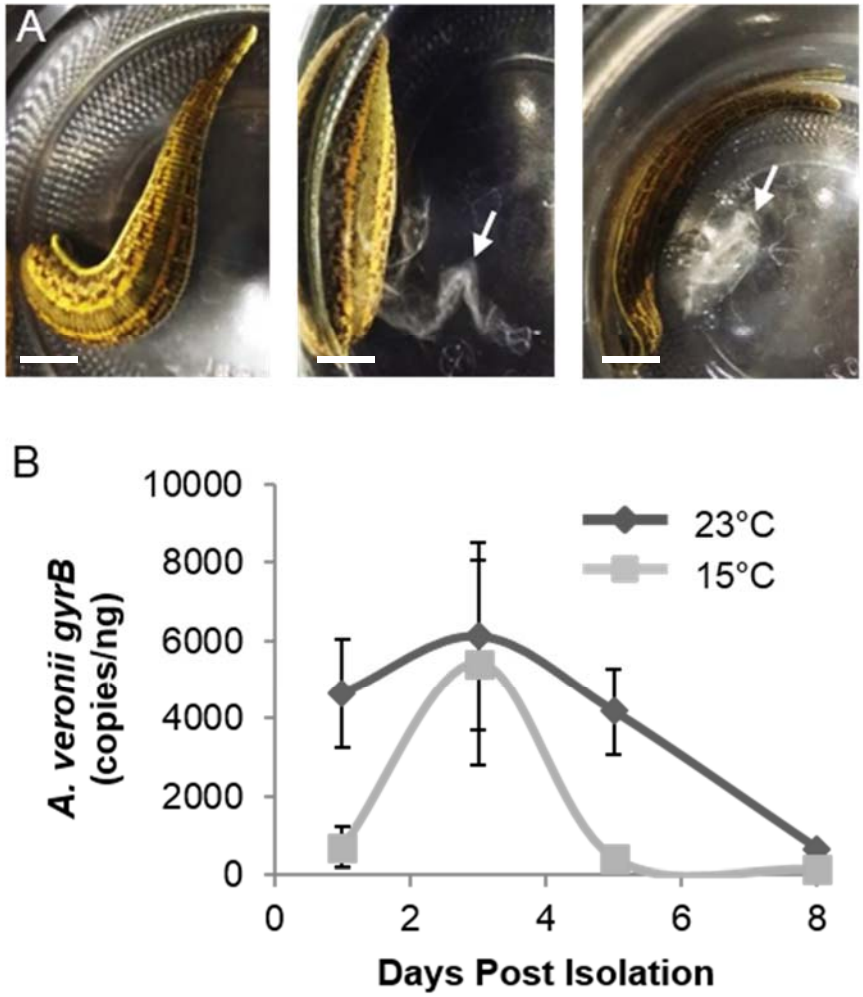

Figure 2-1. Shed mucosal secretions harbor proliferating microbial symbionts. A. A leech undergoing the process of mucosal shedding; arrow indicates mucus. Mucus is sloughed off from the anterior to posterior end of the leech. Scale bar indicates 0.96 cm. B. *A. veronii* density within mucus increases from Day 1 to 3, decreasing thereafter. *A. veronii* density is measured via qPCR using the single-copy *gyrB* gene. Significant differences in overall density between the two temperatures examined were observed (p=0.0086, one-way ANOVA; n=77).

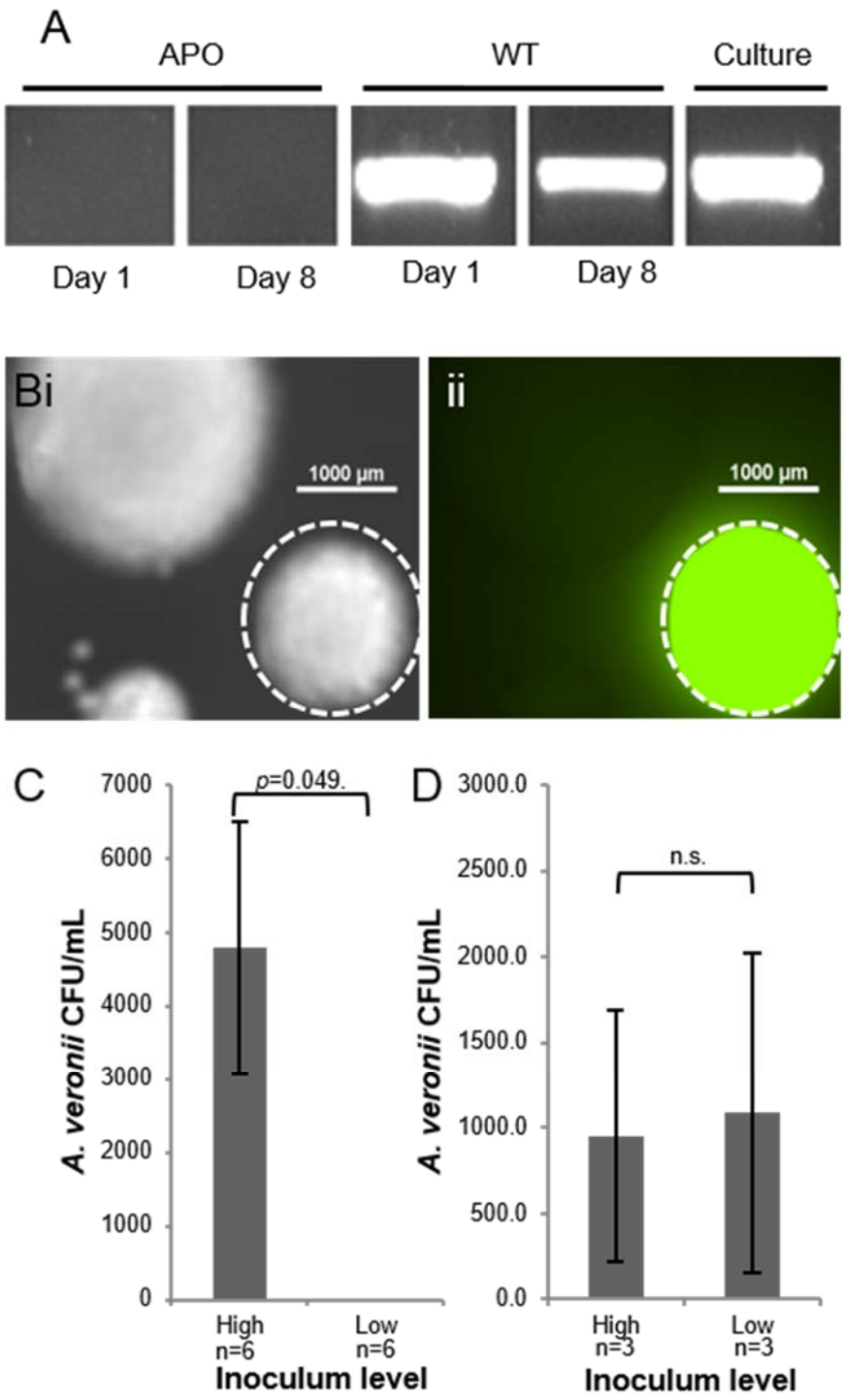

Figure 2-2. Digestive tract A. veronii seed mucosal castings, which are sufficient for symbiont transmission. A. Antibiotic treatment, sufficient to clear gut symbionts (aposymbiotic; APO), also eliminates *A. veronii* from representative mucosal castings as verified through *A. veronii* specific *gyrB* PCR. B. Leeches fed *A. veronii* possess this symbiont with their first subsequent mucosal shed. i) Brightfield imaging of bacterial colonies from the mucosal secretions of an APO leech fed *gfp*-expressing *A. veronii* mutant HM21S::TN7gfp (2 X 10^4 CFU/mL). ii) FITC fluorescent image of i; hatched circles indicate *gfp*-expressing *A. veronii* colony. C. Fluorescent colony counts within shed mucus of leeches that have been fed *gfp*-expressing *A. veronii*. D. Fluorescent colony counts within shed mucus of leeches exposed to mucus harboring *gfp*-expressing *A. veronii*. For C and D, high and low inoculums were matched to symbiont densities observed in mucus (see Fig. 1B); n.s.= not significant.

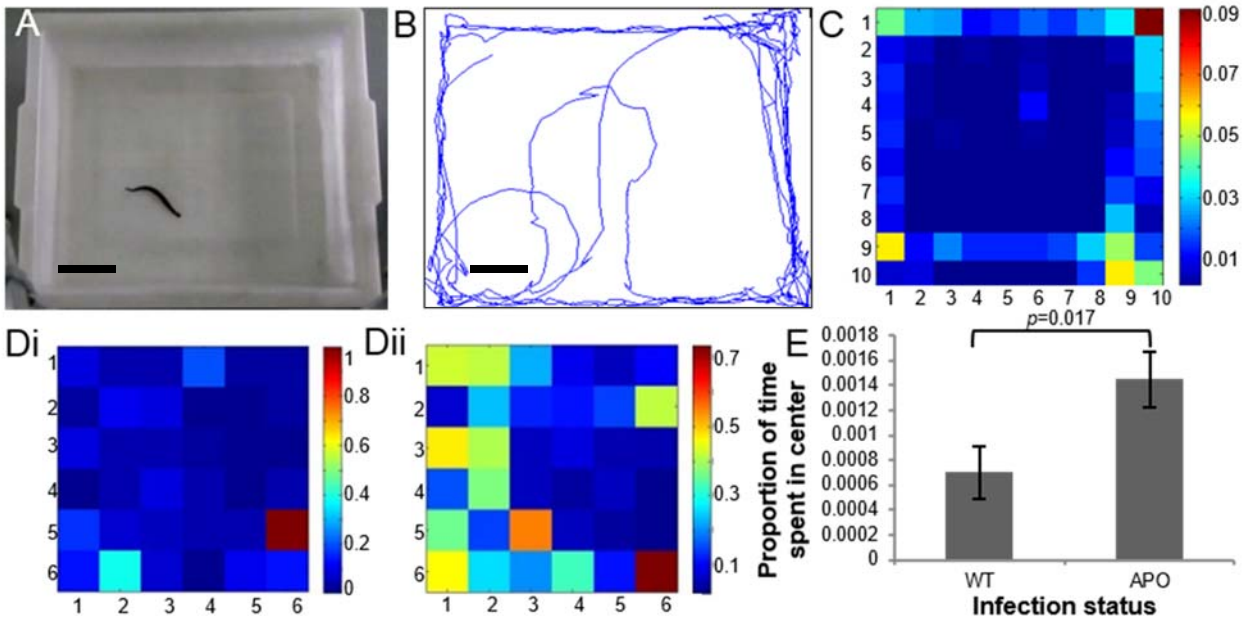

Figure 2-3. Host symbiont status impacts exploratory behavior. A. A leech within the general behavior analysis arena. Scale bar depicts 3.33 cm. B. A representative leech movement track obtained during a single recording session. Scale bar depicts 3.33 cm. C. An example of the 10 x 10 grid used to observe time spent in each section of the arena, with warmer colors indicating a greater proportion of time spent within that square. Each 1 X 1 square represents 4.44 cm. D. The inner 6 x 6 squares of the arena depict time spent in the center between i) WT and ii) APO leeches. Each 1 X 1 square represents 4.44 cm. E. Baseline behavior assays demonstrate that APO leeches spend significantly more time in the center of the arena than WT (p=0.017, Student's t-test, n=33), consistent with more exploratory behavior. Error bars indicate 1 Standard Error of the Mean (SEM).

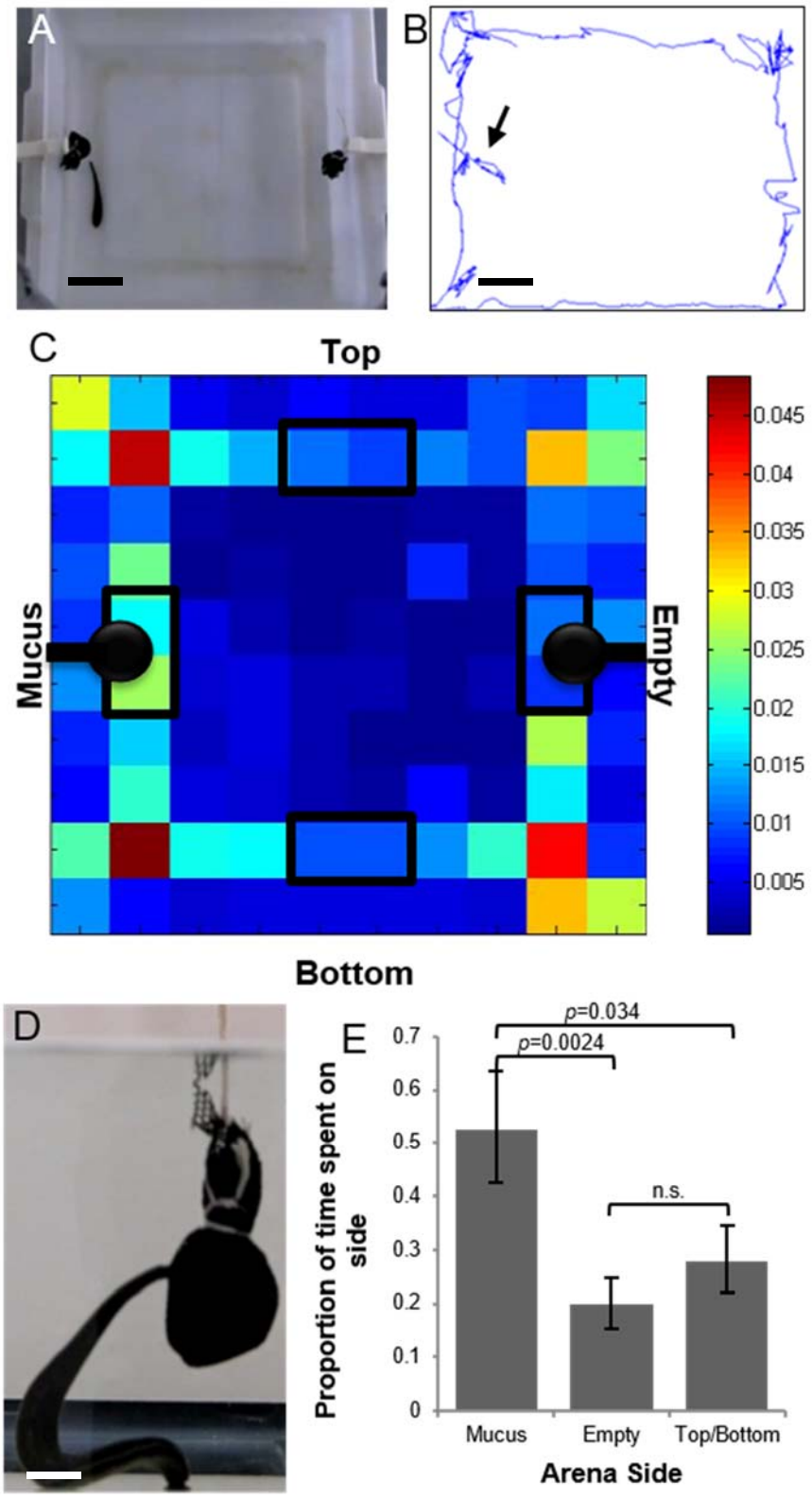

Figure 2-4. Leeches are attracted to mucus shed by conspecifics. A. A leech within the behavioral analysis arena; mesh bags that are either empty or contain mucus are visible on either side of the arena. Scale bar depicts 3.33 cm. B. A representative leech movement track obtained during a single recording session; arrow indicates location of mesh sac containing mucus. Scale bar depicts 3.33 cm. C. Average time distribution of leech movement in the behavior arena, warmer colors indicate a greater proportion of time spent in that square. Black boxes indicate regions of interest measured for each side of the arena representing the top, bottom, and each side in which the mesh sacs (black circles) are placed. Each 1 X 1 sqaure represents 4.44 cm. D. A leech investigating a mucus filled sac. Scale bar depicts 1.5 cm. E. Leeches spend significantly more time at the mesh sac containing mucus (p= 0.0024, one-way ANOVA, n=37).

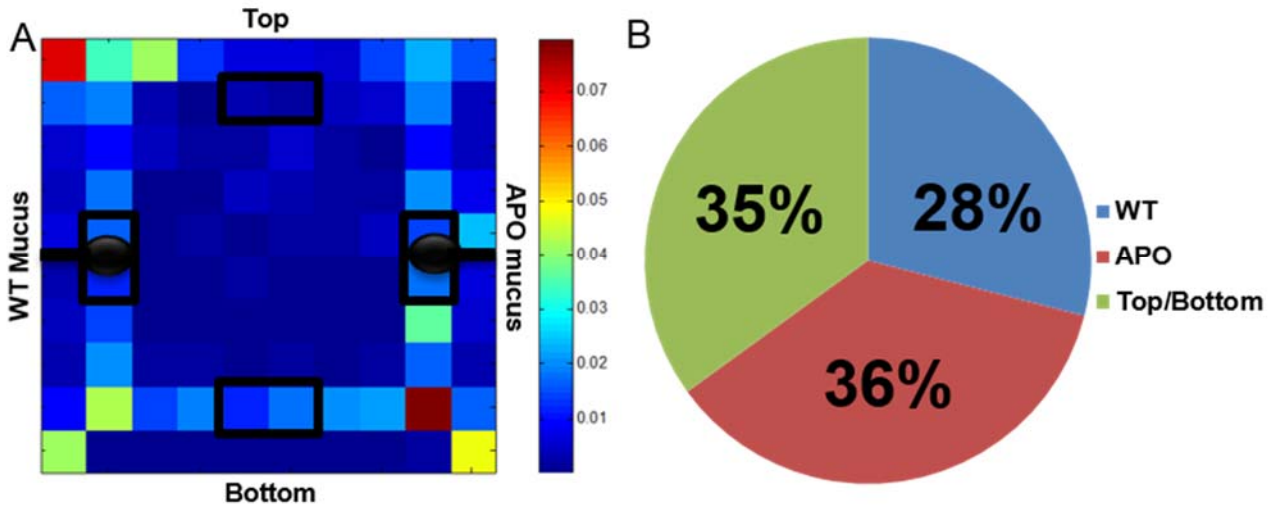

Figure 2-5. Mucus infection does not affect leech attraction. A. Average time distribution of WT leech movement in the behavior arena in the presence of mucus. Black boxes indicate regions of interest measured for each side of the arena: top, bottom, and each side in which mesh sacs (black circles) contain either WT or APO mucus. Each 1 X 1 square represents 4.44 cm. B. There was no significant difference in the time leeches spent near the mesh sac containing either WT or APO mucus (p=0.4585, ANOVA, n=18).

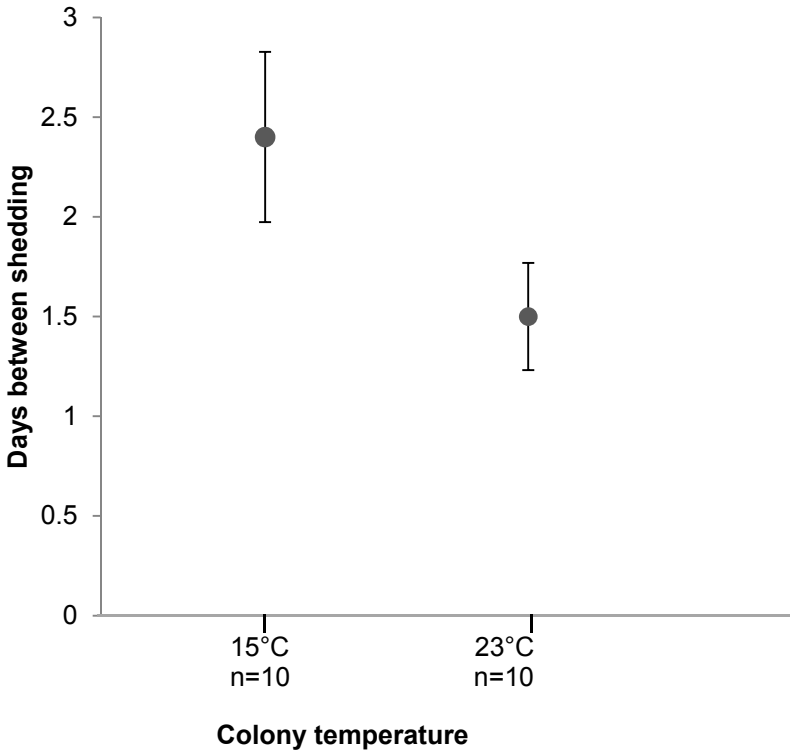

Figure 2-S1. Mucosal secretions are shed at regular intervals. Host mucosal shedding occurs approximately every 2 days, providing a continuous source of symbionts to the environment, *n* = sample size.

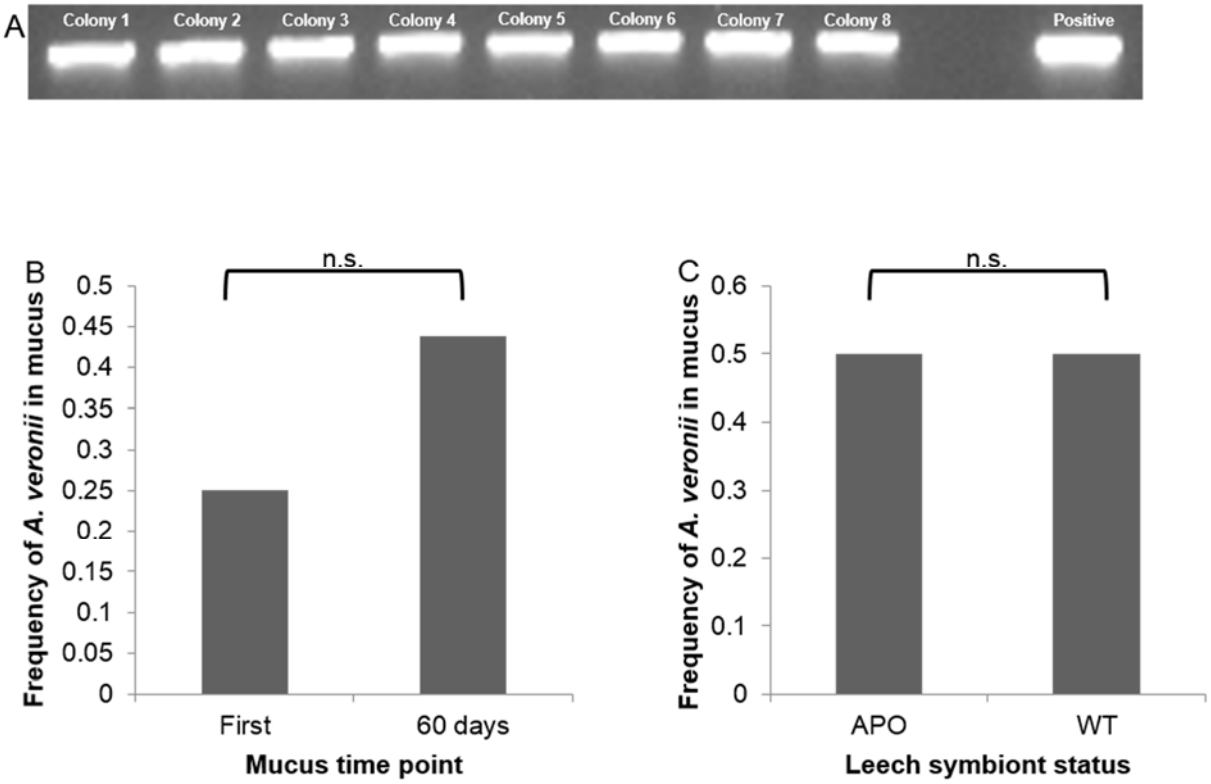

Figure 2-S2. Digestive tract A. veronii seed mucosal castings and are sufficient for symbiont transmission. A. Identity of *gfp*-expressing *A. veronii* isolated from the mucosal secretion of an APO leech fed a high inoculum (2 x 10^4 CFU/mL) was confirmed using *A. veronii*-specific *gyrB* primers. B. There is no significant difference in the frequency of fluorescent *A. veronii* colonies found in the first mucus and a mucus shed 60 days following *per os* feeding (p=0.3865, Pearson's chi-squared test, n= 24). C. There was also no significant difference in the frequency of fluorescent colonies found in the mucus of APO or WT leeches following *per os* administration of *A. veronii* (p=0.7595, Pearson's chi-squared test, n=24). Only the results of high oral administration of *A. veronii* are included in B and C due to the absence of *gfp*-expressing *A. veronii* in the mucosal secretions of leeches receiving low inoculums.

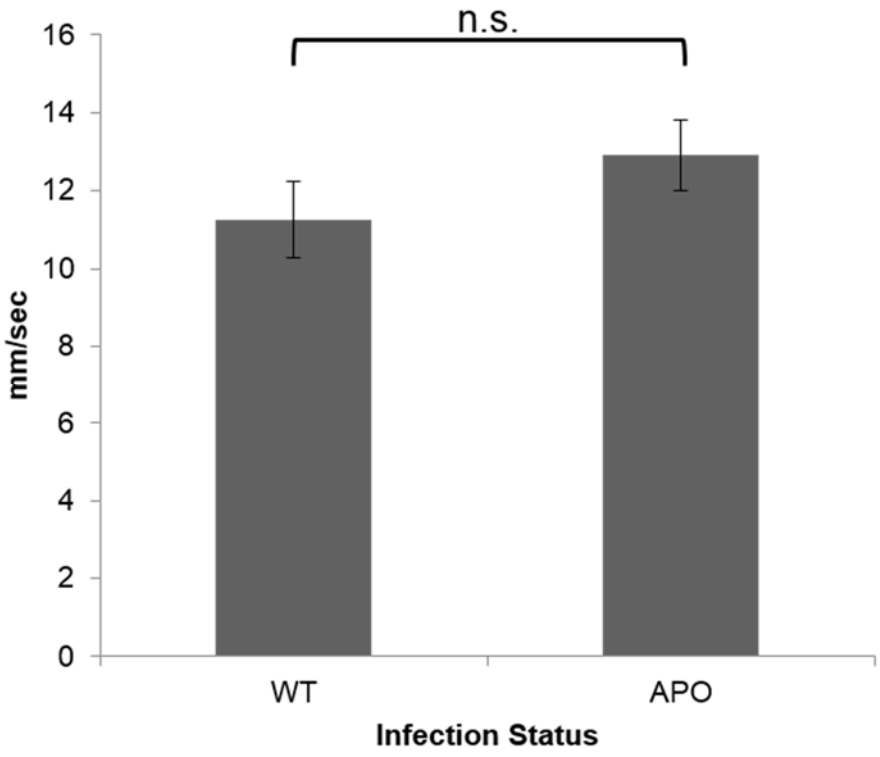

Figure 2-S3. Antibiotic treatment does not affect the speed with which leeches move within the behavioral arena. There was no significant difference in distance traveled (mm)/sec when comparing APO and WT leeches (p=0.36; Student's t-test, n=33); n.s.= not significant.

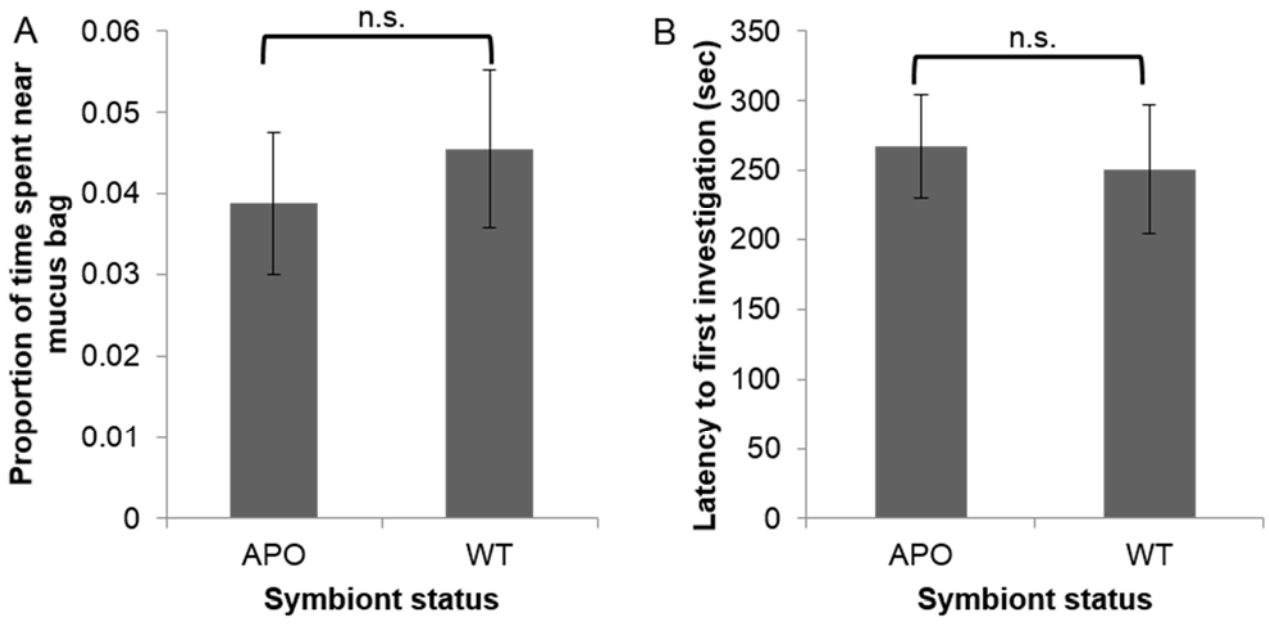

Figure 2-S4. Leeches display attraction to mucus irrespective of leech symbiont status. A. There was no significant difference between APO and WT leeches in the proportion of time spent near the mucus bag (p=0.609, Student's t-test, n=36). B. There also was no significant difference between APO and WT leeches in the latency to first investigate the mucus sac (p=0.78, Student's t-test, n=23).

Table 2-1. Primer list and amplification settings

Primer	Sequence	Amplicon size (bp)	Amplification conditions[a]
Aeromonas veronii gyrase B (*gyrB*)		514	35 cycles of 30 sec at 95°C, 30 sec at 53°C, and 30 sec at 72°C
AvgyrB-F	5'-GCA GAT TGG CGA CAG CAC-3'		
AvgyrB-R	5'-GCA CCT TGA CGG AGA TAA CG-3'		
Aeromonas veronii gyrase B (*gyrB*) q-PCR		75	35 cycles of 5 sec at 98°C, and 5 sec at 53°C
QT-AvgyrB For	5'-ACT CAC CTC GTT GGC TTC C-3'		
QT-AvgyrB Rev	5'-CTT GCT GTA GTC CTC TTT GTC C-3'		

[a]Each amplification reaction was initiated with a 3 min denaturation step at 95°C, excluding the *Aeromonas veronii* gyrase B (*gyrB*) q-PCR, which contained an initial 2 min denaturation step.

Chapter 3: Characterization of shed medicinal leech mucus reveals a diverse microbiota[2]

ABSTRACT

Microbial transmission through mucosal-mediated mechanisms is widespread throughout the animal kingdom. One example of this occurs with *Hirudo verbana*, the medicinal leech, where host attraction to shed conspecific mucus facilitates horizontal transmission of a predominant gut symbiont, the Gammaproteobacterium *Aeromonas veronii*. However, whether this mucus may harbor other bacteria has not been examined. Here, we characterize the microbiota of shed leech mucus through Illumina deep sequencing of the V3-V4 hypervariable region of the 16S rRNA gene. Additionally, Restriction Fragment Length Polymorphism (RFLP) typing with subsequent Sanger Sequencing of a 16S rRNA gene clone library provided qualitative confirmation of the microbial composition. Phylogenetic analyses of full-length 16S rRNA sequences were performed to examine microbial taxonomic distribution. Analyses using both technologies indicate the dominance of the Bacteroidetes and Proteobacteria phyla within the mucus microbiota. We determined the presence of other previously described leech symbionts, in addition to a number of putative novel leech-associated bacteria. A second predominant gut symbiont, the *Rikenella*-like bacteria, was also identified within mucus and exhibited similar population dynamics to *A. veronii*, suggesting persistence in syntrophy beyond the gut. Interestingly, the most abundant bacterial genus belonged to *Pedobacter*, which includes members capable of producing heparinase, an enzyme that degrades the anticoagulant, heparin. Additionally, bacteria associated with denitrification and sulfate cycling were observed, indicating an abundance of these anions within mucus, likely originating from the leech

[2] Modified and reprinted from Ott BM, Rickards A, Gehrke, Rio RVM. Characterization of shed medicinal leech mucus reveals a diverse microbiota. Front. Microbiol., 2015. 5:757. doi: 10.3389/fmicb.2014.00757

excretory system. A diverse microbiota harbored within shed mucus has significant potential implications for the evolution of microbiomes, including opportunities for gene transfer and utility in host capture of a diverse group of symbionts.

INTRODUCTION

Host-generated mucus may harbor both pathogenic and beneficial microbes [1-5] with numerous examples of mucus-mediated microbial transmission occurring throughout the Animal kingdom. For instance, the bobtail squid (*Euprymna scolopes*) uses mucosal secretions for the aggregation of its bioluminescent symbiont, *Vibrio fischeri*, from the surrounding water facilitating its migration into the light organ [6]. Representing basal metazoans, the hydra (*Hydra vulgaris*) contains an external mucosal layer, termed the glycocalyx, where critical symbionts are recruited during early host embryogenesis [7]. Additionally, the mucosal secretions of humans (i.e. sputum and nasal secretions) provide a protective environment, particularly in terms of humidity and salinity [8], enabling the transmission of infectious respiratory agents between individuals.

The sanguivorous European medicinal leech, *Hirudo verbana* (Hirudinida: Hirudinidae) uses a dual mode of transmission for acquiring a predominant gut symbiont, the Gammaproteobacterium *Aeromonas veronii* [9]. Vertical transmission of *A. veronii* occurs during cocoon development likely by the albumenotrophic activity of larvae [10], while colonization of adults is through contact with shed mucus that contains a proliferating *A. veronii* population that originates from the digestive tract [9]. Importantly, leeches are attracted to mucus produced by conspecifics [9], which facilitates symbiont horizontal spread. This mixed mode of transmission has significant implications for the evolution of symbiosis, including enabling the capture of a more genetically diverse symbiont population with an enhanced ability to adapt to

environmental changes. Furthermore, lifestyle options beyond that of mutualism with the leech may be possible for the mucus-inhabiting bacteria.

The shedding of mucus by medicinal leeches occurs every 2-3 days [9] always in an anterior to posterior direction. In addition to catalyzing *A. veronii* horizontal transmission, these mucosal casts have been proposed to serve a multitude of other roles ranging from protection against UV rays and desiccation to facilitating conspecific recognition [11]. Within shed mucus, the *A. veronii* population originates from the leech digestive tract with density maximizing just prior to a new secretion [9]. This suggests the coupling of symbiont population dynamics and host biology, as well as the potential for other microbial inhabitants within shed mucus to provide metabolic and/or structural support enabling *A. veronii* proliferation.

While leech shed mucus was demonstrated to aid in the transmission of a sole gut symbiont, a moderately rich microbiota (~36 taxa) is actually housed within the *H. verbana* GI tract [12]. The leech GI tract is primarily composed of two parts; a crop, where the blood meal is stored, and a smaller intestinum, the actual site of nutrient absorption [13]. In addition to *A. veronii*, a second predominant Bacteroidetes symbiont, a *Rikenella*-like bacterium, is also localized to the leech gut. Within the host GI tract, the *A. veronii* and the *Rikenella*-like symbionts are synergistic [14] based on glycan utilization [15], and the likelihood that *A. veronii* may reduce the oxygen supply, enabling the habitation of the anaerobic *Rikenella*-like symbiont.

In addition to the gut, a second endogenous microbiota has also been described in the leech excretory system, consisting of multiple pairs of nephridia connected to bladders that lie alongside the lateral caeca of the crop. A maximum of six bacterial species reside within the bladder, with infection rates of these taxa varying between different leech individuals [16]. Interestingly, symbiont species display a stratified spatial arrangement within the bladder

suggesting a community organization likely impacted by resource availability and output metabolism. The functional basis of the excretory tract symbionts may lie in the recycling of carbon and nitrogenous waste [16], which may enable the leech host to sustain long gaps, often as long as 6-12 months [17], between blood meals.

In this paper, we characterize the composition and relative abundance of the microbiota within shed leech mucus using Illumina deep sequencing of the V3-V4 hypervariable region of the 16S rRNA gene. Additionally, Restriction Fragment Length Polymorphism (RFLP) typing and Sanger sequencing of a 16S rRNA clone library provided qualitative confirmation. Phylogenetic analyses of full-length 16S rRNA sequences were performed to examine both the microbial diversity and their evolutionary relations within leech shed mucus. Lastly, following the identification of the *Rikenella*-like symbiont within mucus, its population dynamics were compared with the *A. veronii* symbiont [9]. The discovery of a rich microbial community within mucus suggests its utility for genetic mixing and resource partitioning within this setting. This species assemblage raises questions pertaining to microbial dynamics and whether these other bacteria, as a group, may also utilize mucus as a means for horizontal transmission.

MATERIALS AND METHODS

(a) Leech husbandry. Medicinal leeches (*H. verbana*), were obtained from the medical supplier Leeches USA (Westbury, NY, USA), and housed in sterile Leech Strength Instant Ocean H_2O (0.004%, I.O.) at 15°C at constant darkness. Leeches were maintained on defibrinated bovine blood (Hemostat, CA).

(b) Mucus sampling. Day 3 mucus (i.e. 3 days post shedding; isolated at Day 1 and incubated separately in sterile leech strength I.O.) was chosen for both the construction of a mucosal 16S rRNA clone library and Illumina 16S rRNA deep sequencing, as this time point corresponds to

the peak of *Aeromonas* population size [9]. Mucosal samples used for describing the *Rikenella*-like bacterium population dynamics through time were obtained at 1, 3, 5 or 8 d post shedding within sterile Leech Strength I.O. H_2O. All samples were snap frozen at -80°C until further processing.

(c) High-throughput amplicon sequencing analyses. The microbial community of shed mucus was characterized using barcoded Illumina sequencing. Total DNA was extracted from shed mucus using the Holmes-Bonner Protocol [18] and tested for purity on a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA). This DNA served as template for PCR amplification of the V3-V4 hypervariable region of the 16S rRNA gene [19] using the V3Met (5'-CCTACGGGAGGCAGCAG-3') and MetaV4 (5-GGACTACHVGGGTWTCTAAT-3') primers. Amplicons (1x250 bp, paired end) were sequenced on the Illumina MiSeq platform in the West Virginia University Genomics Core Facility following the manufacturer's protocols (Illumina, CA). Sequence quality control was performed using mothur [20] following the MiSeq SOP (http://www.mothur.org/wiki/MiSeq_SOP; date accessed page August 30, 2014). The screen.seqs command was used to trim the sequence when the average quality score over a 50 bp window dropped below 35, and to eliminate any sequences that were not in the 400-500 bp range. The unique.seqs command was used to cluster the sequences that were within 2 bp of similarity to a more abundant sequence. Sequences were aligned to the SILVA-compatible alignment database using align.seqs and then trimmed to a common region (i.e. 6388 to 25316 of *Escherichia coli*) using the filter.seqs command to remove any overhangs. The classify.seqs and remove.lineage commands were utilized to identify and remove mitochondrial, chloroplast, Archaea, Eukarya and unknown contaminants. Bacterial taxonomy was assigned to each sequence in the improved data set using the classify.seqs command, which uses the Naïve

Bayesian classifier of RDP [21]. Following taxonomic assignment, sequences were assigned to operational taxonomic units (hereafter OTUs) at the 3% level of divergence using the cluster.seqs command. Here, a 16S rRNA sequence is considered derived from a known genus if the read similarity was $\geq$ 95% [22]. The read counts at each taxonomy level were normalized to total relative abundance.

(d) Mucosal 16S rRNA clone library. To assess the microbial diversity within leech shed mucus, total DNA was extracted from 3 d old mucosal sheds of two individuals, using the Holmes-Bonner protocol [18], and subjected to PCR using 27F' and 1492R' general eubacteria primers [23, 24] (T_a [annealing temperature] = 50°C; 28 cycles; amplicon ~ 1450 bp). PCR products were cloned using the pGEM-T Easy Vector cloning kit (Promega, WI), with subsequent transformation into JM109 *Escherichia coli* cells (Promega, WI). Inserts were amplified using M13F' and M13R' vector primers (T_a = 46°C; 35 cycles; amplicon ~ 1636 bp) and digested with *HaeIII* restriction endonuclease (NEB, Ipswich, MA, USA) for RFLP typing. Clones with unique restriction profiles were purified and subject to Sanger sequencing using M13 primers with an ABI Genetic Analyzer 3130*xl* at the WVU Department of Biology Genomics Center. The DNA sequences were aligned and assembled into contigs and identified to the highest taxonomic level possible using nucleotide Basic Local Alignment Search Tool (BLASTn, http://blast.ncbi.nlm.nih.gov/Blast.cgi).

(e) Molecular phylogenetic analyses. To examine microbial community diversity and their relations, phylogenetic trees including the 16S rRNA sequences that were identified within mucus samples, close relatives, and previously identified *H. verbana* leech isolates [16, 25] were constructed. DNA sequences were aligned using the Clustal X algorithm with default settings, and refined manually when necessary. Maximum parsimony (MP) analyses were performed with

1000 replicates in PAUP 4.0 [26]. MP heuristic searches utilized the tree-bisection-reconnection (TBR) branch-swapping algorithm with 200 Max trees and starting trees were created using stepwise additions. All MP analyses were performed twice, where gaps were treated either as "missing data" or as a "5th character state", with no differences noted between the results. Lineage support was measured by calculating nonparametric bootstrap (BS) values (n =1000) [27].

A second tree was produced for the same alignment using the Bayesian Markov Chain Monte Carlo method as implemented with MrBayes (3.1.2) [28]. The best-fit model (GTR+I+G) used for Bayesian analyses was statistically selected using the Akaike Information Criterion in MrModeltest version 2.3 [29]. Bayesian analyses were performed with six Markov chains [30] for 5,000,000 generations. Posterior probabilities (PP) were calculated, with the stabilization of the model parameters (i.e. burn-in) occurring around 4,000,000 generations. Every 100th tree following stabilization was sampled to determine a 50% majority rule consensus tree. All trees were made using the program FIGTREE v.1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

(f)* Rikenella-*like bacteria population dynamics. The population dynamics of *Rikenella*-like symbionts within host mucus at 1, 3, 5 and 8 d following secretion were determined by utilizing real-time quantitative PCR (q-PCR). DNA isolation from mucus was performed following a Holmes-Bonner protocol [18]. Analyses were performed in an iCycler iQ Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA) using Bio-Rad SsoFast EvaGreen Supermix, 10 mM of primers *rpoD*F' q-PCR (5'-AGT TGC GGA CAC TCT ACG TG-3') and *rpoD*R' q-PCR (5'-TCC AAG AGC GTG TTG TCT TC-3') (T_a = 55°C; 35 cycles; amplicon = 84 bp) and 2 µL of DNA template as described [10]. Quantification of the amplicons relative to standard curves was performed using Bio-Rad CFX Manager software v 2.0. The respective DNA

concentration (ng/ul) of each sample was used for the normalization of copy numbers. All assays were performed in triplicate and replicates were averaged for each sample.

(g) Statistical analyses. Mean microbial species richness values were obtained using the Shannon-Weaver diversity index [31]. A rarefaction curve was generated using Hurlbert's formulation [32] and vegan version 2.0-10 (Okansen) using the "rarefy" function. The R package "Fossil" was used to generate Chao1, Chao2, ICE and ACE richness estimates. Percent identities of the16S rRNA sequences between different *Pedobacter* isolates and species were generated using MEGA 6.06 (Tamura et al, 2013; Mol. Bio. Evol) implementing the Jukes-Cantor Model for nucleotide substitution (Jukes and Cantor, 1969). Statistically significant differences in *Rikenella*-like symbiont density within mucus through time were determined by performing a one-way ANOVA using JMP 10 (SAS Institute Inc., Cary, NC, USA) software, with a significance value set at $p \leq 0.05$.

(h) Confirmation of reagent purity. The DNA extraction buffer, ultrapure water and PCR kit used for nucleic acid extraction and amplification respectively, were verified to be contaminant-free using PCR amplification with general eubacterial primers.

(i) Nucleotide sequence accession numbers. The complete sequence dataset for the Illumina reads is available from NCBI under BioProject ID PRJNA269158. Individual sequences generated by the Sanger-sequenced 16S rRNA clone library are available at Genbank under accession numbers KP231731-KP231772.

RESULTS

Shed mucus harbors a diverse microbial community.

To obtain an understanding of the microbiome profile within shed mucus (Figure 1A), Illumina deep-sequencing of the V3-V4 hypervariable region of the 16S rRNA gene was

performed. These mucosal secretions, consisting primarily of glucosaminoglycans, are produced by mucus gland cells that are irregularly distributed beneath the external leech epithelia [11, 13] (Figure 1B). Mucus 3 d post shedding was selected for the characterization of the microbiota as this time point corresponds to the peak in density of a predominant leech gut symbiont, *A. veronii* [9], and we were particularly interested to know if other bacteria inhabited mucus at this time.

Quality-filtered reads from the mucosal Illumina dataset were assigned to a reference OTU housed in the phylum Bacteroidetes (~68% of reads), Proteobacteria (~29% of reads), or were categorized as "Other" (3% of reads) (Figure 2A). Of these a total of 2,136,157 reads (~98%) could be assigned to the genus level, with an additional 34,784 sequences unclassifiable to this level. Approximately 4% of all reads (84,918 reads) were to genera that constituted <1% of the mucosal microbiota. Genera that contained ≥1% of total reads included *Bdellovibrio* (2%), *Chitinibacter* (2%), *Curvibacter* (15%), *Methylophilus* (25%), *Polynucleobacter* (7%), and *Zoogloea* (2%) within Proteobacteria (Figure 2B) and *Pedobacter* (52%) within Bacteroidetes (Figure 2C). Interestingly, the most prevalent 16S rRNA OTUs obtained within Bacteroidetes are housed in the *Pedobacter* genus (~52% of OTUs), which has never been identified within the medicinal leech. However, *Bdellovibrio sp*. (~2% of OTUs), which has been previously described within the nephridial system, was found within the Proteobacteria phylum [16]. Additionally, the predominant gut bacteria, *A. veronii* and the *Rikenella*-like bacterium, were also detected, but both consisted of < 1% of total OTU abundance in their respective phyla.

RFLP typing of the mucosal microbial community.

A total of 140 clones from a mucosal 16S rRNA library were binned into RFLP types. A total of 25 unique RFLP types were identified (Table 1), sequenced, and their classification

determined through BLASTn [33]. At least two clones from each RFLP type were sequenced in both directions with no sequence variation observed between clones. Phylogenetic reconstruction, using both Maximum Parsimony (MP) and Bayesian analyses, confirmed the taxonomic identities of the retrieved sequences while also enabling the visualization of microbial diversity within shed mucus. The trees produced by both MP and Bayesian analyses were overall congruent (Figure 3). The 16S rRNA sequences retrieved from the mucosal clone library belonged to either Bacteroidetes or Proteobacteria, with 6 classes (i.e. Bacteroidia, Sphingobacteriia, Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, and Deltaproteobacteria) represented within these phyla (Figure 3). A number of bacterial groups, including an unclassified Bacteroidetes, an unclassified Betaproteobacteria, *Pedobacter, Rikenella, Curvibacter,* and *Aeromonas veronii,* exhibited multiple RFLP types suggesting 16S rRNA nucleotide diversity within these taxa in leech-shed mucus (Table 1).

The 16S rRNA mucosal clone library identified the presence of previously described leech symbionts. Specifically, the mucosal 16S rRNA clone library contained the sequences for; *A. veronii*, *Rikenella*-like bacteria and an unclassified Proteobacteria previously described within the GI tract; *Comamonas*, an unclassified Betaproteobacteria, and *Desulfovibrio* previously identified within the bladder; and an unclassified Bacteroidetes which has been described in both the leech digestive and excretory systems. Additionally, novel (i.e. not previously associated with *H. verbana*) sequences recovered were related to other Bacteroidetes (i.e. *Pedobacter* and a Chitinophagaceae family member) and Proteobacteria, specifically members of the Neisseriaceae family, and *Polynucleobacter*, *Polaromonas*, *Curvibacter*, *Sinorhizobium* and *Methylophilus* genera. Further, the various RFLP types observed within certain taxa (i.e. the unclassified Bacteroidetes, *Pedobacter, Rikenella, Curvibacter,* unclassified Betaproteobacteria and *A.*

veronii) are supported by the phylogenetic branching patterns of the corresponding sequences (Figure 3). For example, the *Rikenell*a-like bacterium found within shed mucus formed a monophyletic group with a *Rikenella*-like symbiont isolated from the leech digestive tract, with these forming a sister clade to an additional *Rikenella*-like symbiont isolate from the GI tract supporting diversity within this genus inside the leech host.

Additionally, while the closely related *Sphingobacterium sp*. symbionts have been isolated within the leech crop [25] and bladders [16], the positioning of this crop isolate in a sister clade to the mucosal *Pedobacter* isolates with strong statistical support (95% BS and 1.0 PP values), demonstrates their distinct identities (Figures 2 & S1). Comparisons of *Pedobacter* full-length 16S rRNA sequences generated by the Sanger-sequenced clone library resulted in identities ranging from 94.3-99.9%, indicating genetic diversity in *Pedobacter* isolates within mucus. Similar comparisons with the most closely related, previously characterized *Pedobacter* species (Figure S2) revealed 16S rRNA sequence identities ranging from 93.9-98.6%, suggesting that mucosal isolates are housed within the *Pedobacter* genus, but may be distinct from known species.

Mean species richness, as determined through the comparison of Shannon-Weaver diversity index values, was higher in the mucosal clone library (2.34) relative to the bladder (1.4) and digestive tract clone libraries (1.11 and 1.38 at 7 d and 90 d following feeding; respectively) (Table 2). In further support, the higher asymptote obtained with the shed mucosal clone library rarefaction curve (Figure 4A) also supports higher species richness [34] within shed mucus relative to the leech digestive tract and bladder microbial communities. However, the characterization of the microbiota within shed mucus appears to be incomplete through RFLP typing of the 16S rRNA clone library, as demonstrated when comparing observed richness

values (Shannon-Weaver diversity indices) that account for only 10-23% of nonparametric estimators of expected richness (ACE, ICE, Chao 1 and Chao 2 estimators). This indicates the significant presence of low abundance sequences and unevenness in the abundance of taxa, as better captured with Illumina deep sequencing.

***Rikenella*-population dynamics.**

Following the detection of the *Rikenella*-like bacteria also within shed mucus, we aimed to determine whether this symbiont may also be found proliferating within this substrate. We assayed the population density of *Rikenella*-like bacteria using quantitative PCR of the *rpoD* gene from mucus samples at two biologically relevant temperatures representing a pond's edge in *H. verbana*'s natural geographic distribution [35], the site of leech mating during the summer months (23°C, [13]), and a pond's base (15°C). No significant differences in *Rikenella* load were observed between the two temperature regimens (p=0.6868, one-way ANOVA, n=39). The *Rikenella*-like symbiont population dynamics (Figure 4B) mirrored those of *A. veronii* [9], reaching maximum abundance three days post shedding at both temperatures. A trend was observed in the *Rikenella*-load when examining time points following shedding (p=0.0518, one-way ANOVA, n=39), with Day 3 showing a higher abundance than Day 8 at both examined temperatures.

DISCUSSION

Many animals harbor mutualistic microbes that provide adaptive advantages to their hosts, often enabling unique lifestyles and habitation within restricted environments (reviewed in [36]). These mutualists can be acquired through vertical or environmental transmission, although a mixture of both routes may occur [37]. With the medicinal leech, acquisition of a predominant

gut symbiont, *A. veronii*, may occur as larvae during cocoon development [10] and following eclosion through contact with conspecific mucosal sheds [9].

Previous research has examined the leech gut microbiota utilizing both Sanger-sequenced 16S rRNA clone libraries [25] and next-generation deep sequencing [12], with the latter showing higher GI tract microbial diversity. In this study, we aimed to understand the taxonomic composition and relative abundance of the microbial community harbored within leech mucosal secretions and elucidate on their potential functional roles. Both independent culture-free molecular techniques, RFLP typing and Sanger sequencing of a 16S rRNA clone library and Illumina sequencing of the 16S rRNA V3-V4 hypervariable region, validated that shed mucus harbored bacteria from the Bacteroidetes and Proteobacteria phyla. The microbiota of leech mucosal secretions also consist of the previously described GI and bladder symbionts *Comamonas sp*., *Desulfovibrio sp*., and *Bdellovibrio sp*., as well as predominant digestive tract symbionts (*A. veronii* and *Rikenella*-like bacteria), in addition to a number of bacteria that have not been previously associated with the leech. The identification of these novel bacteria in association with shed mucosal casts supports the presence of yet to be described microbiotas within leeches, such as a mouth, pharyngeal and/or epidermal community. Additionally, microbial richness (i.e. the number of different types of bacteria characterized) was higher within the mucosal 16S rRNA clone library relative to the digestive and excretory tract clone libraries [16, 25], further suggesting the presence of unexplored microbial communities within the leech. Interestingly, observed richness values (Shannon-Weaver diversity indices) account for only a fraction (i.e. 11-14%) of expected richness estimators, indicating the existence of a large proportion of low abundance sequences that were not discovered in the mucosal 16S rRNA clone

library. This hypothesis is supported by the numerous genera uncovered in low abundance through Illumina deep-sequencing of the 16S rRNA V3-V4 hypervariable region.

In contrast to the 16S rRNA clone library, Illumina deep-sequencing identified a larger percentage of Bacteroidetes OTUs in comparison to the Proteobacteria (68% versus 29%; respectively), while *Aeromonas* consisted of less than 1% of relative OTU abundance. The low number of *Aeromonas* sequences in our Illumina data set is not entirely surprising, as others have also encountered difficulties detecting *Aeromonas* during taxonomic assignment using only two variable regions (V4-V5) [38]. As Illumina is currently limited to short-read lengths, only small segments of the 16S rRNA gene can be characterized and it appears that, at least in regards to *Aeromonas*, longer sequences may be appropriate for proper taxonomic designation. Additionally, biases in OTU detection could also result from the use of different primers [39] and primer efficiency, as well as in the initial amplification steps (i.e. PCR bias [40]). The generation of the Sanger-sequenced clone library was also dependent on RFLP typing through the use of the *HaeIII* restriction enzyme. If different taxa have similar RFLP banding patterns, these would have been placed in the same group. Although we attempted to reduce this occurrence by sequencing multiple clones from each haplotype, it is possible that some unique sequences were not detected. In congruence with our results, a study conducted on the microbiome of *Cephalotes* ants, comparing diversity obtained through 454 pyrosequencing to Sanger-sequenced clone libraries, resulted in similar discrepancies [39]. This study revealed that at higher taxonomic ranks the groups were qualitatively similar, although the proportions of each varied [39].

The most abundant microbial group present within shed mucus, as determined by both sequencing methodologies, belonged to the *Pedobacter* genus, a novel leech associated bacterial group. Here, it is important to note that the abundance of certain bacteria within shed mucus may

not be reflective of density within the leech host as the mucosal substrate may enable the proliferation, or contrastingly the decrease, of certain microbial groups over others. Within mucus, *Pedobacter* sequences are diverse, as exemplified by multiple RFLP banding patterns, comparisons of sequence identity, and when mucosal sequences are placed in *Pedobacter*-focused phylogenetic trees. The genus *Pedobacter* was first described by Steyn et al. (1998) to include heparinase-producing, obligate aerobic, Gram-negative rods discovered in soils and activated sludge. However, members of this genus have since been found in a wide array of environments, to include: freshwater lake sediment [41] and natural [42] and man-made water sources [43]. The capabilities of most *Pedobacter sp*. are not well described, although a characteristic function, degradation of the anticoagulant heparin through the production of heparinase [44, 45], is of particular interest in regards to the ecology of the leech host. The role of heparinase in a putative symbiont of a blood-feeding host is unknown, but may facilitate feeding in conjunction with hirudin, a well-described anticoagulant produced by the leech salivary glands [46-49].

In addition to the highly abundant *Pedobacter*, a number of other bacteria were discovered within the mucus. We can only speculate as to their function in regards to the mucosal niche, however, previous studies on these lineages may help elucidate their roles. For example, the aerotolerant *Desulfovibrio sp*. are typically found in aquatic environments rich in organic materials and contribute towards the reduction of sulfate into sulfide [50]. Interestingly, members of the Comamonadaceae family are also capable of both sulfate cycling [51] (suggesting the presence of sulfate in the mucosal environment), in addition to denitrification in aqueous environments [52]. Members of the Methylophilaceae family are also involved in denitrification [53]. The functional roles of both Comamonadaceae and Methylophilaceae

families indicate a potentially high level of denitrification activity within leech shed mucus. As the mucus is sloughed from the anterior to posterior end of the leech [9], and the leech nephridial system harbors 17 pairs of nephridia and bladders that empty their contents along the entire length of the leech [13], the mucus could contain significant levels of nitrogen and sulfate from leech waste [13, 54-56], encouraging the growth of these denitrifiers.

A number of bacteria described within our leech-exuded mucus samples have recently been implicated as frequently occurring DNA contaminants of microbiomes due to reagent and laboratory contamination [57]. These microbial contaminants prove to be a significant problem particularly within low biomass (Salter et al., 2014), i.e. 10^3 or 10^4 CFU/mL, and low quality/dilute (Lusk, 2014) samples where they tend to eclipse the true microbial community. However, the leech mucosal microbiota consists of a greater biomass than this critical threshold, with culturable microbes estimated to be at least 10^4 CFU/mL, which falls short of accounting for unculturable microbes. Furthermore, by streaking mucus on various nutrient agar plates using traditional culturing approaches, a number of low abundance bacteria (i.e. each comprising <2% of total Illumina reads) were isolated, including *Aeromonas*, *Morganella*, *Stenotrophomonas*, *Delftia*, *Chryseobacterium*, and *Variovorax* species, verifying their presence and viability within mucus. In addition, the presence and proliferation of both *Aeromonas* and *Rikenella*-like bacteria within exuded mucus have been confirmed using qPCR ([9] and current results), independent of Illumina amplicon generation and sequencing.

Interestingly, our most common mucosal community member, *Pedobacter*, is also recognized as a high frequency contaminant [57]. However, we have determined that *Pedobacter* is not universally present within all leech-related samples, e.g. leech epithelial swabs, although the same DNA isolation buffers and procedures are used in sample processing. Additionally,

preliminary metatranscriptome analyses of the exuded mucosal microbiome show that *Pedobacter* contributes a significant proportion of transcriptional activity within this environment (Ott; pers. obs.). These results support our conclusion that *Pedobacter* are likely not resulting from contamination but are bona fide community members of shed mucus.

Following the identification of a second predominant gut symbiont, *Rikenella*-like bacteria, also within mucus, its population dynamics were examined. As the dynamics between *A. veronii* and the *Rikenella*-like bacteria have been examined extensively within the leech host [15, 16], we were interested to see if the relationship between these two symbionts may persist beyond the gut environment. We examined the growth of the *Rikenella*-like bacteria over a period of 8 days and we showed that the population dynamics mirrored those of *A. veronii* within shed mucus [9], suggesting a continuation of their syntrophic activities into this environment. As *Rikenella*-like bacteria in the crop are known to degrade host mucin glycans to short chain fatty acids, which serve as a carbon source for *A. veronii* [15], it is not difficult to extrapolate this functional role to the glucosaminoglycan-rich mucosal environment. Therefore, *Rikenella*-like bacteria may serve a similar role in the maintenance and proliferation of mucosal *A. veronii* as in the gut. In return, it is possible that *A. veronii* ensures a habitable environment for the obligately anaerobic *Rikenella*-like bacteria through the removal of oxygen, indicating a relationship beyond one based solely on nutrition. Mucosal seeding through the digestive tract [9] and the eventual decrease in population may also occur in tandem, with symbionts either dispersing or dying after approximately 3 days post shedding, likely due to the exhaustion of resources.

The microbial community of leech mucus is diverse and harbors a number of novel bacteria that have not been previously associated with the leech. The discovery of this rich microbial community within mucus not only raises questions involving microbial interactions

that may occur within this environment, but also whether these other bacteria utilize mucus as a transmission substrate for the infection of novel hosts. As many members of a host microbiota are thought to be transmitted as assemblages [37, 58-60], future work will explore the function of mucus as a symbiont mixing vessel by identifying loci which may be under different modes of selection (i.e. diversifying versus purifying) and the possibility of gene swapping in this environment. It is feasible that the mucosal environment provides symbionts with the opportunity to transmit to novel hosts as well as increase/mix their genetic repertoire, potentially enhancing their ability to adapt to changing environments.

REFERENCES

1. Krediet, C.J., et al., *Utilization of mucus from the coral Acropora palmata by the pathogen Serratia marcescens and by environmental and coral commensal bacteria.* Appl. Environ. Microbiol., 2009. **75**(12): p. 3851-8.
2. Rohwer, F., et al., *Diversity and distribution of coral-associated bacteria.* Mar. Ecol. Prog. Ser., 2002. **243**: p. 1-10.
3. Sekar, R., et al., *Microbial communities in the surface mucopolysaccharide layer and the black band microbial mat of black band-diseased Siderastrea siderea.* Appl. Environ. Microbiol., 2006. **72**(9): p. 5963-73.
4. Sharon, G. and E. Rosenberg, *Bacterial growth on coral mucus.* Curr. Microbiol., 2008. **56**(5): p. 481-8.
5. Shnit-Orland, M. and A. Kushmaro, *Coral mucus-associated bacteria: a possible first line of defense.* FEMS Microbiol. Ecol., 2009. **67**(3): p. 371-80.
6. Nyholm, S.V., et al., *Establishment of an animal-bacterial association: recruiting symbiotic vibrios from the environment.* Proc. Natl. Acad. Sci. U.S.A., 2000. **97**(18): p. 10231-5.
7. Fraune, S., et al., *In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides.* Proc. Natl. Acad. Sci. U.S.A., 2010. **107**(42): p. 18067-72.
8. Thomas, Y., et al., *Survival of influenza virus on banknotes.* Appl. Environ. Microbiol., 2008. **74**(10): p. 3002-7.
9. Ott, B.M., et al., *Hitchhiking of host biology by beneficial symbionts enhances transmission.* Sci. Rep., 2014. **4**: p. 5825.

10. Rio, R.V., et al., *Symbiont succession during embryonic development of the European medicinal leech, Hirudo verbana.* Appl. Environ. Microbiol., 2009. **75**(21): p. 6890-5.

11. Michalsen, A., M. Roth, and G. Dobos, *Medicinal leech therapy*. 2007, New York, NY: Thieme Medical Publishers.

12. Maltz, M.A., et al., *Metagenomic analysis of the medicinal leech gut microbiota.* Front. Microbiol., 2014. **5**: p. 151.

13. Sawyer, R.T., *Leech biology and behavior*. 1986, Oxford, United Kingdom: Clarendon Press.

14. Kikuchi, Y. and J. Graf, *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**(6): p. 1984-91.

15. Bomar, L., et al., *Directed culturing of microorganisms using metatranscriptomics.* MBio, 2011. **2**(2): p. e00012-11.

16. Kikuchi, Y., L. Bomar, and J. Graf, *Stratified bacterial community in the bladder of the medicinal leech, Hirudo verbana.* Environ. Microbiol., 2009. **11**(10): p. 2758-70.

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

18. Holmes, D.S. and J. Bonner, *Preparation, molecular weight, base composition, and secondary structure of giant nuclear ribonucleic acid.* Biochemistry, 1973. **12**(12): p. 2330-8.

19. Klindworth, A., et al., *Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies.* Nucleic Acids Res., 2013. **41**(1): p. e1.

20. Kozich, J.J., et al., *Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform.* Appl. Environ. Microbiol., 2013. **79**(17): p. 5112-20.

21. Schloss, P.D., et al., *Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities.* Appl. Environ. Microbiol., 2009. **75**(23): p. 7537-41.

22. Schloss, P.D. and J. Handelsman, *Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness.* Appl. Environ. Microbiol., 2005. **71**(3): p. 1501-6.

23. Lane, D.J., *16S/23S rRNA sequencing*, in *Nucleic acid techniques in bacterial systematics*, E.S.M. Goodfellow, Editor. 1990, John Wiley and Sons: Chichester, United Kingdom.

24. Weisburg, W.G., et al., *16S ribosomal DNA amplification for phylogenetic study.* J. Bacteriol., 1991. **173**(2): p. 697-703.

25. Worthen, P.L., C.J. Gode, and J. Graf, *Culture-independent characterization of the digestive-tract microbiota of the medicinal leech reveals a tripartite symbiosis.* Appl. Environ. Microbiol., 2006. **72**(7): p. 4775-81.

26. Swofford, D.L., *PAUP 4.0-Phylogenetic Analysis Using Parsimony. Version 4.* 2002, Sinauer Associates: Sunderland, MA.

27. Felsenstein, J., *Confidence limits on phylogenies: an approach using the bootstrap.* Evolution, 1985. **39**: p. 783-791.

28. Ronquist, F. and J.P. Huelsenbeck, *MrBayes 3: Bayesian phylogenetic inference under mixed models.* Bioinformatics, 2003. **19**(12): p. 1572-4.

29. Nylander, J.A.A., *MrModeltest v2.* 2004, Uppsala University, Uppsala, Sweden: Evolutionary Biology Centre.

30. Larget, B. and D.L. Simon, *Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees.* Mol. Biol. Evol., 1999. **16**: p. 750-9.

31. Shannon, E.H., *A Mathematical Theory of Communication.* Bell Syst. Tech. J., 1948. **27**: p. 379-423, 623-56.

32. Hurlbert, S.H., *The nonconcept of species diversity: a critique and alternative parameters.* Ecology, 1971. **52**(577-586).

33. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res., 1997. **25**(17): p. 3389-402.

34. Gotelli, N.J. and R.K. Colwell, *Estimating Species Richness*, in *Frontiers in Measuring Biodiversity*, M.A. E. and M.B. J., Editors. 2011, Oxford University Press: New York. p. 39-54.

35. Utevsky, S., et al., *Distribution and status of medicinal leeches (genus Hirudo) in the Western Palaearctic: anthropogenic, ecological, or historical effects?* Aquat. Conserv., 2010. **20**(2): p. 198-210.

36. Sachs, J.L., et al., *The evolution of cooperation.* Q. Rev. Biol., 2004. **79**(2): p. 135-60.

37. Ebert, D., *The Epidemiology and Evolution of Symbionts with Mixed-Mode Transmission.* Annu. Rev. Ecol. Evol. Syst., 2013. **44**: p. 623-643.

38. Nelson, M.C., et al., *Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys.* PLoS One, 2014. **9**(4): p. e94249.

39. Kautz, S., et al., *Surveying the microbiome of ants: comparing 454 pyrosequencing with traditional methods to uncover bacterial diversity.* Appl. Environ. Microbiol., 2013. **79**(2): p. 525-34.

40. Shendure, J. and H. Ji, *Next-generation DNA sequencing.* Nat. Biotechnol., 2008. **26**(10): p. 1135-45.

41. An, D.S., et al., *Pedobacter daechungensis sp. nov., from freshwater lake sediment in South Korea.* Int. J. Syst. Evol. Microbiol., 2009. **59**(Pt 1): p. 69-72.

42. Chun, J., J.Y. Kang, and K.Y. Jahng, *Pedobacter pituitosus sp. nov., isolated from Wibong falls.* Int. J. Syst. Evol. Microbiol., 2014.

43. Joung, Y., H. Kim, and K. Joh, *Pedobacter yonginense sp. nov., isolated from a mesotrophic artificial Lake in Korea.* J. Microbiol., 2010. **48**(4): p. 536-40.

44. Payza, A.N. and E.D. Korn, *Bacterial degradation of heparin.* Nature, 1956. **177**(4498): p. 88-9.

45. Steyn, P.L., et al., *Classification of heparinolytic bacteria into a new genus, Pedobacter, comprising four species: Pedobacter heparinus comb. nov., Pedobacter piscium comb. nov., Pedobacter africanus sp. nov. and Pedobacter saltans sp. nov. proposal of the family Sphingobacteriaceae fam. nov.* Int. J. Syst. Bacteriol., 1998. **48 Pt 1**: p. 165-77.

46. Haycraft, J.B., *On the action of a secretion obtained from the medicinal leech on the coagulation of the blood.* Proc. R. Soc. B, 1884. **36**: p. 478-87.

47. Haycraft, J.B., *Über die Einwirkung eines Sekretes des officiellen Blutegels auf die Gerinnbarkeit des Bluts.* Naunyn Schmiedebergs Arch. Exp. Pathol. Pharmacol., 1894. **18**: p. 209-17.

48. Jacoby, Y., *Über Hirudin.* Dtsch Med. Wochenschr., 1904. **30**: p. 786-7.

49. Markwardt, F., *Hirudin as alternative anticoagulant--a historical review.* Semin. Thromb. Hemost., 2002. **28**(5): p. 405-14.

50. Adams, M.E. and J.R. Postgate, *A new sulphate-reducing vibrio.* J. Gen. Microbiol., 1959. **20**(2): p. 252-7.

51. Schmalenberger, A., et al., *The role of Variovorax and other Comamonadaceae in sulfur transformations by microbial wheat rhizosphere communities exposed to different sulfur fertilization regimes.* Environ. Microbiol., 2008. **10**(6): p. 1486-500.

52. Khan, S.T., et al., *Members of the family Comamonadaceae as primary poly(3-hydroxybutyrate-co-3-hydroxyvalerate)-degrading denitrifiers in activated sludge as revealed by a polyphasic approach.* Appl. Environ. Microbiol., 2002. **68**(7): p. 3206-14.

53. Kalyuhznaya, M.G., et al., *Methylophilaceae link methanol oxidation to denitrification in freshwater lake sediment as suggested by stable isotope probing and pure culture analysis.* Environ. Microbiol. Rep., 2009. **1**(5): p. 385-92.

54. Dev, B., *Excretion and Osmoregulation in the Leech, Hirudinaria Granulosa (Savigny).* Nature, 1964. **202**: p. 414.

55. Nicholls, J.G. and S.W. Kuffler, *Extracellular Space as a Pathway for Exchange between Blood and Neurons in the Central Nervous System of the Leech: Ionic Composition of Glial Cells and Neurons.* J. Neurophysiol., 1964. **27**: p. 645-71.

56. Zerbst-Boroffka, I., *Organische Säurereste als wichtigste Anionen im Blut von Hirudo medicinalis.* Z. vergl. Physiol., 1970. **70**: p. 313-321.

57. Salter, S.J., et al., *Reagent and laboratory contamination can critically impact sequence-based microbiome analyses.* BMC Biol., 2014. **12**(1): p. 87.

58. Aagaard, K., et al., *The placenta harbors a unique microbiome.* Sci. Transl. Med., 2014. **6**(237): p. 237ra65.

59. Dominguez-Bello, M.G., et al., *Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns.* Proc. Natl. Acad. Sci. U.S.A., 2010. **107**(26): p. 11971-5.

60. Makino, H., et al., *Mother-to-infant transmission of intestinal bifidobacterial strains has an impact on the early development of vaginally delivered infant's microbiota.* PLoS One, 2013. **8**(11): p. e78331.

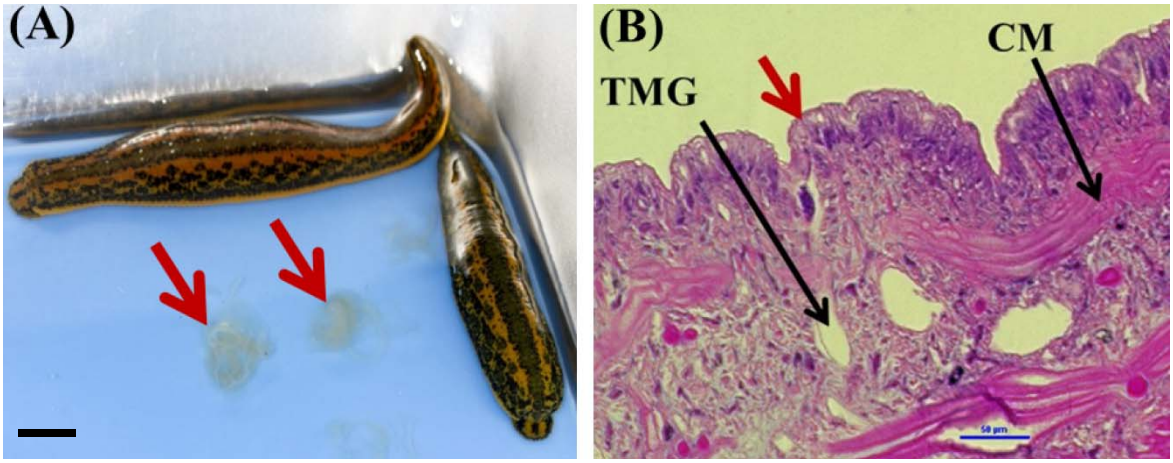

Figure 3-1. Medicinal leech mucosal secretions**.** A. *Hirudo verbana* leeches, with arrows indicating shed mucosal secretions. Scale bar indicates 0.7 cm. B. The epithelial layer of the leech (20X magnification, Hemotoxylin and eosin (H&E) stain). Red arrow indicates tubular mucus gland cell opening. TMG=tubular mucus gland cell (one of two types of mucus gland cells [13]), CM=circular muscle.

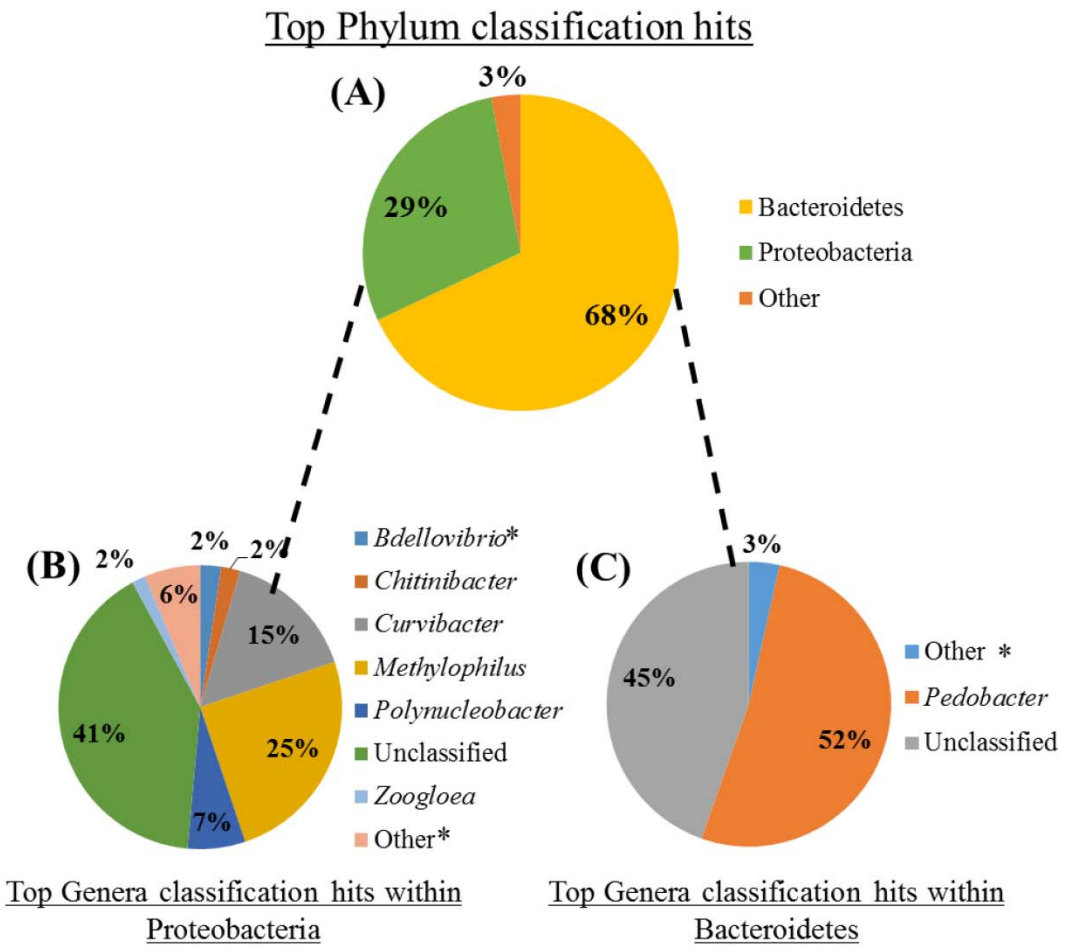

Figure 3-2. Microbiota composition of shed leech mucus obtained through Illumina sequencing of the V3-V4 hypervariable region of 16S rRNA gene**.** A. The relative abundance of total reads, following quality control, within phyla. B. The relative abundance of reads, comprising ≥2% of total reads, of genera housed within Proteobacteria. * indicates previously described within the leech [16,25]. C. The relative abundance of reads, comprising ≥1% of total reads, of genera housed within Bacteroidetes.

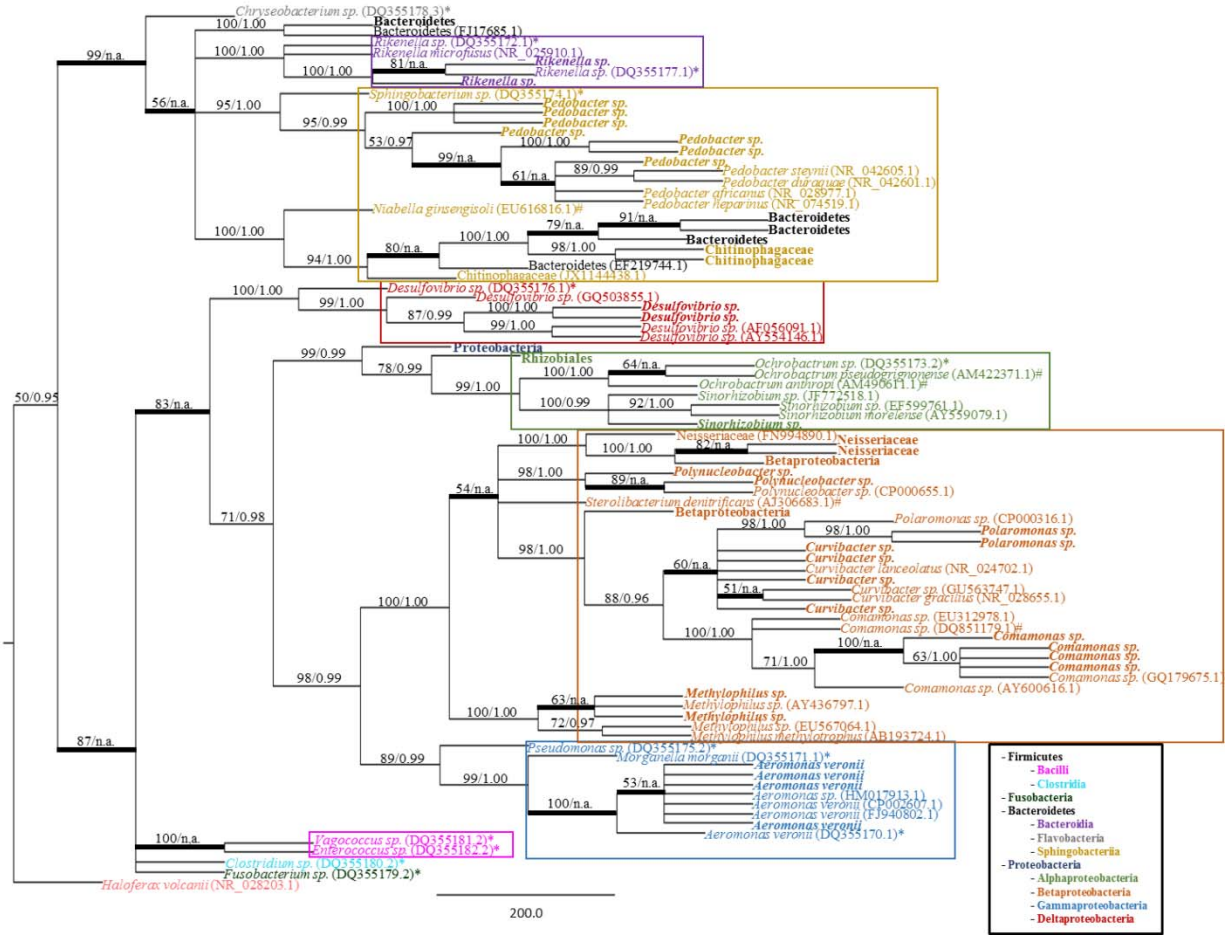

Figure 3-3. Bacterial phylogeny from shed leech mucus. Molecular phylogenetic tree of 16S rRNA gene sequences amplified from adult *H. verbana* mucosal secretions. A MP analysis tree created from approximately 900 aligned nucleotides is shown. Significance values, represented in MP bootstrap and Bayesian PP (BS/PP), are indicated at respective nodes. A bolded branch with "BS/n.a." significance value refers to a branch that was not statistically supported by Bayesian analysis. Branch lengths are measured in number of substitutions over the whole sequence. Representative 16S rRNA sequences obtained within shed mucus are in bold, with other sequences obtained from NCBI indicated by accession numbers. Previously described leech isolates obtained through study of the gut and bladder systems are indicated by an * [25] or # [16], respectively. Color blocks indicate the Class housing the representative sequences.

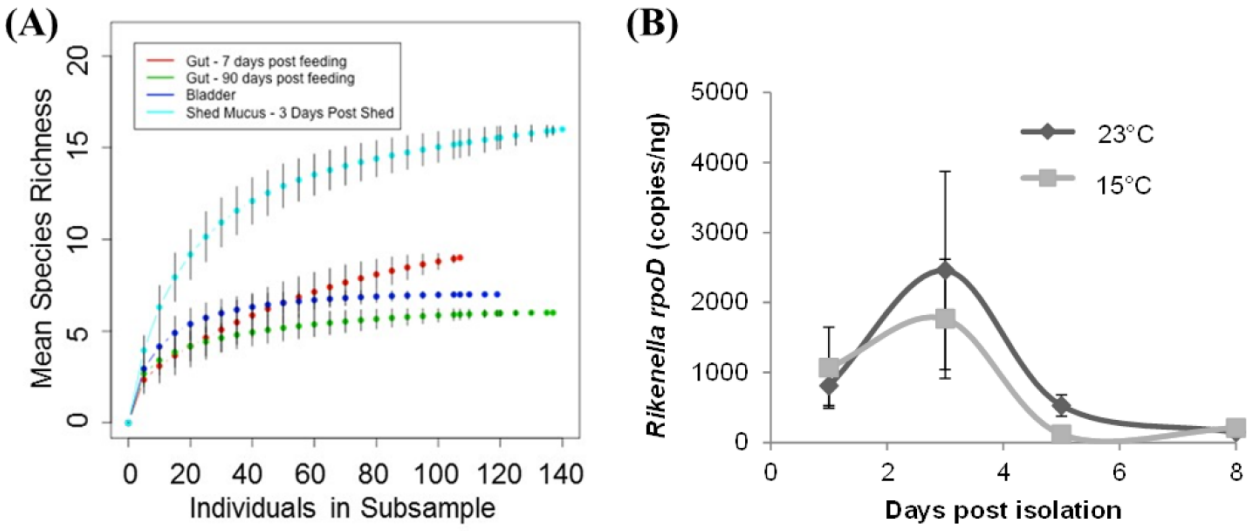

Figure 3-4. Leech mucosal secretions contain a high diversity of microbial symbionts, to include the *Rikenella*-like bacterium**.** A. Rarefaction curves were generated using Hurlbert's formulation [9] and are plotted with the mean species richness (y-axis) as a function of the number of individuals in each subsample (x-axis). The asymptotes rise less steeply as an increasing proportion of OTUs have been encountered. Error bars represent standard error. B. *Rikenella* density within mucus increases from Day 1 to 3, decreasing thereafter, mirroring the *A. veronii* population dynamics [32]. *Rikenella* density was measured via qPCR using the *rpoD* gene. No significant differences were observed either between the two temperature regimens (p=0.6868) or as the mucus aged (p=0.0518, one-way ANOVA, n=39).

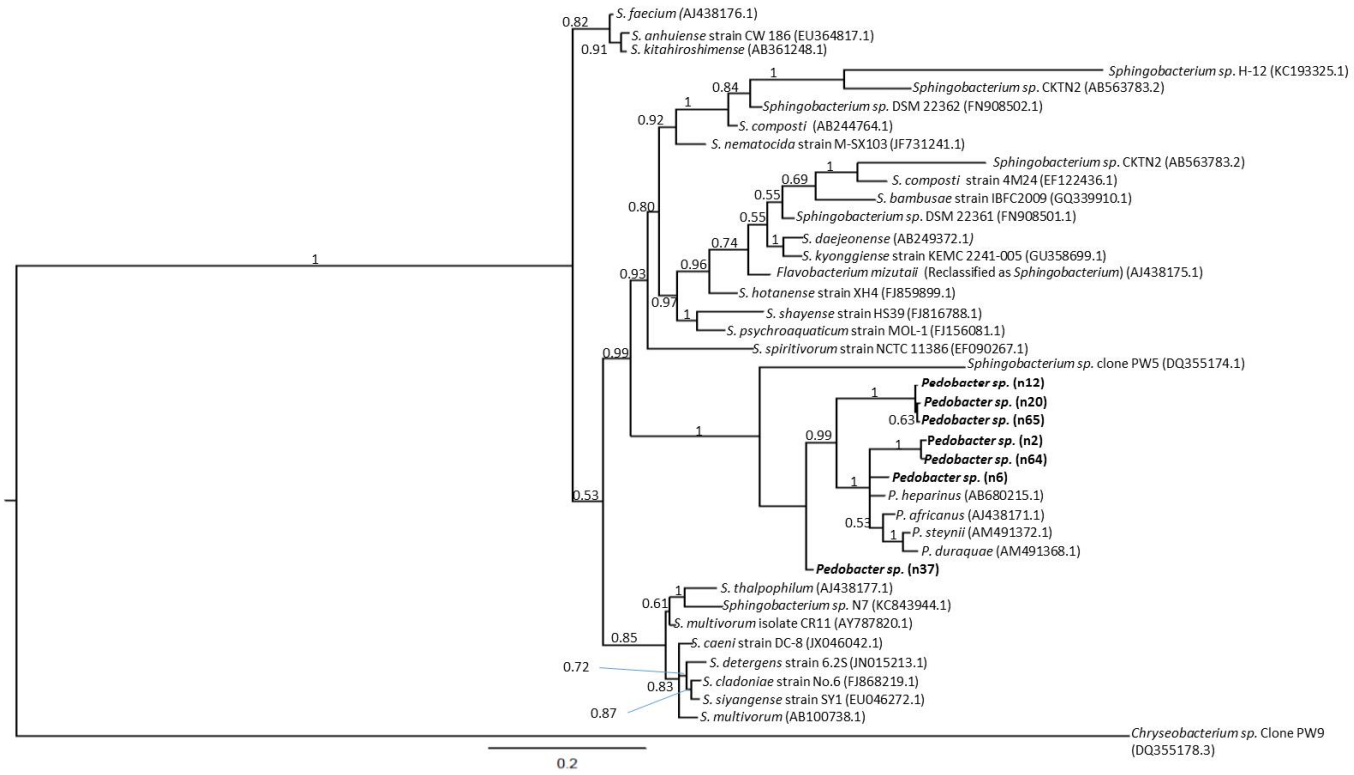

Figure 3-S1. Molecular 16S rRNA phylogenetic tree exhibiting the separation between the novel leech symbiont, Pedobacter sp., and its close relative, *Sphingobacterium sp.* A Bayesian analysis tree created from approximately 1400 aligned nucleotides is shown. Significance values, represented in Bayesian PP, are indicated at respective nodes. Branch lengths are measured in number of substitutions over the whole sequence. Representative *Pedobacter sp.* 16S rRNA sequences obtained within shed mucus are in bold, with other sequences obtained from NCBI indicated by accession numbers.

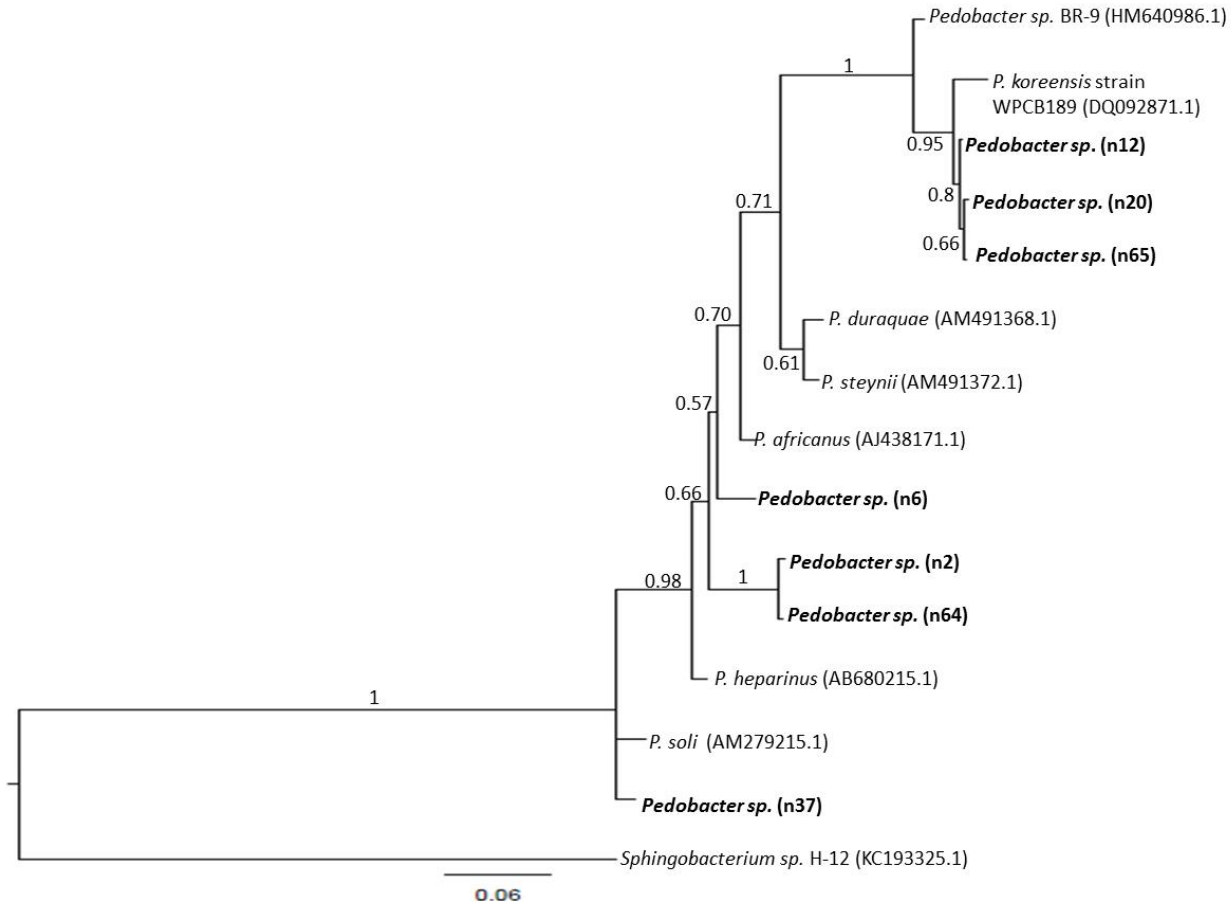

Figure 3-S2. Molecular 16S rRNA phylogenetic tree of the novel leech symbiont, *Pedobacter sp***.** A Bayesian analysis tree created from approximately 1400 aligned nucleotides is shown. Significance values, represented in Bayesian PP, are indicated at respective nodes. Branch lengths are measured in number of substitutions over the whole sequence. Representative *Pedobacter sp.* 16S rRNA sequences obtained within shed mucus are in bold, with other sequences obtained from NCBI indicated by accession numbers.

Table 3-1. 16S rRNA sequences obtained from adult medicinal leech, *H. verbana*, mucosal secretions.

Accession Number	Bacterial Division	Class	Order	Family	Genus	Species	No (%) of Clones
	Bacteroidetes[ab*]						9 (6.4%)
	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	***Rikenella/Alistipes***[b*]		2 (1.4%)
		Sphingobacteriia	Sphingobacteriales	Chitinophagaceae			9 (6.4%)
				Sphingobacteriaceae	*Pedobacter*[**]		32 (23.0%)
	Proteobacteria[b]						1 (0.7%)
	Proteobacteria	Alphaproteobacteria	Rhizobiales				1 (0.7%)
				Rhizobiaceae	*Sinorhizobium/Ensifer group*		1 (0.7%)
		Betaproteobacteria[a*]					5 (3.6%)
		Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Polynucleobacter*		5 (3.6%)
				Comamonadaceae	*Polaromonas*		11 (7.9%)
					Comamonas[a]		9 (6.4%)
					Curvibacter[*]		11 (7.9%)
			Methylophilales	Methylophilaceae	*Methylophilus*		5 (3.6%)
			Neisseriales	Neisseriaceae			3 (2.1%)
		Gammaproteobacteria	Aeromonadales	Aeromonadaceae	***Aeromonas***	***veronii***[b**]	31 (22.1%)
		Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae	***Desulfovibrio***[b]		5 (3.6%)

Isolates appearing in bold were previously described in 16S rRNA clone libraries of leech excretory (a., [16]) and digestive (b., [25]) systems; * and ** indicate two and three different RFLP types identified, respectively.

Table 3-2. Species richness and coverage estimation of 16S rRNA clone libraries generated from *H. verbana* microbial niches.

Richness Estimation					
Sample Identity	**Chao1**	**Chao2**	**ACE**	**ICE**	**Shannon-Weaver**
Shed Mucus-3 days post shed	20.5	20.5	16.69	16.69	2.341
Gut-7 days post feeding[a]	11.25	11.25	11.1	11.1	1.112
Gut-90 days post feeding[a]	6	6	6	6	1.377
Bladder[b]	7	7	7	7	1.398

Isolates previously described in a) [25] (richness estimation was calculated for combined crop and intestinum data); b) [16].

Chapter 4A: A Tale of Transmission: *Aeromonas veronii* Activity within Leech Exuded Mucus[3]

ABSTRACT

Transmission exposes symbionts to new and often stressful environmental conditions. With horizontal transmission, these different conditions represent major lifestyle shifts. We provide the first comprehensive and comparative analysis of the global transcriptional profiles of a symbiont as it shifts between lifestyles during transmission. The Gammaproteobacterium, *Aeromonas veronii*, is transmitted from the gut of the leech to other hosts via host mucosal castings, yet *A. veronii* can also transition from mucosal habitancy to free-living. These three lifestyles are characterized by distinct physiological constraints and consequently lifestyle-specific changes in the expression of stress-response genes. Mucus-bound *A. veronii* had the greatest expression, in both quantity of loci and levels of transcription, of stress-response mechanisms, yet these bacteria are able to proliferate within the mucus, suggesting the availability of nutrients within this environment. We found that *A. veronii* alters transcription of loci in a synthetic pathway that incorporates N-acetylglucosamine (NAG; a major component of mucus) into the bacterial cell wall, enabling proliferation. Our results demonstrate that symbionts undergo dramatic, widespread transcriptional changes throughout the process of transmission that allow them to thrive while encountering new environments which further shape their ecology and evolution.

IMPORTANCE

Transmission is critical to the establishment and persistence of host-associated microbiotas. Transmission can occur via one of three routes; vertical (i.e. parental), horizontal

[3] Reprinted from Ott BM, Dacks AM, Ryan KJ, Rio RVM. Comings and Goings within a Transmission Substrate: *Aeromonas veronii* Activity within Leech Exuded Mucus. 2015. *In prep.*

(i.e. environmental/social acquisition), or mixed mode (i.e. incorporating vertical and horizontal mechanisms). Although most studies examine microbial ecology within purely vertical transmitted associations, little is known about microbial activity within host-derived substrates that enable the, more frequent, horizontal transmission of members of the microbiota. Horizontal transmission is often characterized by dramatic lifestyle transitions from a host-associated to free-living state. In this study, we used RNA-seq and immunostaining to describe how bacteria may adapt to a transmission substrate, while remaining viable and proliferating until a novel niche is encountered. We identify pivotal factors, specifically in stress-response and nutrient acquisition mechanisms, likely mediating the transmission of the majority of host-associated microbiotas, while emphasizing the significance of this "in-between" state towards microbial ecology and evolution.

INTRODUCTION

Mutualistic bacteria provide numerous advantages to their eukaryotic hosts, such as the provisioning of essential nutrients [1, 2], protection from pathogens [3, 4], and immunological priming [5]. These mutualists also serve as some of the best examples of the extended phenotype [6], by providing opportunities for rapid adaptation and resilience towards dynamic ecological conditions including broadening the host dietary range via enzymatic activities (reviewed in [7, 8]), detoxification mechanisms (reviewed in [9, 10]), or enhancing tolerance to environmental stressors such as fluctuating temperatures [11]. Although the persistence of these relations over evolutionary time is crucial for securing adaptive potential, little is known about the microbial ecology within transmission settings.

Mixed mode transmission (MMT), integrating components of both vertical and horizontal microbial acquisition, is the predominant mechanism for obtaining microbial symbionts by a host (reviewed in [12]). A recently described model of MMT involves the acquisition of the

heterotrophic, Gammaproteobacterium *Aeromonas veronii,* a predominant member, along with the Bacteroidetes *Mucinivorans hirudinis* [13, 14], of the European medicinal leech (*Hirudo verbana*) digestive tract microbiota. Specifically, *A. veronii* is found within the leech crop [13], which is the largest component of the digestive tract where the blood meal is stored over a period of weeks to months [15]. Similar relationships between *Aeromonas spp.* and other leeches include those involving additional members of the European medicinal leech sister complex; *H. medicinalis* and *H. orientalis,* with *A. hydrophila* [16](reviewed in [15]) and either *A. jandaei* or *A. veronii* [16], respectively. In the *H. verbana* symbiosis, acquisition of *A. veronii* incorporates imperfect vertical transmission via the leech cocoon [17] complemented by environmental acquisition through regularly secreted mucus that conspecific leeches actively seek in their environment [18]. Although *A. veronii* is a minor microbial constituent of these mucosal secretions, it has a high prevalence, being found in all wild-type mucosal samples [18, 19]. Within the mucus, *A. veronii* is capable of surviving and even proliferating at a tempo that syncs with host shedding frequency [18]. The horizontal transmission of *A. veronii* likely enables additional lifestyles including as a human pathogen (reviewed in [20]), free-living in H_2O [21], and as a zebrafish symbiont [22]. Each lifestyle represents a major shift on the physiological constraints exerted upon *A. veronii* by a variety of ecological factors and thus adjustments to gene expression must occur for persistence under these different conditions.

Shifting niches results in changes in several environmental factors, including ambient temperatures [23] and nutrient resources (reviewed in [24]), with significant impacts on the "transcriptional tuning" of organisms necessitating a balance between several responses, including growth and stress (reviewed in [24]). Stress is a known selection agent [25] for either pre-existing or novel mutations within the population [25]. Several factors likely make the transition from gut to mucosal occupancy stressful for *A. veronii*, including alterations in nutrient

availability (i.e. blood to glucosaminoglycan environment), abrupt shifts in atmospheric conditions (i.e. anaerobic to aerobic, temperature, etc), and the likely rise in competition due to increased community diversity [19]. Notably, *A. veronii* is capable of not only remaining viable after such a sudden lifestyle shift, but also replicating to a density high enough for transmission to another leech individual [18], indicating the availability of nutrients to support population growth. However, the extent to which *A. veronii* gene expression shifts in response to these dramatic environmental changes are unknown.

In this study, we addressed three major questions. First, how does *A. veronii* respond to a lifestyle shift from mutualistic to a free-living state? Second, how is *A. veronii* capable of surviving in the mucosal environment that likely involves numerous stressors? Finally, what nutrient sources enable *A. veronii* to proliferate within shed mucus? To address these questions, high-throughput Illumina-based deep sequencing was used to characterize the *A. veronii* transcriptome within leech mucosal secretions. Differential RNA-seq analyses were used to compare *A. veronii* transcript abundance within leech mucosal secretions to the leech digestive tract and within monocultures [26]. In addition to these transcriptome analyses, lectin staining identified N-acetylglucosamine (NAG; a major glycan element of peptidoglycan) as an abundant component of shed mucus and the bacterial cell membrane. We describe the expression profile of loci involved in the import and processing of NAG that would facilitate its incorporation into peptidoglycan, thereby enabling *A. veronii* proliferation within this niche. While the antimicrobial properties of mucosal secretions have received strong interest [27, 28], this is the first study that focuses on the activity of a microbe within a transmission substrate. Here, we describe the significance of this "in-between" state for microbial ecology and evolution by determining the factors that enable the transition between different lifestyles.

MATERIALS AND METHODS

(a) Leech husbandry. *H. verbana* leeches were obtained from Leeches USA (Westbury, NY, USA) and maintained in sterilized, Leech Strength Instant Ocean H_2O (0.004%, I.O.) in the Department of Biology at West Virginia University at 15°C. Leeches were maintained on defibrinated bovine blood (Hemostat, CA) through an artificial feeding system.

(b) Mucosal sampling. Tanks housing leeches were cleaned to remove all mucus and fresh Leech Strength I.O. was added (denoted as Day 0). Mucosal samples collected the following day were considered 1 d and were aged to 3 d at 15°C within sterile Leech Strength I.O. H_2O.

(c) RNA extraction and library preparation. Total RNA was extracted from three pooled wild-type leech mucus samples using the MasterPure RNA purification kit (Epicentre, Madison WI) following the Fluid Samples protocol. DNA was removed from the RNA samples using the Turbo DNA-*free* kit (Ambion, Austin, TX) using the Rigorous DNase treatment option. DNA removal was confirmed through *Aeromonas*-specific primers AvgyrB-F and AvgyrB-R with the amplification settings in Table 1 and an RNA template lacking a reverse-transcription step.

Total RNA was quantified using the NanoDrop 2000 (Thermo Scientific, Waltham, MA) and integrity was examined using the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA). The pooled-mucus RNA sample with the highest integrity was then processed with the Ribo-Zero Magnetic Kit (Gram-Negative Bacteria) (Epicentre, Madison, WI) using the Individual Washing Procedure. The resulting mRNA-enriched RNA was then purified using the RNeasy MinElute Cleanup Kit (Qiagen, Valencia, CA). Approximately 250 ng of the enriched mRNA was ethanol-precipitated and re-eluted in Elute, Prime, Fragment mix. The eluted mRNA was then processed using the TruSeq RNA Sample Prep Kit v2 (Illumina, San Diego, CA) by the WVU Genomics Core Facility and the libraries were sequenced using the Illumina MiSeq (Illumina, San Diego, CA) (2 x 250 bp) at the WVU Genomics Core Facility.

(d) Transcriptome analyses. Transcriptomes from the *A. veronii* monoculture and leech crop previously-published [26] were used for comparative purposes in our analyses (Table 2). It is important to note that, while paired-end reads and a higher starting quantity of RNA (i.e. 250 ng for mucus compared with 100 ng for gut and culture, due to *A. veronii* being a minor constituent of exuded mucus) were used for the generation of the mucosal RNA-seq library, the preparation, sequencing and bioinformatics analyses of the RNA-seq libraries were identical for all three environments. In support, interlaboratory variation in transcriptome profiling due to technical variation is minimal given proper standardization [29].

The cDNA reads from all 3 libraries were assembled *de novo* into contiguous sequences (contigs) using Trinity [30], with those specific to *A. veronii* identified by mapping to the *A. veronii* strain HM21 genome [31] using BLASTn [32]. Bowtie2 [33], using default parameters, and custom perl scripts (Niel Infante, pers. comm.) were utilized to determine the number of reads corresponding to each contig. The functional annotation of contigs was accomplished using the NCBI non-redundant (nr) database housed on the WVU Shared Research Facilities (SRF) High Performance Computing (HPC) Spruce Knob host. The resulting BLAST files were organized and visualized using MEGAN5 [34], which incorporates Kyoto Encyclopedia of Genes and Genomes (KEGG) and SEED databases. Gene expression values (EV), which normalize transcripts based on gene length [35], were determined using the following formula:

$$\frac{\text{number of reads mapped}}{\text{length of gene (kb)} \times \frac{\text{total number of reads}}{1{,}000{,}000}}$$

to enable analysis of differential gene expression [26] and to minimize systematic variation between samples as well as differences in library composition [36]. Data from this study were deposited in the NCBI Sequence Read Archive under the Accession SRP067591.

Gene ontology (GO) terms for each *A. veronii*-specific contig were determined using Blast2GO, followed by enrichment analysis using Fisher's Exact Test [37] with a false discovery rate (FDR) correction. Differences in EV between KEGG functional categories (i.e. Cellular processes (CP), Environmental information processing (EIP), Genetic information processing (GIP), and Metabolism (Met)) within each environment (i.e. mucus, gut or monoculture) were also analyzed by first transforming values to log(EV+1) and testing for statistical significance using a 1-way ANOVA with SAS JMP (v.10) [38]. Graphs were built using JMP Graph Builder. Fold change was determined as the ratio of EV within the mucosal environment compared with either the culture or gut. Significance was reported with a p-value ≤ 0.05.

(e) Lectin staining and Immunocytochemistry. Lectin staining was used to visualize the distribution of N-acetylglucosamine (NAG), predicted to be the major mucopolysaccharide component of leech mucus [39], and mannose, which was not predicted to be present as mannose is not a component of mucopolysaccharides, and was thus used as a specificity control. A GFP-expressing *Aeromonas* strain (HM21S::Tn7gfp, Graf Lab, University of Connecticut) was orally administered to leeches to replicate a natural shedding scenario into the mucosal secretions [18]. Subsequently, GFP immunocytochemistry was used to visualize *Aeromonas*-localization within the mucus. Mucus castings were placed in a 24-well culture plate and fixed using 4% paraformaldehyde followed by dehydration using an ethanol series (i.e. incubated in 100% ethanol for 5 min, 3 times; followed by a 30 min evaporation step). To ensure lectin binding specificity, a subset of mucosal casts were inoculated with mannose *in vitro*. Slides were then blocked in blocking solution (1% bovine serum albumen (BSA)-0.3% Triton X-100 in PBS) for 1 h, followed by an overnight incubation at room temperature (RT) with 20 µg/mL of Rhodamine-labeled succinylated Wheat Germ Agglutinin (Vector Labs, Burlingame, CA; RL-1022S) and Fluorescein-labeled Concanavalin A (mannose; Vector Labs, Burlingame, CA; FL-

1001) in blocking solution. Slides were washed in a PBS-buffer solution series, blocked for 1 h (see above), followed by another overnight incubation at RT with a 1:1,000 dilution of rabbit anti-GFP primary antibody (Life Technologies, A11122) in PBSAT (1% sodium azide; 2% BSA). Slides were subsequently washed in a PBS-buffer solution series, blocked again for 1 h, followed by another overnight incubation at RT using a 1:1000 dilution of goat anti-rabbit 633 secondary antibody (Life Technologies, A21070) in PBSAT. Following overnight incubation, slides were washed in a PBS-buffer solution series, followed by washes in 40%, 60% and 80% glycerol. Samples were then mounted using a hard-set mounting media containing DAPI (Vector Labs, Burlingame, CA; H-1500). A no-lectin sample was used as a negative control. Mucosal casts supplemented with mannose exhibited a high signal in the targeted wavelength, while those without mannose showed little to no signal (data not shown), indicating that the lectin-based dyes do not bind mucus non-specifically. Slides were visualized using an Olympus Fluoview FV1000 Confocal Microscope equipped with argon and green HeNe lasers and appropriate filters and an Olympus AX70 Fluorescent microscope. Images were examined and adjusted to appropriate contrast and brightness using Olympus Fluoview Software.

(f) RT-PCR of NAG-related genes under varying NAG concentrations. As the mucosal environment was hypothesized to consist of N-acetylglucosamine (NAG) [39], we further examined whether the RNA-Seq library contained loci involved in the breakdown of this carbohydrate. While our transcriptome revealed the expression of a complete pathway for both the import and metabolism of NAG (KEGG, map00520 and map00550), we chose to validate *A. veronii* expression of this pathway by further examining three loci (i.e. *glmM* (phosphoglucosamine mutase), *murA* (UDP-N-acetylglucosamine 1-carboxyvinyltransferase) and *murB* (UDP-N-acetylmurate dehydrogenase)) under varying NAG concentrations. Primers

were generated and verified for species-specificity using Primer BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and through PCR under the settings in Table 1.

An artificial *in vitro* culture assay was used as a model for analyzing the expression within NAG-enriched environments, such as leech mucus. As LB-broth contains N-acetylglucosamine, *A. veronii* was cultured in minimal media made with M9 salts [40] in which the sole carbon source was glucose. Overnight cultures were generated with the GFP-expressing *A. veronii* strain, HM21S::Tn7gfp + Streptomycin (Str; 100 µg/mL) + Kanamycin (Km; 100 µg/mL) and grown at 30°C.

To ensure that living cells were not cannibalizing dead cell debris and exclusively salvaging the NAG remnants for incorporation into peptidoglycan, live cells were selected using a BD FACSAria (BD Biosciences, San Jose CA) at the WVU Flow Cytometry Core Facility. First, overnight cultures were incubated with propidium iodide, which only stains dead cells, then separated from cells with GFP-expression and lacking propidium iodide staining. These aliquots were run a second time through the flow cytometer to ensure samples were clear of dead cell debris. Subsequently, cultures consisting of ~5,000,000 live cells were allowed to recover and grow at 30°C, shaking at 225 rpm, in minimal media + Str + Km containing no NAG (S-9002; Vector Labs, Burlingame, CA), a low concentration of NAG (90 nmol/mL), or a high concentration of NAG (360 nmol/mL) [41] for 42 h, which signified entry into mid-log phase, verified through OD_{600} nm readings. Cells were then centrifuged at 4,000 *xg* for 5 min at 4°C and re-suspended in RNA*later* (Life Technologies, Carlsbad, CA) and placed at -20°C until further processing.

RNA was isolated using the TRIzol protocol (Invitrogen, Carlsbad, CA) and verified free of DNA contamination through PCR amplification with an RNA only template. First-strand cDNA synthesis was performed with 500 ng RNA, a 2 mM primer cocktail of AVgfp199-R,

AVglmM-R, AVMmurA-R, and AVmurB-R (Table 1) and Superscript II Reverse Transcriptase (Life Technologies). Second-strand synthesis was then performed with 2 uL cDNA template and complementary 5'-end gene primers with the settings described in Table 1. Amplicons were then analyzed by agarose gel electrophoresis and visualized with Kodak one-dimensional image analysis software. The expression level of *A. veronii* gfp199 was used as a loading control.

RESULTS

A. veronii undergo transcriptional tuning within different lifestyles.

To determine changes in *A. veronii* activity within leech mucosal secretions compared with other environments associated with different lifestyles, we analyzed a RNA-seq library of a pooled 3 d old mucus sample (n=3). This time point corresponds with a peak in *A. veronii* density within mucus following host shedding [18]. In total, approximately 9.3 million paired-end reads were generated, with the average length of each read being 208 ± 40 bp (Table 2). Overall read quality was high for the sample set based on FASTQC analysis. Using the Trinity software package [30], approximately 8 million reads were assembled *de novo* into 128,964 contigs. Single-end read *A. veronii* libraries generated within the crop and culture environments [26] were processed using the same pipeline and used for comparative analyses with the mucus sequences.

The contigs were then mapped to the *A. veronii* HM21 reference genome sequence [31] using Bowtie 2 [33]. A total of 626 contigs (constituting 301,818 reads), averaging 1355 ± 54 bp, matched *A. veronii* sequences within the mucus. Similarly, 200 contigs generated for the leech crop and 960 contigs for culture were mapped as stated above, with average sizes of 406 ± 23 bp and 420 ± 17 bp, respectively. To validate that our assembly matched previously reported results [26] regarding gut and monoculture environments, we searched for and found all genes reported in the study referenced, indicating consistency across our analyses. Further, the Expression

Values (EVs) of canonical housekeeping loci (e.g. *gyrB*, *rpoD*, and *tufB*) were similar across the three libraries further validating RNA-seq normalization.

The *A. veronii*-specific contigs were then functionally annotated by BLAST (using the NCBI non-redundant database) and gene loci were identified within each contig. If a gene was transcribed in multiple environments, it tended to be more highly expressed in the mucosal environment (Figs. 1A and B). Specifically, 70 and 18 loci were more highly expressed within exuded mucus relative to the monoculture and gut, respectively. In comparison, 18 and 3 loci were more highly expressed within culture (Fig. 1A) and gut (Fig. 1B) relative to mucus. All *A. veronii* specific contigs were then analyzed using Blast2GO [37], which determined functional classification and assigned gene ontology (GO) terms to each contig. By qualitatively examining the resulting GO terms, we were able to concluding that the same gene sets (i.e. GO terms) were observed across all three environments, which supports the validity of comparing transcriptomic datasets obtained from different studies [42]. To further determine which GO terms were significantly enriched within the three environments, we used a Fisher's Exact Test, with a false discovery rate (FDR) correction. All *A. veronii* expressed genes that were enriched in at least one environment fell into 1 of 31 general GO terms (Fig. 1C). We focused on those significantly enriched within the mucosal environment (i.e. 12 out 31) (Fig. 1C, indicated with red boxes). While many GO terms describe general biological processes (e.g. cellular activity, cell binding and transcription), an enrichment of reads corresponding to genes involved in oxidoreductase activity, iron/sulfur binding, hydrolase activity, heat shock response, amino acid metabolism and biosynthesis, and membrane transport, are indicative of the environmental resources and pressures encountered by *A. veronii* within mucus.

Mucosal secretions represent a stressful environment.

To ensure evolutionary persistence, microbes must be able to adapt to environmental changes. As a previous study compared differences in *A. veronii* gene expression between the leech host crop and within monoculture [26], we used these data sets to identify loci that may be necessary for the transition and survival of *A. veronii* during horizontal transmission, specifically shifting from localization in the host digestive tract to exuded mucus [18]. To elucidate the impact of lifestyle shifts on *A. veronii* transcriptional activity, the EVs of transcribed genes within all three environments were determined. The *A. veronii* genes that were solely expressed within a given environment (gut=53; culture=300; mucus=105) were identified and categorized based on KEGG Class (Fig. 2A). First, using a 1-way ANOVA, we determined that there was a higher likelihood of finding a uniquely expressed (i.e. solely expressed within a given environment) gene in the 1:10 EV bin than any other bin (p=0.0007), primarily driven by a high diversity of genes at low EVs within monoculture. Additionally, the mean EV was significantly (p<0.0001) higher in the gut for uniquely expressed metabolism-related genes, followed by the mucus. For Environmental information processing (EIP), the mean expression level of uniquely expressed loci was also significantly (p=0.0002) higher in the gut, with mucus and culture not differing from each other. Analyses regarding the Genetic information processing (GIP) and Other (p<0.0001) categories followed a similar pattern as EIP. However, no uniquely expressed *A. veronii* loci involved in Cellular processes (CP) were identified within the gut environment. As such, a Student's *t*-test determined that the mean expression of CP-related genes that were uniquely expressed was significantly (p<0.0001) higher in the mucus than in the culture. Finally, when we examine if there are significant differences in mean EV values between KEGG classes for uniquely expressed genes within specific environments, we observed no significant differences in the gut (1-way ANOVA, p=0.0591), however monoculture had significantly (1-way ANOVA, p=0.0352) higher expression in GIP compared with the other KEGG classes.

Interestingly, there was significantly (1-way ANOVA, p=0.0170) higher mean expression in CP-related genes. Upon further investigation, we observed that this significance is driven by a single gene, catalase (*katE*), within the mucosal environment.

While identifying genes exclusively expressed within each environment elucidates what processes may be unique to that habitat, this alone does not provide a comprehensive understanding of the lifestyle shift experienced by *A. veronii*. Therefore, we continued with analyses that incorporated all genes expressed within each environment. As we were mainly interested in genes within the four main KEGG classes, we compared the EVs within these classes to see which groups contained the highest expression within given environments. In every environment, GIP contained the highest levels of expression (Fig. 2C). The majority of GIP genes overexpressed (i.e. genes that had an EV > 1,000) within the culture and gut environments were associated with housekeeping functions such as transcription and translation (e.g. large subunit ribosomal proteins and RNA polymerases). However, within mucus, a number of GIP genes associated with stress responses were overexpressed (Fig. 2D; Table 3). Additionally, an overexpressed stress-related gene, glutathione peroxidase (Fig. 2E; Table 3), binning to metabolism was also identified within mucus. Interestingly, upregulation of glutathione peroxidase indicates that *A. veronii* is challenged by oxidative stress in the mucus, as this enzyme uses lipids to reduce free H_2O_2 into H_2O [43]. In support of the oxidative stress encountered by *A. veronii* within mucus, cysteine desulfurase (Fig. 2E) was also overexpressed. Cysteine desulfurase catalyzes the formation of sulfur and alanine from cysteine [44]. The sulfur from this reaction can then form Fe-S clusters [44], which are frequently used in oxidative stress response.

While glutathione peroxidase and cysteine desulfurase are associated with oxidative stress, genes associated with other forms of stress, including temperature and osmotic stress,

were also overexpressed. For example, the ATP-dependent helicases DeaD and RhlE (Fig. 2D-E) are known as DEAD-box RNA helicases, and are involved in various cell processes at lower temperatures, such as ribosome biogenesis [45], RNA degradation [45] and translation initiation (reviewed in [46]). Additionally, the chaperone DnaK assists cells with protein folding after a heat shock event [47], indicating that this is associated with thermotolerance, while the chaperone GroEL is useful for the cell to adapt to osmotic stressors and protein misfolds [47].

Beyond managing the multiple levels of stress, *A. veronii* must also be able to use nutrients within the host-secreted mucus for survival and proliferation. The overexpression of the isocitrate lyase, *aceA* (Fig. 2E), hints at the presence of glycans in the mucosal environment [48]. Isocitrate lyase is a key enzyme in the conversion of isocitrate to succinate, a critical component of the TCA cycle, and glyoxylate, a substrate involved in numerous metabolic pathways including the production of malate. The expression of malate dehydrogenase also confirms the presence of malate and its metabolism into oxaloacetate, the latter of which can feed back into the TCA cycle, or a number of other pathways. Interestingly, isocitrate lyase is also active within the leech gut [48] and indicates not only a similarity in available nutrients but also reveals a cross-over of *A. veronii* genetic function and some redundancy between the gut and mucosal environments.

Mucus consists of N-acetylglucosamine that may enable A. veronii proliferation.

In addition to investigating activity related to lifestyle adaptation, we also wanted to determine what nutrient sources enable *A. veronii* proliferation within the stressful environment of mucus. As shed mucus is thought to consist of glucosaminoglycans [39], we searched for genes involved in the metabolism of this substance within the *A. veronii* mucosal RNA-Seq library. The mucosal transcriptome contained reads corresponding to loci involved in both the import of extracellular N-acetylglucosamine (NAG; a major component of glycosaminoglycans)

(Fig. 3A) and processing of NAG towards the generation of N-acetylmuramic acid (MurNAc). NAG and MurNAc are significant components of the glycan backbone of peptidoglycan (i.e. the bacterial cell wall), which is necessary for cellular proliferation. As such, we hypothesized that *A. veronii* scavenges NAG for a nutrient source from the surrounding mucosal environment for use in peptidoglycan synthesis.

In order for *A. veronii* to scavenge environmental NAG, we first confirmed that mucus contains this monosaccharide. Mucosal secretions were examined for the presence of NAG using a Succinylated-wheat germ agglutinin (WGA-S) probe ([49]; Fig. 3B). This staining demonstrates that mucus consists of an abundant amount of NAG, and can be used to visualize the structure of exuded mucus (Fig. 3Bi). To determine that *A. veronii* is within the vicinity of NAG for import, a GFP-expressing *A. veronii* strain was orally administered to leeches to ensure natural shedding into the mucus [18] and immunocytochemistry was used to visualize the GFP. Bacterial rods can be seen throughout the mucus (Fig. 3Bii) with small clusters of GFP-fluorescent rods (Fig. 3Biii; Movie S1), consistent with the low density of *A. veronii* within the mucosal microbiota [19].

To determine if *A. veronii* will alter its transcriptional profile in response to nutrient availability, key genes involved in the import and processing of NAG (Fig. 3A) were examined using semi-quantitative RT-PCR within an '*in vitro*' culture system as a model for environments containing varying concentrations of NAG. To control for glucose as the sole carbon source, *A. veronii* cultures were grown in minimal media [40] and to reduce the likelihood that *A. veronii* were acquiring NAG from cell wall debris by cannibalizing dead cells, cultures were incubated with propidium iodide (which stains dead cells) through flow cytometry to isolate live cells (Fig. 3C, columns designated with "P"). The recovery media was then supplemented with low or high

levels of NAG. An additional set of cultures were supplemented with similar NAG concentrations, however these cultures did not undergo flow cytometry (Fig. 3C, "N").

Figure 3C shows that *glmM*, *murA*, and *murB* had increased expression levels when incubated with increasing concentrations of NAG. Interestingly, control samples (i.e. no NAG) that were not subjected to flow cytometry to remove dead cell debris had a high level of expression for all genes examined relative to their flow cytometry counterparts, suggesting cannibalistic activity (Fig. 3C). This could mean that *A. veronii* can use both the surrounding mucosal environment and the cellular remnants of other microbial community members as a source of NAG. These results indicate that in a NAG-enriched environment, these genes are turned on to process NAG into MurNAc, likely making peptidoglycan.

DISCUSSION

In this study, we identify the transcriptional responses that enable the transition between different microbial lifestyles, in this case from a mutualistic to a free-living state. Additionally, we demonstrate that this symbiont is capable of not only surviving in a stressful environment, but also proliferating to densities high enough to use this "in between" state as a mechanism for transmission, likely by using host exuded mucus for nutritional resources.

The leech mucosal transcriptome consists of 129,964 contigs, 626 of which were identified as *A. veronii*-specific. The low proportion of *A. veronii* transcripts within the total transcriptome (i.e. ~1% of total size) corresponds with the relatively low density of *A. veronii* within shed mucus, as characterized by previous Illumina deep sequencing of the microbiota composition [19]. Comparisons with previous studies of *A. veronii* in other distinct habitats [26] enabled our understanding of how this microbe adapts to a major lifestyle shift, i.e. from a digestive tract symbiont to being shed into exuded mucus. Bacteria cope with environment shifts via several mechanisms, including lateral gene transfer (LGT) and elevated mutation rates

coupled with high effective population sizes [50, 51], however less is known concerning adaptive transcriptional tuning.

The expression of loci both unique to *A. veronii* living in the mucus and shared among the gut and/or in monoculture provides a summary of the activities and conditions within each environment. The majority of genes found to be expressed in all three environments (Fig. 1A-B) are more highly expressed within the mucosal environment. This indicates that, while certain functions are redundant with the culture and gut, these roles are amplified within mucus likely due to alterations in nutrient resources and increased competition from the diverse surrounding microbiota [19]. Interestingly, *A. veronii* in culture express a higher quantity of unique genes at lower EVs. This suggests that *A. veronii,* under low stress (i.e. rich monoculture), can afford to exert energy on auxiliary gene products when residing in an ideal environment, such as adenylosuccinate synthase, haloacid dehydrogenase, and chorismate synthase. In contrast, these products may be available to mucosal-bound *A. veronii* from the surrounding microbial community in a form of nutrient cross-feeding (reviewed in [52]) or to gut-bound *A. veronii* from the host [48] or the secondary leech gut symbiont, *M. hirudinis* [14, 48].

Conversely, while genes uniquely expressed within the gut environment are low in number, the EVs of these genes tend to be large enough to prove significantly higher in every KEGG category in comparison to other environments (Fig. 2A), with the exception of cellular processes. While the majority of these genes are involved in housekeeping functions (e.g. ribosomal proteins), genes related to phage-shock were also found to be highly expressed, confirming results found in the original gut transcriptome study [26]. Phage-shock genes are important to the gut environment as a response to osmotic stress, high temperatures, stationary growth phase and filamentous phage infection [53]. As phage-shock genes were not expressed

within the mucosal environment, we can conclude that *A. veronii* does not require such a response, and likely does not encounter some of these stressors.

While residing in mucus, *A. veronii* does not spend energy on a number of costly syntheses, as in culture, or use phage-shock gene products to respond to its surrounding environments, as within the leech gut. Instead, *A. veronii* shifts its energy balance in favor of membrane transport (Fig. 1C), likely indicating that the surrounding microbial community and/or composition of the mucus provides necessary metabolites. Additionally, a significant increase in expression related to cellular processes, mainly driven by catalase activity, as well as the enrichment of genes involved in iron/sulfur binding and responses to temperature fluctuations, denotes a stressful environment. Generally, microbes respond to stress either by acclimating (e.g. transferring limited resources from growth to survival pathways [54]), becoming dormant (reviewed in [55]), or dying. In mucus-bound *A. veronii*, many over-expressed genes (i.e. a corresponding EV > 1,000) are related to stress-response processes, indicating that *A. veronii* adapts to the environmental transition by shifting resources, at least partly, to survival pathways.

While mucosal secretions of aquatic animals are known to harbor high levels of antimicrobial peptides [27, 28], the overexpression of genes related to the processing of these compounds was not observed in the *A. veronii* mucosal transcriptome. However, fluctuating temperature is a stressful condition encountered by *A. veronii* during transition from the host to the mucus. A number of genes enriched and over-expressed in the mucosal environment were related to heat-shock, such as *dnaK* [47] (Fig. 1C, Fig. 2D-E, Table 3), as well as function within colder environments, such as *deaD* and *rhlE* [45, 56](reviewed in[46]) (Fig. 2D-E, Table 3). This indicates that *A. veronii* is experiencing fluctuations in temperature not observed in the gut or monoculture environments, a reasonable conclusion as the monoculture was maintained at a steady 30°C [26]. Interestingly, this also indicates that the leech gut environment is more stable

when the host is maintained at room temperature [26] than free-floating mucus maintained at 15°C (reflective of a pond's base within the geographic range of the *H. verbena* host).

Additionally, a shift in osmolarity within the mucosal environment, relative to the gut or culture, is supported by a higher expression of genes related to osmotic stress (e.g. *groEL* [47]; Fig. 2D-E, Table 3) within mucus. The chaperonin *groEL* is also well known for its increased expression in the obligate symbionts of many invertebrates, such as aphids [57], tsetse [58] and weevils [59]. This high expression compensates for lower protein stability resulting from an AT-bias typically selected for in many endosymbionts [60]. However, as *A. veronii* is able to maintain a larger genomic inventory, similar to that of a generalist, and lacks an AT-biased genome [31], the overexpression of *groEL* is likely due to a variety of other stressors that may impact protein folding in the mucosal environment.

Another abiotic factor affecting *A. veronii* within the mucus is oxidative stress, as indicated by a significant bias toward a single, over-expressed gene, catalase (Fig. 2E), consistent with a high abundance of H_2O_2 [61]. Catalase may protect against the oxidative damage produced by reactive oxygen species (ROS), driving the increased need for chaperones in the mucosal environment. Interestingly, catalase was also observed within the coral mucus metagenome [62], and is known to provide *Vibrio fischeri*, a mutualist of sepiolid squids, protection from ROS within host-secreted mucus (reviewed in [63]), indicating that mucosal secretions of other aquatic organisms are also characterized with high levels of oxidative stress. Additionally, the enrichment of genes related to iron/sulfur dependent enzymes (Fig. 1C) by *A. veronii* in leech mucus, supported by the overexpression of cysteine desulfurase (Fig. 2D-E, Table 3) indicates the catabolism of sulfur and alanine from cysteine [44], followed by the formation of Fe-S clusters [44] which can respond to oxidative stress. Finally, the overexpression

of glutathione peroxidase (Fig. 2D-E, Table 3), which uses lipids to reduce hydrogen peroxide to water [43], further supports the presence of oxidative stress within the mucosal environment.

Even within the presence of oxidative and temperature pressures, *A. veronii* is capable of survival and proliferation [18], indicating that there are sufficient nutrients within the mucus for growth. While the enrichment of amino acid metabolism and biosynthesis genes in the mucus supports protein synthesis, we decided to focus on carbon sources in the mucus, as heterotrophic bacteria rely on external sources of organic carbon for growth. The overexpression of isocitrate lyase *aceA* transcripts (Fig. 2E) suggests not only a source of nutrition, but also an overlap of metabolic function between the gut and mucus [48]. As expression of isocitrate lyase within the gut suggested that *A. veronii* was using acetate or other fatty acids as carbon and energy sources [48], it is likely that similar compounds are also available in the mucus. Interestingly, the acetate and short chain fatty acids in the gut were generated by *Mucinivorans hirudinis* from host mucin glycans [48]. This indicates that *M. hirudinis*, which also survives and proliferates within mucus [19], may be playing a parallel role in both environments. Alternatively, one of the many additional mucosal microbiota members may provide this service. In either case, these results indicate a redundancy that exists between the gut and mucosal environments.

While little is known regarding the available nutrients within mucus, we determined that these castings contain the glucosaminoglycan, N-acetylglucosamine (NAG; Fig. 3B), which is a major component of peptidoglycan [64]. The *A. veronii* mucosal transcriptome revealed a complete pathway involved in the import of extracellular NAG and its subsequent incorporation into the production of N-acetylmuramic acid (MurNAc) (Fig. 3A). Consistent with this conclusion, tRNA-synthetases (i.e. alanyl, lysyl, and glutamyl) for the amino acids that compose the peptide fraction of peptidoglycan were identified within the *A. veronii* mucosal transcriptome. Additionally, *A. veronii* form small clusters within the vicinity of NAG (Fig. 3B).

The small numbers of fluorescent cells confirm the low relative abundance of *A. veronii* in the mucosal microbiota [19], yet also highlights the efficiency of mucus as a transmission platform [18]. We also found that *A. veronii* increase their expression of several peptidoglycan synthesis genes in responses to high NAG concentrations (Fig. 3A, C) confirming that *A. veronii* may alter nutritional consumption in response to changing environmental conditions. When the expression of NAG-related genes within *in vitro* culture assays was examined, the ability of *A. veronii* to cannibalize dead cells and use the NAG in their cell walls was also observed (Fig. 3C). This is interesting, as in other NAG-limited environments, *A. veronii* may be capable of scavenging other dead community members to promote its proliferation.

In conclusion, these RNA-Seq analyses indicate that transmission substrates, such as mucosal castings which underlie transmission of many host-associated microbiotas [65, 66](reviewed in[67]), expose symbionts to a variety of novel stressors that impact the evolution of the microbiota. Most organisms experience different selective pressures during their lifespan, with horizontally transmitted microbes, in particular, experiencing dramatic shifts in their exposure to a wide variety of biotic and abiotic factors. Our results demonstrate that as symbionts shift between lifestyles, they undergo widespread alterations in gene expression, which we term “transcriptional tuning”. In particular, stress-related genes are upregulated for the shift from a sheltered existence within a host to an environmentally exposed transmission substrate, addressing ecological pressures while permitting the acquisition of nutrients. These results provide insight into how mucosal-derived bacterial symbionts of other aquatic animals, such as corals and the Hawaiian bobtail squid, may be acquired. The widespread nature of our results emphasizes the significance of lifestyle shifts within transmission settings towards shaping microbial evolution.

REFERENCES

1. Douglas AE. *Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera*. Annu. Rev. Entomol., 1998. **43**:p. 17–37.
2. Graf J. *The effect of symbionts on the physiology of Hirudo medicinalis, the medicinal leech*. Invert. Reprod. Dev., 2002. **41**:p. 269–75.
3. Moreira LA et al. *A Wolbachia symbiont in Aedes aegypti limits infection with dengue, Chikungunya, and Plasmodium*. Cell, 2009. **139**:p. 1268–78.
4. Oliver KM et al. *Facultative bacterial symbionts in aphids confer resistance to parasitic wasps*. Proc. Natl. Acad. Sci. U.S.A., 2003. **100**:p. 1803–7.
5. O'Hara AM, Shanahan F. *The gut flora as a forgotten organ*. EMBO Rep., 2006. **7**:p. 688–93.
6. Dawkins R. *The Extended Phenotype* 1982. Oxford University Press, Oxford, United Kingdom.
7. Flint HJ et al. *Microbial degradation of complex carbohydrates in the gut*. Gut Microbes, 2014. **3**:p. 289–306.
8. Hansen AK, Moran NA. *The impact of microbial symbionts on host plant utilization by herbivorous insects*. Mol. Ecol., 2014. **23**:p. 1473–96.
9. Werren JH. *Symbionts provide pesticide detoxification*. Proc. Natl. Acad. Sci. U.S.A., 2012. **109**:p. 8364–5.
10. Haine ER. *Symbiont-mediated protection*. Proc. Biol. Sci., 2008. **275**:p. 353–61.
11. Dunbar HE et al. *Aphid thermal tolerance is governed by a point mutation in bacterial symbionts*. PLoS Biol., 2007. **5**:p. e96.
12. Ebert D. *The Epidemiology and Evolution of Symbionts with Mixed-Mode Transmission*. Annu. Rev. Ecol. Evol. Syst., 2013. **44**:p. 623–43.

13. Worthen PL et al. *Culture-independent characterization of the digestive-tract microbiota of the medicinal leech reveals a tripartite symbiosis*. Appl. Env. Microbiol., 2006. **72**:p. 4775–81.

14. Nelson M et al. *Mucinivorans hirudinis gen. nov., sp. nov., an anaerobic, mucin-degrading bacterium isolated from the digestive tract of the medicinal leech, Hirudo verbana.* Int. J. Syst. Evol. Microbiol., 2015.

15. Sawyer RT. *Leech Biology and Behaviour* 1986. Clarendon Press, Oxford, United Kingdom.

16. Siddall ME et al. *Bacterial symbiont and salivary peptide evolution in the context of leech phylogeny*. Parasitology, 2011. **138**:p. 1815–27.

17. Rio R V et al. *Symbiont succession during embryonic development of the European medicinal leech, Hirudo verbana*. Appl. Env. Microbiol., 2009. **75**:p. 6890–5.

18. Ott BM et al. *Hitchhiking of host biology by beneficial symbionts enhances transmission*. Sci. Rep., 2014. **4**:p. 5825.

19. Ott BM et al. *Characterization of shed medicinal leech mucus reveals a diverse microbiota*. Front. Microbiol., 2015. **5**.

20. Janda JM, Abbott SL. *The genus Aeromonas: taxonomy, pathogenicity, and infection*. Clin. Microbiol. Rev., 2010. **23**:p. 35–73.

21. Figueras MJ et al. *Aeromonas rivuli sp. nov., isolated from the upstream region of a karst water rivulet*. Int. J. Syst. Evol. Microbiol., 2011. **61**:p. 242–8.

22. Janda JM, Abbott SL. *Evolving concepts regarding the genus Aeromonas: an expanding Panorama of species, disease presentations, and unanswered questions*. Clin. Infect. Dis., 1998. **27**:p. 332–44.

23. Barbier M et al. *From the Environment to the Host: Re-Wiring of the Transcriptome of*

Pseudomonas aeruginosa from 22°C to 37°C. PLoS One, 2014. **9**:p. e89941.

24. López-Maury L et al. *Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation*. Nat. Rev. Genet., 2008. **9**:p. 583–93.
25. Bjedov I et al. *Stress-induced mutagenesis in bacteria.*. Science, 2003. **300**:p. 1404–9.
26. Bomar L, Graf J. *Investigation into the Physiologies of Aeromonas veronii in vitro and Inside the Digestive Tract of the Medicinal Leech Using RNA-seq*. Biol. Bull., 2012. **223**:p. 155–66.
27. Subramanian S et al. *Comparison of antimicrobial activity in the epidermal mucus extracts of fish*. Comp. Biochem. Physiol. B. Biochem. Mol. Biol., 2008. **150**:p. 85–92.
28. Shnit-Orland M, Kushmaro A. *Coral mucus-associated bacteria: a possible first line of defense*. FEMS Microbiol. Ecol., 2009. **67**:p. 371–80.
29. 't Hoen PAC et al. *Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories*. Nat. Biotechnol., 2013. **31**:p. 1015–22.
30. Grabherr MG et al. *Full-length transcriptome assembly from RNA-Seq data without a reference genome*. Nat. Biotechnol., 2011. **29**:p. 644–52.
31. Bomar L et al. *Draft Genome Sequence of Aeromonas veronii Hm21, a Symbiotic Isolate from the Medicinal Leech Digestive Tract*. Genome Announc., 2013. **1**:p. e00800–13.
32. Camacho C et al. *BLAST+: architecture and applications*. BMC Bioinformatics, 2009. **10**:p. 421.
33. Langmead B, Salzberg SL. *Fast gapped-read alignment with Bowtie 2*. Nat. Methods, 2012. **9**:p. 357–9.
34. Huson DH et al. *Integrative analysis of environmental sequences using MEGAN4*. Genome Res., 2011. **21**:p. 1552–60.
35. Li P et al. *Comparing the normalization methods for the differential analysis of Illumina*

high-throughput RNA-Seq data. BMC Bioinformatics, 2015. **16**:p. 347.

36. Dillies M-A et al. *A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis*. Brief Bioinform., 2013. **14**:p. 671–83.

37. Conesa A et al. *Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.*. Bioinformatics, 2005. **21**:p. 3674–6.

38. SAS Institute Inc. *JMP*. 10. Cary, N.C.

39. Michalsen A et al. *Medicinal Leech Therapy* 2007. Thieme Medical Publishers, New York, New York.

40. Sambrook J, Russell DW. *Molecular Cloning* 2001. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

41. Kawakubo M et al. *Natural antibiotic function of a human gastric mucin against Helicobacter pylori infection*. Science, 2004. **305**:p. 1003–6.

42. Suárez-Fariñas M et al. *Evaluation of the Psoriasis Transcriptome across Different Studies by Gene Set Enrichment Analysis (GSEA)*. PLoS One, 2010. **5**:p. e10247.

43. Gaber A et al. *NADPH-dependent glutathione peroxidase-like proteins (Gpx-1, Gpx-2) reduce unsaturated fatty acid hydroperoxides in Synechocystis PCC 6803*. FEBS Lett., 2001. **499**:p. 32–6.

44. Flint DH. *Escherichia coli contains a protein that is homologous in function and N-terminal sequence to the protein encoded by the nifS gene of Azotobacter vinelandii and that can participate in the synthesis of the Fe-S cluster of dihydroxy-acid dehydratase*. J. Biol. Chem., 1996. **271**:p. 16068–74.

45. Lehnik-Habrink M et al. *DEAD-Box RNA helicases in Bacillus subtilis have multiple functions and act independently from each other*. J. Bacteriol., 2013. **195**:p. 534–44.

46. Redder P et al. *Bacterial versatility requires DEAD-box RNA helicases*. FEMS Microbiol.

Rev., 2015. **39**:p. 392–412.

47. Susin MF et al. *GroES/GroEL and DnaK/DnaJ have distinct roles in stress responses and during cell cycle progression in Caulobacter crescentus*. J. Bacteriol., 2006. **188**:p. 8044–53.

48. Bomar L et al. *Directed culturing of microorganisms using metatranscriptomics*. MBio, 2011. **2**:p. e00012–11.

49. 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. Env. Microbiol., 2007. **73**:p. 1984–91.

50. Delaux P-M et al. *Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution*. PLoS Genet., 2014. **10**:p. e1004487.

51. Gómez-Valero L et al. *The evolutionary fate of nonfunctional DNA in the bacterial endosymbiont Buchnera aphidicola*. Mol. Biol. Evol., 2004. **21**:p. 2172–81.

52. Seth EC, Taga ME. *Nutrient cross-feeding in the microbial world*. Front. Microbiol., 2014. **5**:p. 350.

53. Rowley G et al. *Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens*. Nat. Rev. Microbiol., 2006. **4**:p. 383–94.

54. Schimel J et al. *Microbial Stress-Response Physiology and its Implications for Ecosystem Function*. Ecology, 2007. **88**:p. 1386–94.

55. Wood TK et al. *Bacterial persister cell formation and dormancy*. Appl. Environ. Microbiol., 2013. **79**:p. 7116–21.

56. Iost I, Dreyfus M. *DEAD-box RNA helicases in Escherichia coli*. Nucleic. Acids Res., 2006. **34**:p. 4189–97.

57. Fares MA et al. *The Evolution of the Heat-Shock Protein GroEL from Buchnera, the*

Primary Endosymbiont of Aphids, Is Governed by Positive Selection. Mol. Biol. Evol., 2002. **19**:p. 1162–70.

58. Aksoy S. *Molecular analysis of the endosymbionts of tsetse flies: 16S rDNA locus and over-expression of a chaperonin*. Insect Mol. Biol., 1995. **4**:p. 23–9.

59. Charles H et al. *A molecular aspect of symbiotic interactions between the weevil Sitophilus oryzae and its endosymbiotic bacteria: over-expression of a chaperonin*. Biochem. Biophys. Res. Commun., 1997. **239**:p. 769–74.

60. Moran NA. *Accelerated evolution and Muller's rachet in endosymbiotic bacteria*. Proc. Natl. Acad. Sci. U.S.A., 1996. **93**:p. 2873–8.

61. Barnes AC et al. *Peroxide-inducible catalase in Aeromonas salmonicida subsp. salmonicida protects against exogenous hydrogen peroxide and killing by activated rainbow trout, Oncorhynchus mykiss L., macrophages*. Microb. Pathog., 1999. **26**:p. 149–58.

62. Vega Thurber R et al. *Metagenomic analysis of stressed coral holobionts*. Environ. Microbiol., 2009. **11**:p. 2148–63.

63. Nyholm S V, Nishiguchi MK. *The Evolutionary Ecology of a Sepiolid Squid-Vibrio Association: from Cell to Environment*. Vie Milieu Paris, 2008. **58**:p. 175–84.

64. Esko JD et al. *Proteoglycans and Sulfated Glycosaminoglycans* 2009. Cold Spring Harbor Laboratory Press.

65. Warfel JM et al. *Airborne transmission of Bordetella pertussis*. J. Infect. Dis., 2012. **206**:p. 902–6.

66. Ritchie KB. *Regulation of microbial populations by coral surface mucus and mucus-associated bacteria*. Mar. Ecol. Prog. Ser., 2006. **322**:p. 1–14.

67. Bright M, Bulgheresi S. *A complex journey: transmission of microbial symbionts*. Nat.

Rev. Microbiol., 2010. **8**:p. 218–30.

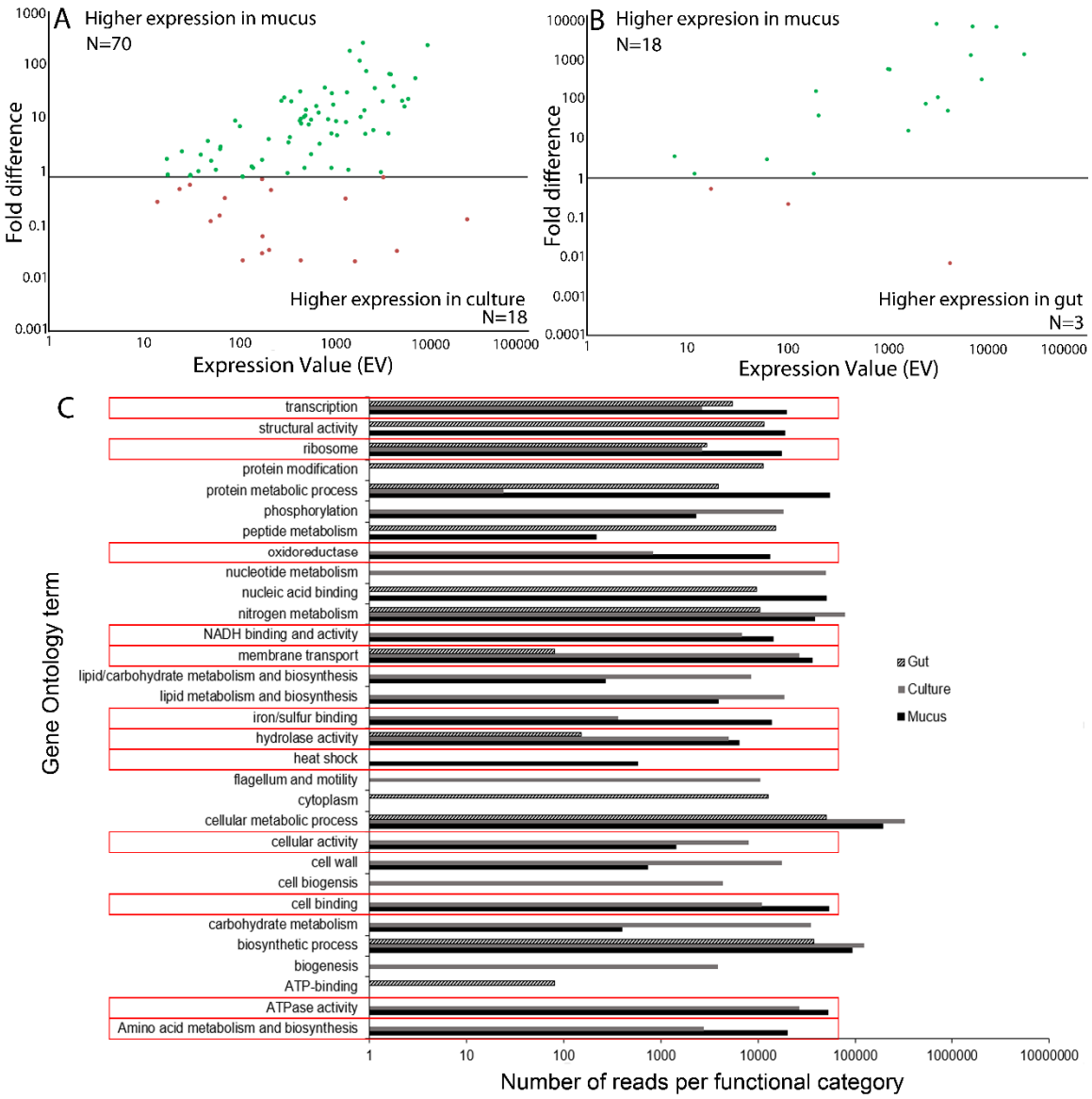

Figure 4A-1. *A. veronii* transcript expression based on RNA-seq analyses. A) Relative expression for each shared contig, based on Expression Value (EV), within mucus and culture environments. B) Relative expression for each shared contig, based on EV comparing the mucus and gut environments. Green indicates contigs with higher expression in the mucus, while red indicates contigs with higher expression in the compared environment (i.e. culture for A; gut for B). C) Gene ontology (GO) terms differentially expressed as determined by enrichment analysis using Fisher's Exact Test. GO terms with a red box indicate categories that are significantly overexpressed in the mucus compared with other environments (Fisher's Exact Test, $p<0.0031$).

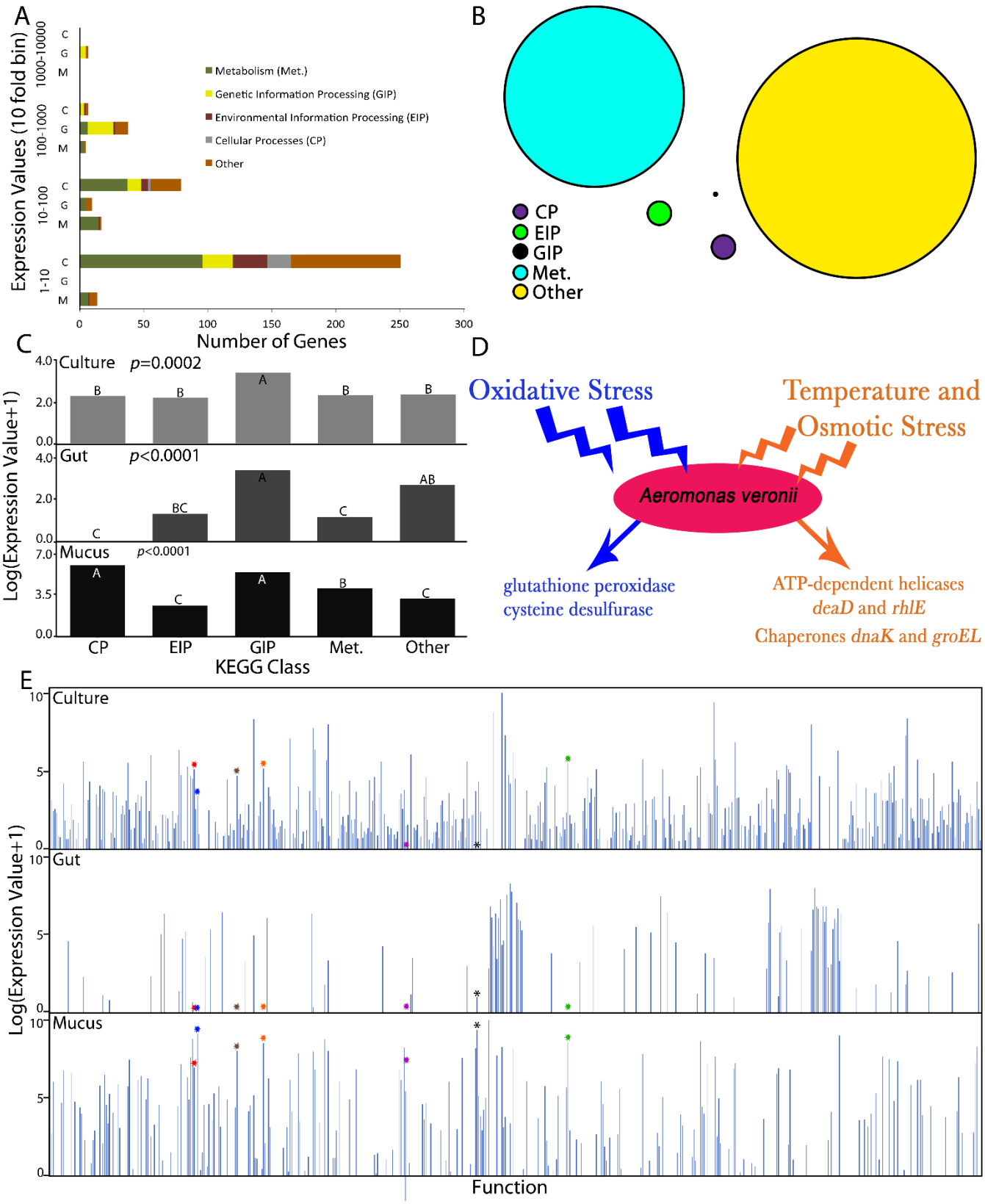

Figure 4A-2. Comparison of *A. veronii* transcriptome within three habitats. A) Proportion of *A. veronii* genes expressed exclusively within exuded mucus (M), gut (G), or monoculture (C). Genes are categorized based on expression value (EV) in 10-fold increments as well as KEGG Class. The "Other" category contains genes that are not classified by KEGG into the four main functional classes. B) *A. veronii* genes, binned into KEGG functional categories, expressed only within mucosal casts. Diameter corresponds with relative transcript abundance. C) A 1-way ANOVA was performed to test for significant differences in mean EV among categories containing all expressed genes (plotted on a Log(EV+1) scale) within each environment. Letters indicate statistically significant differences in expression between KEGG categories. D) Mucosal secretions represent a stressful environment. Genes overexpressed within mucus (EV >1000) indicate the presence of oxidative, temperature and osmotic stressors. E) Histogram showing the distribution of Log(EV+1) for every gene captured within the RNA-Seq libraries. Red sunburst indicates the ATP-dependent RNA helicase *deaD* transcript; blue sunburst indicates ATP-dependent RNA helicase *rhlE*; green sunburst indicates molecular chaperon *dnaK*; brown sunburst indicates the chaperonin *groEL*; orange sunburst indicates cysteine desulfurase; and purple sunburst indicates glutathione peroxidase. The black asterisk indicates isocitrate lyase and the red asterisk indicates catalase.

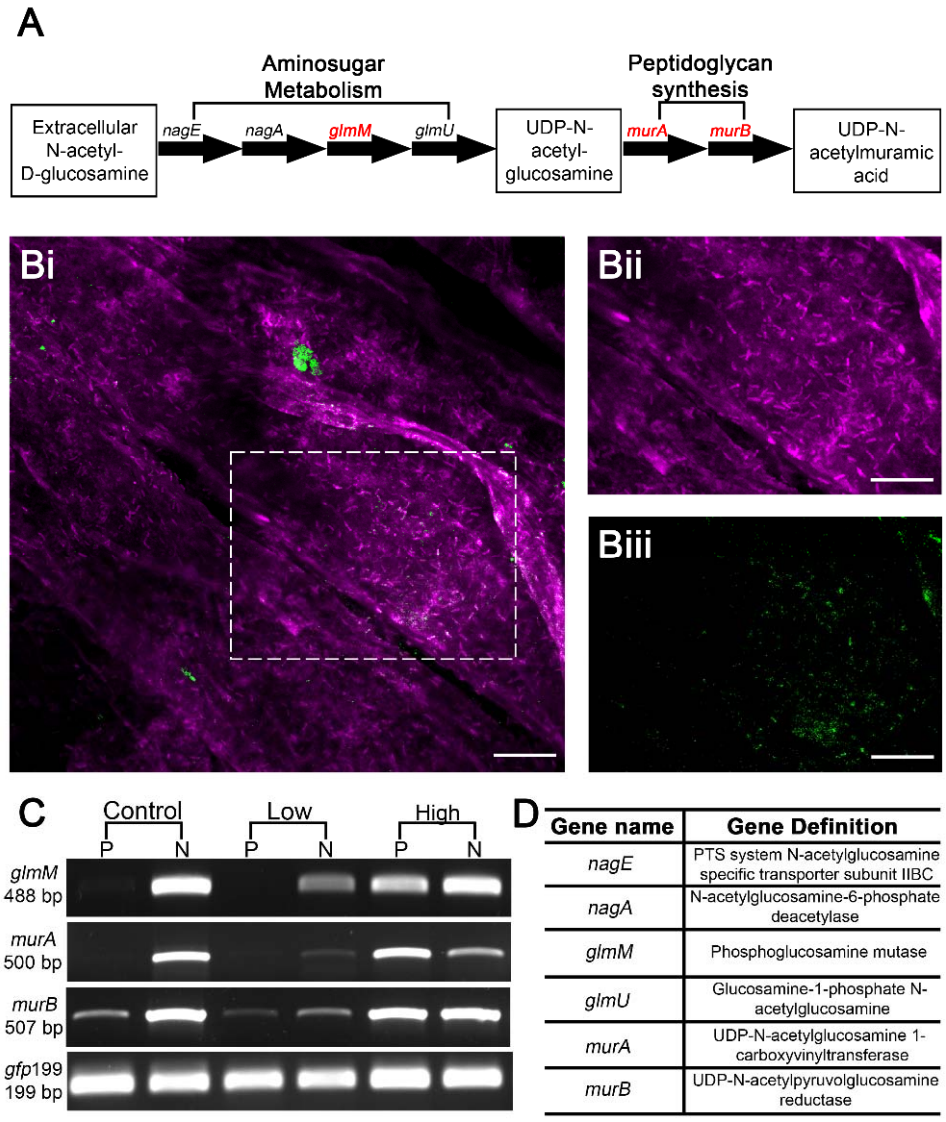

Figure 4A-3. Mucus consists of N-acetylglucosamine that may enable *A. veronii* proliferation. A) Metabolic pathway depicting the influx of NAG and its conversion to N-acetylmuramic acid, both of which compose the glycan portion of peptidoglycan. Loci depicted in red were chosen for expression analyses (refer to (C) and (D)). B) Lectin staining and immunocytochemistry images of mucosal secretions. Bi) Mucosal secretions stained using WGA-S (60X oil). Dotted box indicates a cluster of GFP-expressing *A. veronii* cells. Bii) indicates only WGA-S staining within the dotted box. Biii) depicts only the immunostaining of GFP protein within the dotted box. Arrowheads indicates examples of bacterial cells (Bii) with their corresponding GFP signal

(Biii). Scale bar depicts 10 μM. C) RT-PCR gel electrophoresis images showing the expression of key NAG-related genes in samples that did not undergo flow cytometry (N) and in samples post flow cytometry (P). Cultures were grown in three environments: control (no NAG), low (90 nmol/mL NAG), or high (360 nmol/mL NAG). D) The full names of the genes shown in (A), and analyzed in (C).

Table 4A-1. Primer list and amplification settings.

Primer	Sequence (5'-3')	Amplicon size (bp)	Amplification conditions[a]
Aeromonas veronii gfp (gfp199)		199	30 sec at 55°C; 30 sec at 72°C*
gfp199-F	GCA GAT TGG CGA CAG CAC GT		
gfp199-R	CAA TGT TGT GGC GAA TTT TG		
General eubacterial primers (*16S rRNA*)		1400	30 sec at 50°C; 1.5 min at 72°C
27F'	AGA GTT TGA TCM TGG CTC AG		
1492R'	TAC GGY TAC CTT GTT ACG ACT T		
Aeromonas veronii gyrase B (*gyrB*)		514	30 sec at 53°C; 30 sec at 72°C
AVgyrB-F	GCA GAT TGG CGA CAG CAC		
AVgyrB-R	GCA CCT TGA CGG AGA TAA CG		
Phosphoglucosamine mutase (*glmM*)		488	30 sec at 52°C; 30 sec at 72°C*
AvglmM-F	CAG ATT ATG TGG CCG GAG TT		
AvglmM-R	ATG AAT CTG GAT GGG GTG AA		
UDP-N-acetylglucosamine 1-carboxyvinyltransferase (*murA*)		500	30 sec at 52°C; 30 sec at 72°C*
AvmurA-F	TAG CCA CGG TCG ATG TGA TA		
AvmurA-R	ATC GAA ACC GGT ACC TTC CT		
UDP-N-acetylmurate dehydrogenase (*murB*)		507	30 sec at 53°C; 30 sec at 72°C*
AvmurB-F	CAG GTA CCA CCG GAT TAT TG		
AvmurB-R	AGG GGA TTA CCG CTG AAG AT		

[a]Each amplification reaction was initiated with a 3 min denaturation step at 95°C. *40 cycles of 30 sec at 95°C were used when analyzing NAG-related gene expression *in vitro* as a loading control.

Table 4A-2. Next-generation sequencing and *de novo* assembly statistics

Illumina MiSeq data (raw)	
Number of sequenced bases (Gbp)	1.064
Number of paired end reads	9,356,767
Average length of read (bp)	208 ± 40
Assembly (Mucus metatranscriptome)	
Number of reads assembled	8,140,387
Total transcriptome size (bp)	77,274,024
Number of contigs generated	128,964
Average contig size (bp)	599 ± 755
N25 contig length (bp)	1,893
N50 contig length (bp)	797
N75 contig length (bp)	382
***Aeromonas*-specific (Mucus)**	
Number of reads assembled	301,818
Total transcriptome size (bp)	848,739
Number of contigs assembled	626
Average contig size (bp)	1355 ± 54
***Aeromonas*-specific (Culture)[a]**	
Number of reads assembled	4,314,777
Total transcriptome size (bp)	402,746
Number of contigs assembled	960
Average contig size (bp)	420 ± 17
***Aeromonas*-specific (Gut)[a]**	
Number of reads assembled	3,435,397
Total transcriptome size (bp)	81,208
Number of contigs assembled	200
Average contig size (bp)	406 ± 23

[a]Raw data obtained from Bomar and Graf, 2012

Table 4A-3. Stress-related genes over-expressed in the mucus.

			Expression Value			Fold Change	
Function	**Kegg Class**	**KO**	**Mucus**	**Gut**	**Culture**	**Gut**	**Culture**
molecular chaperon DnaK	GIP	K04043	5365.789863	0	254.5522701	∞	21.07932434
chaperonin GroEL	GIP	K04077	3153.165011	0	119.1977161	∞	26.45323344
ATP-dependent RNA helicase DeaD	GIP	K05592	1055.642763	0	175.7359135	∞	6.006983672
ATP-dependent RNA helicase RhlE	GIP	K11927	9360.359259	0	31.08526087	∞	301.1188904
glutathione peroxidase	Met.	K00432	1379.377862	0	0	∞	∞
cysteine desulfurase	Met.; GIP	K04487	5039.419963	0	185.4926594	∞	27.16775951

Chapter 4B: Draft Genome Sequence of *Pedobacter* sp. Strain Hv1, an Isolate From Medicinal Leech Mucosal Castings[4]

ABSTRACT

The *Pedobacter sp.* Hv1 strain was isolated from the medicinal leech, *Hirudo verbana,* mucosal castings. These mucosal sheds have been demonstrated to play a role in horizontal symbiont transmission. Here, we report the draft 4.9 Mbp genome sequence of *Pedobacter* sp. strain Hv1.

INTRODUCTION/RESULTS

The genus *Pedobacter* was reclassified from the genus *Sphingobacterium* in 1998 [1] due to various phenotypic variations. Members of the *Pedobacter* genus have been described from water and soil samples, as well as from the microbiotas of nematodes, dung beetles and mosquitoes [2–4]. Recently, strain Hv1 was identified as the most abundant bacterium within the mucosal casts of medicinal leeches (*Hirudo verbana*) [5]. Here, we present a draft genome of *Pedobacter* sp. Hv1 strain.

The genome of *Pedobacter* Hv1 was sequenced using the Illumina MiSeq (1,880,901 paired-end reads). Raw reads were assembled using the CLC Genomics Workbench (Aahrus, Denmark), yielding 30 total contigs with lengths of >0.5 kb and an N50 of 665,494 bp. Due to low coverage, three contigs were immediately eliminated. The *oriC* was detected using Ori-Finder [6, 7] in Contig 12 (positions 65,484-65,930). An additional assembly was generated using AbySS [8] with varying kmer lengths, which were aligned with the CLC assembly using Mauve [9, 10]. However, the gaps could not be resolved. BLASTn detected 4 contigs with ~100% sequence identity and 0 gaps compared to regions in larger contigs. Subsequently,

[4] Reprinted from Ott BM, Beka L, Graf J, Rio RVM. Draft Genome Sequence of *Pedobacter* sp. Strain Hv1, and Isolate from Medicinal Leech Castings. Genome Announc., 2015. **3**(6): e01469-15. doi:10.1128/genomeA.01469-15

Tandem Repeats Finder [11] did not identify long terminal repeats typically associated with mobile elements, and the lack of obvious GC skew eliminates recent gene duplication events. Mapping reads with BWA (Burrow-Wheeler Alignment tool) [12] and IGV (Integrative Genomics Viewer) [13, 14] eliminated these contigs due to high probability of an assembly error, resulting in 23 total contigs with lengths >0.85 kb. The updated contigs totaled 4,900,269 bp with an average coverage of 78X; N50 of 666,494 bp. The G + C content is 37.2%.

Functional annotation of the assembled genome was performed with RAST 2.0 [15, 16], and manual annotation and curation, which predicted 4,421 protein encoding genes and 38 RNAs. RNAs were further classified using tRNAscan-SE-1.23 [17], which predicted 35 tRNA genes (including tRNAs for all 20 amino acids, with 4 each for leucine and arginine), and the RNAmmer 1.2 server [18], which identified one 5S, one 23S, and one 16S rRNA gene. Genes related to the utilization and biosynthesis of carbohydrates, including N-acetylglucosamine, mannose and trehalose decomposition, suggests the capability of adapting to varying metabolites. Hv1 contains a number of loci involved in osmotic and oxidative stress response, such as superoxide dismutase [Cu-Zn] precursor, cytochrome c551 peroxidase, and the universal stress protein family 4.

The Hv1 genome contains loci involved in Type I, III and IV secretions systems. Of particular interest, given the diverse microbiota from which Hv1 was originally isolated, is the presence of the conjugative transposon (e.g.various *tra* loci), associated with the Type IV secretion system and the transfer of DNA from environmental sources [19]. Interestingly, this operon is missing in the closest, fully sequenced relative, *Pedobacter heparinus* [20] (RAST, closest neighbor tool; similarity score, 542), while the more distantly related *P. saltans* [21] houses the same genetic components of the *tra* operon (with the exception of *tra*G and *tra*P).

Nucleotide sequence accession numbers. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession LLWP00000000. The version described in this paper is version LLWP01000000.

REFERENCES

1. Steyn PL et al. *Classification of heparinolytic bacteria into a new genus, Pedobacter, comprising four species: Pedobacter heparinus comb. nov., Pedobacter piscium comb. nov., Pedobacter africanus sp. nov. and Pedobacter saltans sp. nov. proposal of the family Sphingobac*. Int J Syst Bacteriol, 1998. **48 Pt 1**:p. 165–177.
2. Baquiran J-P et al. *Culture-Independent Investigation of the Microbiome Associated with the Nematode Acrobeloides maximus*. PLoS One, 2013. **8**:p. e67425.
3. Estes AM et al. *Brood ball-mediated transmission of microbiome members in the dung beetle, Onthophagus taurus (Coleoptera: Scarabaeidae).*. PLoS One, 2013. **8**:p. e79061.
4. Charan SS et al. *Comparative analysis of midgut bacterial communities of Aedes aegypti mosquito strains varying in vector competence to dengue virus.*. Parasitol Res, 2013. **112**:p. 2627–37.
5. Ott BM et al. *Characterization of shed medicinal leech mucus reveals a diverse microbiota*. Front Microbiol, 2015. **5**.
6. Gao F, Zhang C-T. *Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes.*. BMC Bioinformatics, 2008. **9**:p. 79.
7. Luo H et al. *Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes.*. Front Microbiol, 2014. **5**:p. 482.
8. Simpson JT et al. *ABySS: a parallel assembler for short read sequence data.*. Genome Res, 2009. **19**:p. 1117–23.
9. Darling ACE et al. *Mauve: multiple alignment of conserved genomic sequence with rearrangements.*. Genome Res, 2004. **14**:p. 1394–403.

10. Darling AE et al. *progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement.*. PLoS One, 2010. **5**:p. e11147.

11. Benson G. *Tandem repeats finder: a program to analyze DNA sequences.*. Nucleic Acids Res, 1999. **27**:p. 573–80.

12. Li H, Durbin R. *Fast and accurate long-read alignment with Burrows-Wheeler transform.*. Bioinformatics, 2010. **26**:p. 589–95.

13. Thorvaldsdóttir H et al. *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.*. Brief Bioinform, 2013. **14**:p. 178–92.

14. Robinson JT et al. *Integrative genomics viewer.*. Nat Biotechnol, 2011. **29**:p. 24–6.

15. Meyer F et al. *The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes.*. BMC Bioinformatics, 2008. **9**:p. 386.

16. Aziz RK et al. *The RAST Server: rapid annotations using subsystems technology.*. BMC Genomics, 2008. **9**:p. 75.

17. Lowe TM, Eddy SR. *tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence*. Nucleic Acids Res, 1997. **25**:p. 955–964.

18. Lagesen K et al. *RNAmmer: consistent and rapid annotation of ribosomal RNA genes.*. Nucleic Acids Res, 2007. **35**:p. 3100–8.

19. Cascales E, Christie PJ. *The versatile bacterial type IV secretion systems.*. Nat Rev Microbiol, 2003. **1**:p. 137–49.

20. Han C et al. *Complete genome sequence of Pedobacter heparinus type strain (HIM 762-3T)*. Stand Genomic Sci, 2009. **1**:p. 54–62.

21. Liolios K et al. *Complete genome sequence of the gliding, heparinolytic Pedobacter saltans type strain (113T)*. Stand Genomic Sci, 2011. **5**:p. 30–40.

Chapter 4C: Biofilm dynamics: a description of microbiota communication and the investigation of a novel medicinal leech symbiont[5]

ABSTRACT

Despite the growing appreciation for the prevalence of mixed transmission (incorporating horizontal and vertical mechanisms) for establishing microbial symbioses, molecular mechanisms that enable these infections are not well understood. The European medicinal leech, *Hirudo verbana*, provides a simple model for examining mixed transmission through a host-secreted substrate and deeper examination of the mucosal microbial community structure using Illumina 16S rRNA sequencing has revealed a diverse microbiota. In this study, we further examine the microbial dynamics within this environment through the use of Illumina-based RNA-seq analyses of shed mucus. In depth investigation into this metatranscriptome revealed an environment open to communication, cultivated by a diverse microbiota. Communication between community members was demonstrated by genetic exchange among community members through the expression of various secretion systems (e.g. Type IV), while genes related to quorum sensing indicate inter- and intraspecies communication and a coordination of gene expression. Additionally, investigation into the most prevalent mucosal microbial symbiont, *Pedobacter*, provides a deeper understanding of how this symbiont may form and retain its relationship with the leech host. The presence of a cooperative microbiota may further support mucus as an environment capable of not only producing genetically variable microbes, but also assisting in transmission of crucial symbionts. By understanding the features that enable mixed transmission, we may further elucidate critical factors in the development of strategies promoting

[5] Reprinted from Ott BM, Rickards A, Driscoll T, Michael JA, Pollio A, Rio RVM. Biofilm dynamics: a description of microbiota communication and the investigation of a novel medicinal leech symbiont. *In prep*, 2015.

beneficial symbioses and prove that transmission substrates are not static environments, but ecosystems worthy of study.

INTRODUCTION

Microbes are widespread throughout our planet, residing in soils, oceans, and even beneath the arctic glaciers [1]. Most interact with each other and other life forms, and there is no known complex organism that lacks an associated microbiome. These microbial symbionts are critical to the survival and proliferation of their host by assisting in nutrient provisioning [2–5], physiological development [6, 7], immunological priming [8, 9], providing protection from pathogens [10–13], contributing to predator evasion [14], etc. As such, the transmission of these symbionts is a crucial aspect of host biology. Despite an increasing appreciation for the high prevalence of mixed transmission (incorporating vertical, or from a parental route, and horizontal, or environmental, mechanisms) for establishing microbial symbioses (reviewed in [15]), features that enable these infections are not well understood. As many symbioses are very complex, with multiple microbial species cohabitating and interacting simultaneously, it is difficult to determine the acquisition mechanism and subsequent impact of transmission on the symbiosis. Recently, however, a simple model system composed of the medicinal leech host (*Hirudo verbana*) and its gut symbiont, the Gammaproteobacterium *Aeromonas veronii*, has been described, allowing for further investigation into factors involved in transmission.

The European medicinal leech (*Hirudo verbana*) has a relatively simple anatomy allowing it to be an exceptional model system for research on microbiota structure and their functional roles towards host biology. The ectoparasitic leech attaches for a blood meal using its anterior sucker, where the jaws are then used to penetrate the skin of a host. Blood first enters into the pharynx (i.e. thoat) before entering into the digestive tract, which is made up of two

main components: the crop, where blood is stored and erythrocytes are condensed, and the small intestinum, where cell lysis and nutrient absorption occur. It is the crop where the gut symbiont, *A. veronii*, resides [16]. While a portion of leech individuals obtain this beneficial gut symbiont vertically via the cocoon albumen (i.e. yolk-like substance) during the larval state [17], by early adulthood all leeches harbor this bacterium [17], indicating the complementation of incomplete vertical transmission by a horizontal mechanism.

Investigation into this horizontal transmission mechanism revealed that regularly shed mucus, which consists of double-layered mucopolysaccharides with a protein cuticula base [18] produced by globose glands irregularly distributed through the leech skin [19], act as a vehicle for *A. veronii* transmission [20]. In addition, these castings carry a diverse microbial community [21], which may not only compete with *A. veronii* for resources, but may also cooperate with this symbiont through nutrient provisioning (Ott *et al*., *in prep*; Chapter 4A).

The base matrix structure of mucus (i.e. mucopolysaccharides with a protein cuticula), along with the presence of the relatively diverse microbial community [21], indicate that these mucosal secretions can be considered biofilms (reviewed in [22, 23]). One major aspect of biofilm sociobiology is community communication, generally occurring through quorum sensing mechanisms [24, 25](reviewed in [26]). Quorum sensing is the regulation of gene expression as a function of population density in a bacterial community (reviewed in [27]). Individual bacterial cells will produce signaling molecules and corresponding receptors to communicate with the surrounding bacterial cells (reviewed in [26]). Additionally, there are certain molecules specific for a communication within a bacterial species (e.g. the LuxR/LuxI system; reviewed in [26]), as well as molecules used for interspecies communication (e.g. Autoinducer-2; [28–30]). This

allows for "secret" communication within members of the same species, while also allowing for biofilm formation in multispecies conglomerations (i.e. microbiotas; reviewed in [26]).

In addition to quorum sensing, high levels of lateral gene transfer (LGT) are also observed in biofilms (reviewed in [23]), likely resulting from high population levels of different microbial species. One mechanism used for LGT by bacteria is the type IV secretion system (T4SS), a secretion apparatus comprised of 12 different proteins and powered by a battery of VirB4 ATPase subunits [31]. While T4SSs are known to translocate a wide variety of virulence factors (reviewed in [32]), DNA is also transferred between cells using the F-pilus of the T4SS when cells are not in direct contact [33].

Beyond bacterial communication, of particular interest regarding the mucosal microbial community is its most abundant member, *Pedobacter* [21]. The genus *Pedobacter* was reclassified from the genus *Sphingobacterium* in 1998 [34]. Phenotypic variations ranging from antibiotic resistance, enzyme production, and the ability to metabolize amino acids and carbohydrates, as well as differences in fatty acid content, led Steyn *et al.* to suggest the new *Pedobacter* genus. Despite the recent discovery of this genus, 49 species have already been described, in addition to the recent genome announcement of a *Pedobacter* leech isolate [35]. However, little is known about these species and their functions. Many *Pedobacter* species are mainly isolated from soil and water samples from a variety of ecosystems ranging tundra soils in Norway [36] to freshwater sediments in South Korea [37]. Several species are known to generate the enzyme heparinase, which breaks down the anticoagulant heparin [38]. Further, heparin is released by mast cells in response to breaks in the skin, hence it has been suggested that heparin plays a role in the immunological response to a pathogen or other foreign body [39]. The role of heparinase in a symbiont of a blood-feeding host is unknown, but may serve to protect the

microbiota or potentially drive the aggregation of red blood cells within the leech crop following blood meal acquisition.

In this study, we addressed two major questions; 1) what mechanisms are used by the mucosal microbiota for communication; and 2) what does a deeper investigation into a major community member reveal regarding its association with the leech host and its genomic repertoire? To investigate these questions, a high-throughput, metatranscriptomic analysis was used to explore genes expressed within leech mucus that were related to quorum sensing and type IV secretion systems, as well as the activity of a single microbe, *Pedobacter*. Semi-quantitative PCR was also used to determine the presence of *Pedobacter* in a number of leech-specific niches. In addition, comparative genomic analyses with two other *Pedobacter* species (i.e. an ancestral strain and a closer phylogenetic relative) were used to examine the genomic repertoire of our novel *Pedobacter species*. Here, we present evidence that leech mucosal secretions are not a static environment, but rather contain a communicative, and possibly cooperative, microbial community, the study of which can provide a plethora of information on novel leech symbionts.

MATERIALS AND METHODS

(a) Leech husbandry. *H. verbana* leeches were obtained from Leeches USA (Westbury, NY, USA) and maintained in sterilized, Leech Strength Instant Ocean H_2O (0.004%, I.O.) in the Department of Biology at West Virginia University at 15°C. Leeches were maintained on defibrinated bovine blood (Hemostat, CA) through an artificial feeding system.

(b) Mucosal sampling. Tanks housing leeches were cleaned to remove all mucus and fresh Leech Strength I.O. was added (denoted as Day 0). Mucosal samples collected the following day

were considered 1 d and were aged to 3 d, 5 d and 8 d at 15°C within sterile Leech Strength I.O. H_2O.

(c)RNA-seq analysis of mucosal secretions. RNA extraction and library preparation were performed as described in Ott *et al*. (*in prep*; Chapter 4A). Subsequently, FastQC analysis was performed on the raw metatranscriptome data to determine Illumina read quality. cDNA reads were assembled *de novo* into contiguous sequences (contigs) using Trinity [40]. The functional annotation of contigs was accomplished using the NCBI non-redundant (nr) database housed on the WVU Shared Research Facilities (SRF) High Performance Computing (HPC) Spruce Knob host. The resulting BLAST file was organized and visualized using MEGAN5 [41], which incorporates KEGG [42, 43] and SEED [44] databases. Data from this study were deposited in the NCBI Sequence Read Archive under the Accession SRP067591.

Contigs specific to microbial community communication (i.e. quorum sensing and type IV secretion systems) were identified using the SEED and KEGG classification systems in MEGAN5. Clustering diagrams were generated with SAS JMP (v.10) [45] using Ward's minimum variance method without standardizing data. This ensured that the values depicted in the graph directly correlated with the raw read count.

Pedobacter-specific contigs were identified by mapping to the *Pedobacter heparinus* genome [46] using BLASTn [47, 48]. Bowtie2 [49], using default parameters, and custom perl scripts (Niel Infante, pers. comm.), were used to determine the number of reads corresponding to each contig. Gene expression levels were determined using reads per kilobase of transcript per million mapped reads (RPKM). Additionally, gene ontology (GO) terms for each *Pedobacter*-specific contig were determined using Blast2GO [50] and refined manually.

(d) Pedobacter localization. Semi-quantitative PCR was used to determine the presence of *Pedobacter* within leech-associated niches. Primers and amplification settings can be found in Table 1. Primers were determined to be specific to *Pedobacter sp*. Hv1 [35] using MEGA (v.6)[51].

(e) Phylogenetic analyses of Pedobacter diversity. *Pedobacter* diversity was examined by constructing a phylogenetic tree using 16S rRNA sequences of numerous *Pedobacter* species including those with and without known heparinase activity. DNA sequences were aligned using the Clustal X algorithm with default settings, and refined manually when necessary. The Bayesian Markov Chain Monte Carlo method was implemented with MrBayes (3.1.2) [52]. The best-fit model (GTR+I+G) used for Bayesian analyses was statistically selected using the Akaike Information Criterion in MrModeltest version 2.3 [53]. Bayesian analyses were performed with six Markov chains [54] for 3,000,000 generations. Posterior probabilities (PP) were calculated, with the stabilization of the model parameters (i.e., burn-in) occurring around 2,400,000 generations. Every 100th tree following stabilization was sampled to determine a 50% majority rule consensus tree. All trees were made using the program FIGTREE v.1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

(f) Comparative Genome Analyses. Comparative analyses of *Pedobacter heparinus* [46] and *Pedobacter saltans* [55] with *Pedobacter sp*. Hv1 [35] were completed by first annotating all genomes using RAST [56, 57], which generated unique peg IDs for each identified coding sequence (CDS). The annotated list of CDSs was processed using reciprocal best BLAST [58] and visualized on a circular plot using Circos [59]. The analysis of synteny in the Circos diagrams reflects the "best-guess" determination of contig order in *Pedobacter sp*. Hv1 by comparison with the respective reference.

Additionally, FastOrtho [60], a modified version of OrthoMCL [61], was used to identify orthologous groups of CDSs unique to each *Pedobacter* species, shared between two *Pedobacter species*, and shared between all three *Pedobacter* species. Translated amino acid sequences binning to the largest OrthoMCL group shared among all three *Pedobacter* species were aligned and examined through construction of a Bayesian tree as mentioned above, with 5,000,000 generations and a burn-in occurring around 4,000,000 generations.

Genes shared between *P. sp* Hv1 and *P. saltans* (and not *P. heparinus*) were determined by manual data mining. After identification of an operon of interest, individual CDSs were gathered and ordered based on the peg ID. Separate contigs were maintained for *P. sp*. Hv1, as the genome has not been fully assembled [35] and contig order is not certain. CDS gaps were confirmed using BLASTx and BLASTn to ensure that functional and sequence differences were consistent with genes not related to the conjugative transposon loci.

RESULTS

Communication among mucus microbial community members revealed through metatranscriptome analysis.

To elucidate mechanisms used for communication by the mucosal microbial community, we first investigated all microbial activity related to quorum sensing in mucosal secretions using the SEED database [44] through the MEGAN5 [41] graphical user interface (GUI) portal (Figure 1). This resulted in the identification of 5 functional categories related to quorum sensing expressed by 17 different genera (Figure 1). Three other groups, which were only identified to the Order and Family level, also expressed genes within at least 1 of these 5 categories (Figure 1). Interestingly, high expression of genes related to Autoinducer 2 synthesis indicate significant interspecies communication, as this molecule is known to regulate quorum sensing between

different species within a community [28–30]. Additionally, *Pedobacter*, the mostly abundant microbe within mucosal microbial community [21], contributes the highest levels of expression in genes related to Autoinducer 2 synthesis (n=697 reads).

While mechanisms related to quorum sensing can be critical for individual microbial cells and a few subsequent generations, the biofilm environment can also enable high levels of long-term information acquisition through mechanisms such as lateral gene transfer (LGT)(reviewed in [23]). Additionally, as it has been previously hypothesized that LGT may occur in significantly high levels within the mucosal environment [21], we investigated gene expression related to LGT, specifically in type IV secretion systems (T4SS), a common mechanism for cell-to-cell gene transfer in bacteria (reviewed in [32, 62, 63]). A total of 30 genes were expressed and identified related to T4SS, expressed by 33 different genera and 7 other groups that were only identified to the Phylum, Subphylum, Order and Family levels. Interestingly, no one species expresses all genes related to the production of the T4SS apparatus. Also, *Pedobacter* produces relatively high levels of T4SS-related genes, though not at the highest quantity or at the highest levels of expression. However, identification of genes related to T4SS and quorum sensing, along with its high population levels [21], implicate *Pedobacter* as a major member of the microbial community.

Novel Pedobacter species as a new leech-associated symbiont.

To further investigate *Pedobacter*'s importance in the leech symbiotic system, the relationship between *Pedobacter* and the leech host was first examined through localization assays. To accomplish this, we used semi-quantitative PCR and *Pedobacter*-specific *gyrB* primers (Table 1) to examine the presence of *Pedobacter* in multiple aspects of the leech anatomy (i.e. anterior sucker, posterior sucker, pre-pharynx, pharynx, gut, exterior epidermis

(skin), and cocoon samples) as well as the water leeches are housed in, with the mucosal environment as a positive control. While the leech suckers, pharynx, and gut environments did not produce a signal (data not shown), a faint signal was observed in 66% of pre-pharynx samples, moderate signal was observed in 50% of epidermal samples (ventral and dorsal sides), and a strong signal was observed in both spent leech water (i.e. 1 w old leech water) and the foam produced by the leech "mother" (i.e. the leech depositing the cocoon) (Figure 3; Supplemental Figure 1). Interestingly, cocoon yolk (i.e. albumen; nutrition provided to the larvae during development) did not produce a signal (Figure 3). Mucus exhibited a very strong signal at all 4 time points examined (i.e. days 1, 3, 5 and 8; positive control), while fresh water (i.e. sterile leech strength instant ocean; negative control) produced no signal (Figure 3). These results indicate that *Pedobacter* is a member of skin and oral microbiotas that are not yet described.

As *Pedobacter* is harbored on the leech skin, this microbe is likely seeded into the mucus as the castings are shed [20]. To determine how *Pedobacter* can subsequently survive within the mucosal environment, a *Pedobacter*-specific transcriptome was obtained by mapping all contigs to the *Pedobacter heparinus* genome sequence [46] using Bowtie 2 [49]. Using Blast2GO, the resulting *Pedobacter*-specific contigs were first assigned to a specific gene through BLASTx (NCBI; non-redundant (nr) database), followed by gene ontology (GO) annotation. These GO terms were then binned into general functional categories and compared based on read count (Figure 4), revealing relative gene expression within the mucosal environment. Based on these results, we see an increased expression in ATP binding and metabolism relative to other functional GO categories, which may indicate high energy consumption. Additionally, a low expression in antibiotic metabolism is also observed, suggesting that *Pedobacter* encounters antibiotic compounds, at least to some degree. Expression of genes related to iron/sulfur binding

supports the hypothesis made in Ott *et al*., (*in prep*; Chapter 4A) that mucosal secretions contain high levels of oxidative stress, as iron/sulfur dependent enzymes can be used to respond to oxidative stress [64]. Interestingly, when we examine individual genes based on expression level, the overexpression (i.e. RPKM > 1,000) of TonB-dependent receptors (RPKM=9,580; Supplemental Table 1) indicates that iron is likely being imported into the bacterial cell, as these receptors are known to transport ferric chelates across the cell wall (reviewed in [65]). As such, it is likely that there is also a source of iron within mucosal secretions that *Pedobacter*, and possibly other community members, can utilize.

Comparative genome analyses between three Pedobacter species show high levels of genome rearrangement and diversity.

While the *Pedobacter* genus is reasonably diverse (Figure 5), there is little to no information regarding genomic synteny (i.e. order of genes) or retained functional capabilities across species. To investigate this, the genomes of three species were chosen: *Pedobacter heparinus* [46], *Pedobacter saltans* [55], and the recently released draft genome of leech isolate, *Pedobacter sp*. Hv1 [35]; and analyzed through a variety of pipelines.

First, the 23 contigs of the *P. sp*. Hv1 genome [35] was aligned with both *P. heparinus* (i.e. the closer phylogenetic relative; Figure 5) and *P. saltans* (i.e. the "ancestral" species; Figure 5) using a reciprocal best BLAST approach [60] to compare coding sequences (CDS), with the results plotted on circular diagrams. Through visual examination, significant differences in both genomic synteny and content between *P. heparinus*-*P. sp.* Hv1 (Figure 6) and *P. saltans*-*P. sp.* Hv1 (Figure 7) are immediately apparent. First, *P. saltans* (36.6%) and *P. sp*. Hv1 (37.2%) have closer GC contents compared with *P. heparinus* (42.05%) and *P. sp*. Hv1. Additionally, while a number of CDS gaps (i.e. a CDS in missing one of the species) between *P. heparinus* and *P. sp*.

Hv1 genomes are present (Figure 6), this is dwarfed by the quantity observed between *P. saltans* and *P. sp.* Hv1 (Figure 7). Similarly, synteny, which is inversely correlated to the central "hairball" in each diagram, is much higher between *P. heparinus* and *P. sp*. Hv1 compared with *P. saltans* and *P. sp.* Hv1. While this is not entirely surprising, as *P. heparinus* and *P. sp*. Hv1 are more phylogenetically related to each other than either is to *P. saltans* (Figure 5), there are still significantly high levels of genomic rearrangement given the 16S rRNA similarity cut-off observed for members of the same genus (95-97%; reviewed in [66]).

In addition to visual inspection, orthologous groups shared among all three species, among two species, or that were unique to a single species, were identified with FastOrtho [60], and the number of CDSs were plotted in a venn diagram (Figure 8). The core genome is revealed as the CDSs shared among all three species (n=1808), which contains the highest quantity of CDSs. Interestingly, beyond the core genome, each species contains more unique genes than those shared with another species (*P. heparinus* n=1177; *P. saltans* n=1252; *P. sp.* Hv1 n=1552), although *P. heparinus* and *P. sp.* Hv1 share more CDSs (n=500) than either share with *P. saltans* (n=426, n= 247; respectively)(Figure 8). In summary, genome-wide comparisons across these three *Pedobacter* species indicate high levels of genome rearrangement and significant differences in genomic content.

While whole genome comparisons show overall genomic similarity, investigation into individual gene groups can elucidate groups of genes, and corresponding functional capabilities, which are lost or retained across species. To examine such genes, we first re-inspected the orthologous groups (OrthoMCLs) generated by FastOrtho described above. The largest OrthoMCL (OrthoMCL0) was one shared between all three species and contained CDSs related to TonB-dependent receptors, which are known to be highly expressed by *P. sp.* Hv1 in the

mucosal environment (see above). To further investigate the functional similarity of these genes across species, all translated amino acid sequences that binned to OrthoMCL0 were used to generate a phylogenetic tree (Figure 8, Sub-A). When examining individual terminal nodes, in most cases sequences isolated from different species significantly diverge from each other. However, all sequences from one species do not necessarily separate into separate clusters, but rather will group with TonB-dependent receptors from other species. Interestingly, while 18 and 19 CDSs were identified in *P. saltans* and *P. heparinus*, respectively, only two TonB-dependent receptor CDSs were found in *P. sp* Hv1, which did not cluster together.

Additionally, a separate gene group related to the conjugative transposon (Tra) operon was observed only in *P. saltans* and *P. sp*. Hv1 (Figure 8, Sub-B). As each genome examined in this study was annotated through RAST [56], every CDS had an assigned peg ID. Using these peg IDs, we were able to reconstruct the Tra operon in *P. saltans* and *P. sp.* Hv1 (Figure 8, Sub-B). In *P. saltans*, there is a duplication event of the Tra operon, with one copy in the positive strand (+ve) and one copy in the negative strand (-ve). A clear duplication event is also present in Contigs 9 and 11 of *P. sp.* Hv1. Insertions and deletions not related to the Tra operon can be observed in both *P. saltans* and *P. sp*. Hv1 Tra operons, which may explain the lack of certain Tra genes in both species (Table 3).

Finally, due to the localization of *P. sp.* Hv1 in the leech pre-pharynx (Figure 3), we searched for the presence of heparinase-related genes, which encode enzymes that degrade the anticoagulant, heparin [67]. Heparinases produced by an oral symbiont may assist the leech during feeding. While heparinase genes were found in both *P. heparinus* and *P. saltans*, as well as other *Pedobacter* species (Figure 5), no heparinase genes were found in *P. sp*. Hv1.

DISCUSSION

In this study, we described microbial community dynamics within leech mucosal secretions through examination of communication mechanisms (i.e. quorum sensing and type IV secretion systems) and we demonstrate that further investigation into a major mucosal microbiota player (i.e. *Pedobacter*) reveals a novel leech symbiont.

As mucus provides a matrix for the recruitment and aggregation of a diverse microbiota [21, 68], this environment can be defined as a biofilm (reviewed in [22, 23]). One major aspect of biofilm sociobiology is community communication, generally occurring through quorum sensing mechanisms [24, 25](reviewed in [26]). To determine whether the mucosal microbial community expressed genes related to quorum sensing, the entire metatranscriptome was assembled *de novo* (Table 2) and the resulting contiguous sequences (contigs) were assigned a function using BLASTx [47] and sorted using the SEED [44] and KEGG [42, 43] databases through MEGAN5 [41]. This resulted in 5 genes expressed by ~20 different microbial groups (Figure 1). The expression of LuxR-transcriptional activators indicates an intraspecies population-density signal [69, 70], while the type III secretion system, two-component sensor kinase, BarA, is involved in the regulation of cellular motility and pathogenesis [71]. Additionally, the relatively high expression of genes related to Autoinducer 2 (AI-2) synthesis indicates that a number of microbial community members engage in interspecies communication, as the AI-2 signal molecule plays a major role in this form of bacterial correspondence (reviewed in [28–30]). However, low levels of AI-2 transport and processing can be observed (Figure 1), which may indicate that, while a number of microbes are sending out signals, they are not processing signals in return. Alternatively, these observations could be an artifact of sequencing, and deeper sequencing of the mucosal metatranscriptome may be required to determine if this low signal is real. In addition to deeper sequencing, future studies should investigate what

specific, quorum-dependent activities may be activated as a result of these signals and how an increase or decrease in signal can alter the microbial community metatranscriptome and even the population dynamics within the mucus.

The relatively higher microbial diversity in the leech exuded mucus relative to the leech crop [72, 73] could also provide ample opportunities for lateral gene transfer (LGT) [74–78] through type IV secretion systems (T4SS) (reviewed in [32, 62, 63]). This form of communication can impact mucus community members for many generations by allowing acquisition of genes necessary for adaptation to prospective habitats. As such, we pursued genes related to T4SS activity using the same bioinformatics techniques mentioned above (i.e. for quorum sensing). From this data, we observed a high quantity of genes expressed by double the number of microbial groups (Figure 2) involved in quorum sensing. What is more interesting, however, is that no one member of the microbiota expresses all genes related to the physical components and assembly of the T4SS apparatus (Figure 2). This may be an artefact of sequencing depth, or it may be a result of genetic pleiotropy, or genes that encode for proteins involved in different cellular processes [79]. To distinguish between these two possibilities, deeper sequencing of the mucosal metatranscriptome, as proposed with quorum sensing, may help elucidate the validity of our findings.

An interesting aspect regarding both communication mechanisms examined in this study is the participation of the major mucosal community member, *Pedobacter*. This microbe was first described within leech mucosal secretions, and is also the most abundant member of the mucus microbiota [21]. As such, we decided to pursue this novel microbe in more depth. First, we showed that *Pedobacter* is not only a member of the mucosal microbial community, but is also associated with currently unknown skin and oral microbiotas (i.e. epidermal swab and pre-

pharynx samples, respectively; Figure 3). To expand on our investigation of a skin niche, our studies show that not all skin samples used in this study contained *Pedobacter*. This could be explained if *Pedobacter* form clusters on the skin rather than being evenly distributed, and thus may have been missed during our swabbing procedure. Alternatively, this uneven isolation of the *Pedobacter* skin population may result if a mucosal casting was recently shed, removing most of the skin microbiota in the process. To investigate this, future experiments should examine leech skin swabs immediately after a secretion is shed and 24 hours later as a new secretion is being produced [20].

Interestingly, *Pedobacter* is also found in the foam produced by the leech "mother" (i.e. the leech parent that deposits the cocoon) to protect the cocoon (Supplemental Figure 1). As these secretions are produced in the mouth of the mother during deposition and smeared over the cocoon [19], this may indicate the origin of *Pedobacter* in the foam (i.e. via the pre-pharynx). Alternatively, *Pedobacter* is also found in "spent" leech water (i.e. water housed with leeches for ~ 1 week), indicating that the surrounding environment may also inoculate the foam. Future experiments should investigate the viability of *Pedobacter* in both the leech pre-pharynx and the foam of cocoons. *Pedobacter* presence in the cocoon could be considered a form of vertical transmission (i.e. from the pre-pharynx of the mother) or horizontal transmission (i.e. from the surrounding H_2O), or a mixture of both, if emerging leech juveniles acquire *Pedobacter* from the foam.

While *Pedobacter* was found in the oral and skin microbiotas, as well as associated with the protective foam of the leech cocoon, we did not observe any signal indicating that *Pedobacter* was present in the gut. Although metagenomic studies of the digestive tract microbiota do show the presence of *Pedobacter* ([72], supplemental data), the quantity is so low

that we were unable to obtain a signal via semi-quantitative PCR. This low presence within the gut implies that *Pedobacter* is a transient microbe and does not reside within the gut, but as part of an oral microbiota, may be swallowed and thus observed using deep sequencing techniques.

In addition to localization studies, we also examined the *Pedobacter* transcriptome within mucosal castings to elucidate what processes this microbe relies on in stressful environmental conditions (Ott *et al*., *submitted*; Chapter 4A). To accomplish this, *Pedobacter*-specific contigs were identified by mapping to the *Pedobacter heparinus* genome [46] using BLASTn [47, 48]. Expression levels were based on number of reads (obtained using Bowtie2 [49], and RPKM values. Additionally, gene ontology (GO) terms for each *Pedobacter*-specific contig were determined using Blast2GO [50]. The resulting functional categories were plotted against the number of reads binned to each category. By comparing relative expression levels between these categories, we determined that *Pedobacter* relies heavily on ATP-binding and metabolism, which may indicate a high level of energy generation and/or consumption. Additionally, expression of genes related to antibiotic metabolism and iron/sulfur binding support the hypothesis that mucus is a stressful environment, likely resulting from a competitive microbial community and oxidative stress, respectively. As iron/sulfur complexes can be used to respond to oxidative stress [64], this may be a mechanism employed by *Pedobacter* to combat to such stressors. The presence of oxidative stress is not surprising, as studies examining the dynamics of a major gut symbiont, *Aeromonas veronii*, within mucosal secretions (Ott *et al*., *submitted*; Chapter 4A) also indicate increased levels of oxidative stress compared with other leech-associated environments. Additionally, the expression of iron/sulfur binding may indicate the availability of these compounds within the mucosal environment. High expression levels (i.e. RPKM > 1,000) of TonB-dependent receptors (RPKM=9,580; Supplemental Table 1) support

this hypothesis, as these receptors are involved in iron, vitamin B12, nickel and carbohydrate acquisition (reviewed in [65]). We can strengthen these conclusions, however, if the mucosal transcriptome were compared with other leech-associated niches, such as the skin, oral and free-living (i.e. surrounding leech water) lifestyles. This would elucidate any shifts in expression patterns based on the available nutrients, surrounding microbial communites, and the environmental stressors (which are not mutually exclusive). Therefore, future studies should examine *Pedobacter* activity within all of these different environments to gain a more complete picture of how *Pedobacter* survives and proliferates.

While these findings do provide some information regarding the relationship *Pedobacter* has with the leech host and the mucosal microbial community, this microbe is still a novel member of the *Pedobacter* genus (Figure 5), representatives of which are isolated from various niches around the world (e.g. tundra soil in norway [36] to fresh water sediment in South Korea [37]). Investigation into its genomic repertoire may allow for a deeper understanding in how this microbe is able to establish a relationship with the leech host. To accomplish this, we obtained the genomes of two other *Pedobacter* species, *P. heparinius* (i.e. a closer phylogenetic relative; [46]) and *P. saltans* (i.e. the "ancestral" species; [55]), to use in comparative analyses with a recently released, draft genome of our *Pedobacter* isolate, *P. sp.* Hv1 [35]. A reciprocal best blast approach was first implemented, which compared the coding sequences (CDS) of *P. heparinus* and *P. saltans* with *P. sp.* Hv1 to determine homologous sequences across the genomes. These analyses were then visualized using circular diagrams, showing that while closer phylogenetic relatedness (i.e. *P. heparinus*-*P. sp.* Hv1 comparison (Figure 6) versus *P. saltans*-*P. sp.* Hv1 (Figure 7)), correlates with a greater similarity in genomic content and synteny (i.e. the order of genes in the genome), genomic rearrangements are still rampant considering the

phylogenetic similarity cut off that defines a genus (i.e. 16S rRNA similarity of 95-97%; reviewed in [66]). While the reciprocal best blast technique used a best-guess approach to determine contig arrangement in the *P. sp*. Hv1 draft genome (as it still consists of 23 separate contigs), the number of genomic rearrangements still seems high (Figure 6 and Figure 7; *P. heparinus* and *P. saltans*, respectively). However, future studies should confirm this high level of rearrangement among *Pedobacter* species by completing the *P. sp*. Hv1 genome using a primer-walking technique or through long-read, PacBio sequencing.

As visual interpretations provide limited information, FastOrtho [60] was used to generate orthologous groups (OrthoMCLs) of CDSs, which were used quantify the number of CDSs shared between all 3 species, between 2 species, and those that were unique to each species. While the core genome (i.e. CDSs shared between all 3 species) contained the highest number of CDSs (n=1,808; Figure 8), there were also a high number of CDSs unique to each species as well (i.e. *P. sp.* Hv1, n=1,552; *P. saltans*, n=1,252; *P. heparinus*, n=1,177; Figure 8). This indicates that these species are also highly divergent regarding genome content, supporting our visual interpretations.

While whole genome analyses give an overall picture of genomic similarity, it was also important to determine which specific genes are retained across species and which genes are lost over evolutionary time, as this highlights the genomic repertoire required for each species to exist in its own niche. By reevaluating our OrthoMCL groups, we determined that the largest group of CDSs, which also composed part of the *Pedobacter* core genome, were related to TonB-dependent receptors. This is interesting, as our transcriptomic study indicated that *P. sp*. Hv1 highly expresses these receptors within the mucosal environment (Supplemental Table 1). Additionally, as TonB-dependent receptors are involved in iron, vitamin B12, nickel and

carbohydrate acquisition (reviewed in [65]), this may be a mechanism required for its symbiotic relationship with the leech and/or its involvement in the mucosal microbial community. Therefore, we investigated the functional similarity of these genes across species using a phylogenetic analysis. To accomplish this, translated sequences of all CDSs binned to this OrthoMCL were aligned, and a Bayesian analysis was performed. When examining individual terminal nodes in the resulting phylogenetic tree (Figure 8, Sub-A), most sequences isolated from different species diverged significantly from each other. However, sequences from one species do not necessarily separate into different clusters, but may also group with TonB-dependent receptors from other species (Figure 8, Sub-A). Interestingly, while 18 and 19 CDSs were identified in *P. saltans* and *P. heparinus*, respectively, only two TonB-dependent receptor CDSs were found in *P. sp* Hv1, which did not cluster together. As these genes do not exhibit separation among species, which would indicate a specificity in function for a particular ecological niche [80, 81], but rather show signs of LGT and recombination across niches, these TonB-dependent receptors may be housekeeping genes [81]. Future studies should investigate the importance of these TonB receptors in the different *Pedobacter* species and determine why *P. sp*. Hv1 only has two copies compared with the numerous copies found in *P. heparinus* and *P. saltans.*

In addition to TonB-dependent receptors, we also identified a conjugative transposon (Tra) operon. Interestingly, this operon, which promotes DNA transfer between cells by conferring donor ability to the host bacterial cell (reviewed in [82]), is absent in *P. heparinus*. However, the Tra operon is not only present in the ancestral *P. saltans* and the leech isolate, *P. sp.* Hv1 (Figure 8, Sub-B), but duplication events are also apparent in both genomes. Although we were unable to consolidate individual *P. sp.* Hv1 contigs, as this genome is still only a draft

[35], a clear duplication event is still evident in Contigs 9 and 11 (Figure 8, Sub-B). As the Tra-related CDSs are not located near the beginning or end of the contigs in question, and as the CDSs surrounding the Tra sequences were functionally distinct (confirmed through BLASTx and BLASTn), we can assume the observed duplication event is real. Additionally, the presence of CDS insertions (not related to the Tra operon) and deletions in both *P. saltans* and *P. sp*. Hv1 Tra operons is intriguing, as some representatives of the Tra operon are missing in both species (Table 3; reviewed in [82]). It is possible that the gaps observed in the *P. saltans* and *P. sp.* Hv1 Tra operons could be associated with the missing Tra genes. Furthermore, the presence and frequency of indels within this operon could be used to elucidate the evolutionary age of the Tra operon insertion into the *Pedobacter* genome, as well as examine whether this operon was obtained by an ancestor of *P. saltans* and *P. sp.* Hv1, indicating a loss of this gene segement in *P. heparinus*, or if *P. sp*. Hv1 acquired the operon more recently, indicating a loss and subsequent gain of function. Future studies should also investigate what function the Tra operon serves in *P. sp*. Hv1. Is it possible that the presence of this operon allows for LGT? Would loss of the operon eliminate its observed activity regarding T4SS, and could this impact its ability to survive in the mucus? Also, would extra copies increase its involvement in the T4SS system? By answering these questions, we may be able to better understand *P. sp.* Hv1's role as a major mucosal microbiota member.

Finally, we were highly interested in whether heparinase genes are shared amongst all three *Pedobacter* species. The presence of heparinase-related genes would explain the localization of *Pedobacter* in the leech pre-pharynx. As heparinase is an enzyme that degrades the anticoagulant, heparin, production of this enzyme by an oral symbiont may assist the leech host during feeding. Although this study and previous studies demonstrate the presence of

heparinase-related genes within *P. heparinus* [46] and *P. saltans* [55], extensive searches (e.g. reciprocal best blast, manual searches, and basic BLASTx) did not reveal any copies of heparinase-encoding genes within *P. sp.* Hv1. However, this does not necessarily signify the absence of homologous genes in this genome. As the genome is still a draft, it is possible that heparinase-related genes exist within the gaps between contigs. Completing would the genome allow for the potential identification of these genes, and generating PCR primers for heparinase genes may confirm their presence of absence within the microbe. Although primer building for these genes has been attempted (data not shown), it was determined that *Pedobacter*-related heparinase genes are highly divergent, making primer generation in conserved regions difficult. However, further investigation into heparinase gene sequences using Bayesian analysis and multiple levels of multiple sequence alignment (i.e. doing a pairwise alignment, and add one additional sequence at a time) may help identify conserved regions.

Although this data is very preliminary, each result spurs numerous future experiments that will not only elucidate microbial community communication, but also further our understanding of how a major microbial community member can be just as important as an entire microbial community. Hopefully, this research will eventually provide many exciting years for future students while also further elucidating how a microbial transmission substrate is not a static environment, but an ecosystem worthy of study.

REFERENCES

1. Skidmore ML et al. *Microbial Life beneath a High Arctic Glacier*. Appl. Environ. Microbiol., 2000. **66**:p. 3214–320.
2. Douglas AE. *Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera*. Annu. Rev. Entomol., 1998. **43**:p. 17–37.
3. Edwards MA et al. *Microbiologic Assay for the Thiamine Content of Blood of Various Species of Animals and Man*. Am. J. Clin. Nutr., 1957. **5**:p. 51–5.
4. Graf J. *The effect of symbionts on the physiology of Hirudo medicinalis, the medicinal leech*. Invert. Reprod. Dev., 2002. **41**:p. 269–75.
5. Snyder AK et al. *The tsetse fly obligate mutualist Wigglesworthia morsitans alters gene expression and population density via exogenous nutrient provisioning*. Appl. Environ. Microbiol., 2012. **78**:p. 7792–7.
6. Hooper L V. *Bacterial contributions to mammalian gut development*. Trends Microbiol., 2004. **12**:p. 129–34.
7. McFall-Ngai MJ. *Unseen forces: the influence of bacteria on animal development*. Dev. Biol., 2002. **242**:p. 1–14.
8. O'Hara AM, Shanahan F. *The gut flora as a forgotten organ*. EMBO Rep., 2006. **7**:p. 688–93.
9. Round JL et al. *The Toll-Like Receptor 2 Pathway Establishes Colonization by a Commensal of the Human Microbiota*. Science, 2011. **332**:p. 974–7.
10. Moreira LA et al. *A Wolbachia symbiont in Aedes aegypti limits infection with dengue, Chikungunya, and Plasmodium*. Cell, 2009. **139**:p. 1268–78.
11. Hedges LM et al. *Wolbachia and virus protection in insects*. Science, 2008. **322**:p. 702.

12. Teixeira L et al. *The bacterial symbiont Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster*. PLoS Biol., 2008. **6**:p. e2.

13. Oliver KM et al. *Facultative bacterial symbionts in aphids confer resistance to parasitic wasps*. Proc. Natl. Acad. Sci. U.S.A., 2003. **100**:p. 1803–7.

14. McFall-Ngai MJ. *The development of cooperative associations between animals and bacteria: Establishing detente among domains*. Am. Zool., 1998. **38**:p. 593–608.

15. Ebert D. *The Epidemiology and Evolution of Symbionts with Mixed-Mode Transmission*. Annu. Rev. Ecol. Evol. Syst., 2013. **44**:p. 623–43.

16. 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**:p. 1–7.

17. Rio R V et al. *Symbiont succession during embryonic development of the European medicinal leech, Hirudo verbana*. Appl. Env. Microbiol., 2009. **75**:p. 6890–5.

18. Michalsen A et al. *Medicinal Leech Therapy* 2007. Thieme Medical Publishers, New York, New York.

19. Sawyer RT. *Leech Biology and Behaviour* 1986. Clarendon Press, Oxford, United Kingdom.

20. Ott BM et al. *Hitchhiking of host biology by beneficial symbionts enhances transmission*. Sci. Rep., 2014. **4**:p. 5825.

21. Ott BM et al. *Characterization of shed medicinal leech mucus reveals a diverse microbiota*. Front. Microbiol., 2015. **5**.

22. Nadell CD et al. *The sociobiology of biofilms*. FEMS Microbiol. Rev., 2009. **33**:p. 206–24.

23. Stoodley P et al. *Biofilms as complex differentiated communities*. Annu. Rev. Microbiol., 2002. **56**:p. 187–209.

24. McLean RJ. et al. *Evidence of autoinducer activity in naturally occurring biofilms*. FEMS Microbiol. Lett., 2006. **154**:p. 259–63.

25. Rickard AH et al. *Autoinducer 2: a concentration-dependent signal for mutualistic bacterial biofilm growth*. Mol Microbiol, 2006. **60**:p. 1446–56.

26. Miller MB, Bassler BL. *Quorum sensing in bacteria*. Annu. Rev. Microbiol., 2001. **55**:p. 165–99.

27. Marques JC et al. *Processing the interspecies quorum-sensing signal autoinducer-2 (AI-2): characterization of phospho-(S)-4,5-dihydroxy-2,3-pentanedione isomerization by LsrG protein*. J. Biol. Chem., 2011. **286**:p. 18331–43.

28. Bassler BL et al. *Cross-species induction of luminescence in the quorum-sensing bacterium Vibrio harveyi*. J. Bacteriol., 1997. **179**:p. 4043–5.

29. Xavier KB, Bassler BL. *LuxS quorum sensing: more than just a numbers game*. Curr. Opin. Microbiol., 2003. **6**:p. 191–7.

30. Federle MJ. *Autoinducer-2-based chemical communication in bacteria: complexities of interspecies signaling*. Contrib. Microbiol., 2009. **16**:p. 18–32.

31. Low HH et al. *Structure of a type IV secretion system*. Nature, 2014. **508**:p. 550–3.

32. Wallden K et al. *Type IV secretion systems: versatility and diversity in function*. Cell. Microbiol., 2010. **12**:p. 1203–12.

33. Babic A et al. *Direct visualization of horizontal gene transfer*. Science, 2008. **319**:p. 1533–6.

34. Steyn PL et al. *Classification of heparinolytic bacteria into a new genus, Pedobacter,*

comprising four species: Pedobacter heparinus comb. nov., Pedobacter piscium comb. nov., Pedobacter africanus sp. nov. and Pedobacter saltans sp. nov. proposal of the family Sphingobacteriaceae. Int. J. Syst. Bacteriol., 1998. **48 Pt 1**:p. 165–77.

35. Ott BM et al. *Draft Genome Sequence of Pedobacter sp. Strain Hv1, an Isolate from Medicinal Leech Mucosal Castings*. Genome Announc., 2015. **3**:p. e01469–15.
36. Yin Y et al. *Genome sequence of Pedobacter arcticus sp. nov., a sea ice bacterium isolated from tundra soil*. J. Bacteriol., 2012. **194**:p. 6688.
37. An DS et al. *Pedobacter daechungensis sp. nov., from freshwater lake sediment in South Korea*. Int. J. Syst. Evol. Microbiol., 2009. **59**:p. 69–72.
38. Shaya D et al. *Crystal structure of heparinase II from Pedobacter heparinus and its complex with a disaccharide product*. J. Biol. Chem., 2006. **281**:p. 15525–35.
39. Nader HB et al. *New insights on the specificity of heparin and heparan sulfate lyases from Flavobacterium heparinum revealed by the use of synthetic derivatives of K5 polysaccharide from E. coli and 2-O-desulfated heparin*. Glycoconj. J., 1999. **16**:p. 265–70.
40. Grabherr MG et al. *Full-length transcriptome assembly from RNA-Seq data without a reference genome*. Nat. Biotechnol., 2011. **29**:p. 644–52.
41. Huson DH et al. *Integrative analysis of environmental sequences using MEGAN4.*. Genome Res., 2011. **21**:p. 1552–60.
42. Kanehisa M, Goto S. *KEGG: kyoto encyclopedia of genes and genomes*. Nucleic Acids Res., 2000. **28**:p. 27–30.
43. Kanehisa M et al. *Data, information, knowledge and principle: back to metabolism in KEGG*. Nucleic Acids Res., 2014. **42**:p. D199–205.

44. Overbeek R et al. *The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)*. Nucleic Acids Res., 2014. **42**:p. D206–14.

45. SAS Institute Inc. *JMP*. 10. Cary, N.C.

46. Han C et al. *Complete genome sequence of Pedobacter heparinus type strain (HIM 762-3T)*. Stand. Genomic Sci., 2009. **1**:p. 54–62.

47. Camacho C et al. *BLAST+: architecture and applications*. BMC Bioinformatics, 2009. **10**:p. 421.

48. Altschul SF et al. *Basic local alignment search tool*. J. Mol. Biol., 1990. **215**:p. 403–10.

49. Langmead B, Salzberg SL. *Fast gapped-read alignment with Bowtie 2*. Nat. Methods, 2012. **9**:p. 357–9.

50. Conesa A et al. *Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research*. Bioinformatics, 2005. **21**:p. 3674–6.

51. Tamura K et al. *MEGA6: Molecular Evolutionary Genetics Analysis version 6.0*. Mol. Biol. Evol., 2013. **30**:p. 2725–9.

52. Ronquist F, Huelsenbeck JP. *MrBayes 3: Bayesian phylogenetic inference under mixed models*. Bioinformatics, 2003. **19**:p. 1572-4.

53. Nylander JAA. *MrModeltest v2*. 2004. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.

54. Larget B, Simon DL. *Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees*. Mol. Biol. Evol., 1999. **16**:p. 750–9.

55. Liolios K et al. *Complete genome sequence of the gliding, heparinolytic Pedobacter saltans type strain (113T)*. Stand. Genomic Sci., 2011. **5**:p. 30–40.

56. Aziz RK et al. *The RAST Server: rapid annotations using subsystems technology*. BMC

Genomics, 2008. **9**:p. 75.

57. Meyer F et al. *The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes*. BMC Bioinformatics, 2008. **9**:p. 386.

58. Ward N, Moreno-Hagelsieb G. *Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: how much do we miss?* PLoS One, 2014. **9**:p. e101850.

59. Krzywinski M et al. *Circos: an information aesthetic for comparative genomics*. Genome Res., 2009. **19**:p. 1639–45.

60. Gillespie JJ et al. *Genomic diversification in strains of Rickettsia felis Isolated from different arthropods*. Genome Biol. Evol., 2015. **7**:p. 35–56.

61. Li L et al. *OrthoMCL: identification of ortholog groups for eukaryotic genomes*. Genome Res., 2003. **13**:p. 2178–89.

62. Fronzes R et al. *The structural biology of type IV secretion systems*. Nat. Rev. Microbiol., 2009. **7**:p. 703–14.

63. de la Cruz F et al. *Conjugative DNA metabolism in Gram-negative bacteria*. FEMS Microbiol. Rev., 2010. **34**:p. 18–40.

64. Zheng L et al. *Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis*. Proc. Natl. Acad. Sci. U.S.A., 1993. **90**:p. 2754–8.

65. Noinaj N et al. *TonB-dependent transporters: regulation, structure, and function*. Annu. Rev. Microbiol., 2010. **64**:p. 43–60.

66. Janda JM, Abbott SL. *16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls*. J. Clin. Microbiol., 2007. **45**:p. 2761–4.

67. Payza AN, Korn ED. *Bacterial degradation of heparin*. Nature, 1956. **177**:p. 88–9.

68. Tolker-Nielsen T et al. *Development and dynamics of Pseudomonas sp. biofilms*. J. Bacteriol., 2000. **182**:p. 6482–9.

69. Greenberg E. *Quorum sensing in gram-negative bacteria*. ASM News, 1997.

70. Hastings JW, Greenberg EP. *Quorum Sensing: the Explanation of a Curious Phenomenon Reveals a Common Characteristic of Bacteria*. J. Bacteriol., 1999. **181**:p. 2667–8.

71. Teplitski M et al. *Pathways Leading from BarA/SirA to Motility and Virulence Gene Expression in Salmonella*. J. Bacteriol., 2003. **185**:p. 7257–65.

72. Maltz MA et al. *Metagenomic analysis of the medicinal leech gut microbiota*. Front. Microbiol., 2014. **5**:p. 151.

73. Worthen PL et al. *Culture-independent characterization of the digestive-tract microbiota of the medicinal leech reveals a tripartite symbiosis*. Appl. Env. Microbiol., 2006. **72**:p. 4775–81.

74. Raz Y, Tannenbaum E. *The influence of horizontal gene transfer on the mean fitness of unicellular populations in static environments*. Genetics, 2010. **185**:p. 327–37.

75. Beaber JW et al. *SOS response promotes horizontal dissemination of antibiotic resistance genes*. Nature, 2004. **427**:p. 72–4.

76. Foster PL. *Stress-induced mutagenesis in bacteria*. Crit. Rev. Biochem. Mol. Biol., **42**:p. 373–97.

77. Bjedov I et al. *Stress-induced mutagenesis in bacteria*. Science, 2003. **300**:p. 1404–9.

78. Velkov V V. *How environmental factors regulate mutagenesis and gene transfer in microorganisms*. J. Biosci., 1999. **24**:p. 529–59.

79. van de Peppel J, Holstege FCP. *Multifunctional genes*. Mol. Syst. Biol., 2005. **1**:p. E1–E2.

80. Abzhanov A et al. *The calmodulin pathway and evolution of elongated beak morphology*

in Darwin's finches. Nature, 2006. **442**:p. 563–7.

81. Silver AC et al. *Complex evolutionary history of the Aeromonas veronii group revealed by host interaction and DNA sequence data*. PLoS One, 2011. **6**:p. e16751.

82. Firth N et al. *Structure and function of the F factor and mechanism of conjugation*, p. 237–401. *In* Neidhard, FC, Curtiss III, R, Ingraham, JL, Lin, ECC, Low, KB, Magasanik, B, Reznikoff, WS, Riley, M, Schaechter, M, Umbarger, HE (eds.), Escherichia coli and Salmonella: cellular and molecular biology, 1996., 2nd ed. ASM Press, Washington, D.C.

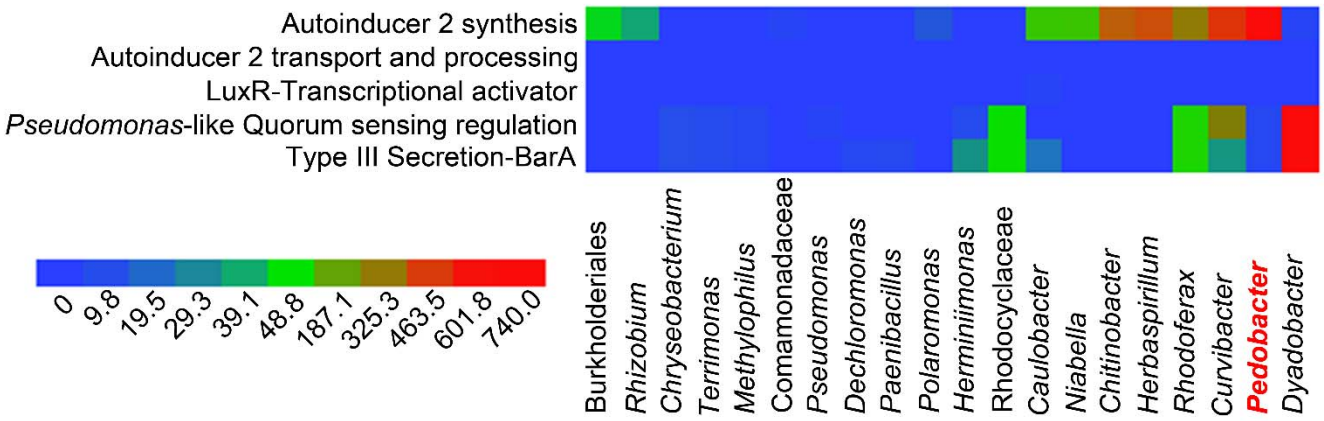

Figure 4C-1. Mucosal microbial community members engaging in quorum sensing. A clustering diagram depicting all community members (x-axis) expressing genes involved in quorum sensing (y-axis; i.e. mechanisms known to be involved in microbe-to-microbe communication) as determined by SEED analysis. Color bar indicates the number of reads identified in each gene category produced by each microbial member, with blue indicating the lowest and red indicating the highest expression levels.

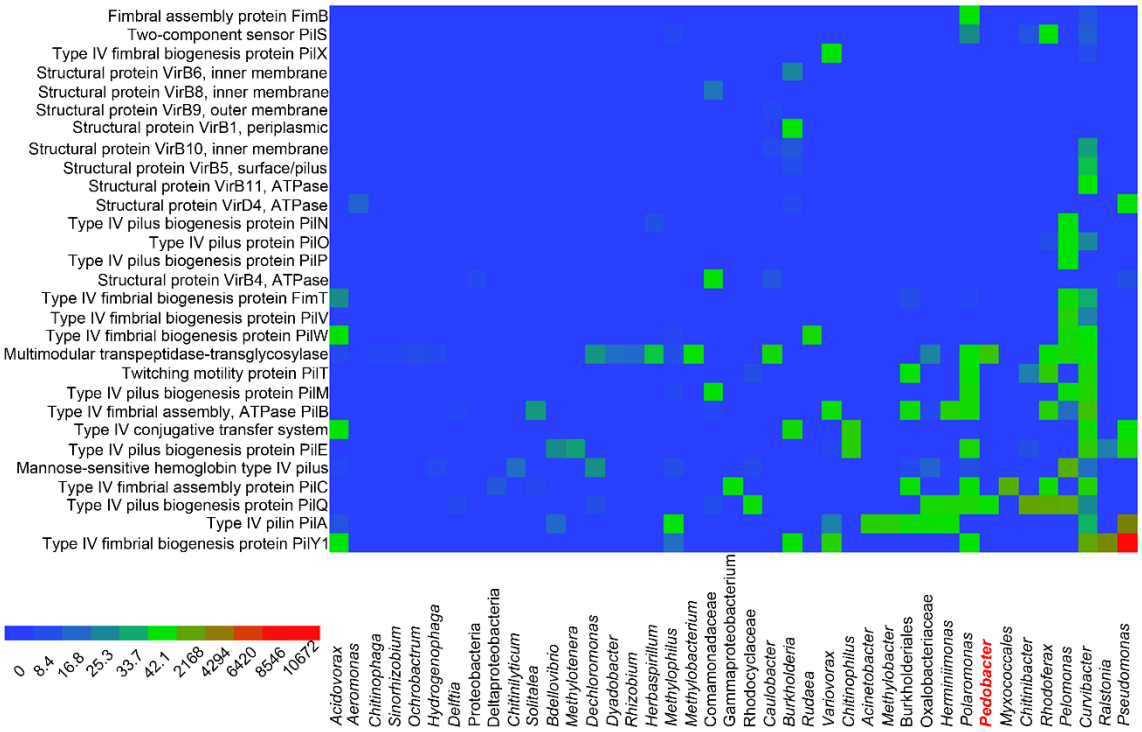

Figure 4C-2. Mucosal microbial community members involved in type IV secretion systems (T4SS). A clustering diagram depicting all community members (x-axis) expressing genes involved in T4SS (y-axis; i.e. a secretion system known to be involved in DNA transfer between microbes) as determined by both SEED and KEGG analysis. Color bar indicates the number of reads identified in each gene category produced by each microbial member, with blue indicating the lowest and red indicating the highest expression levels.

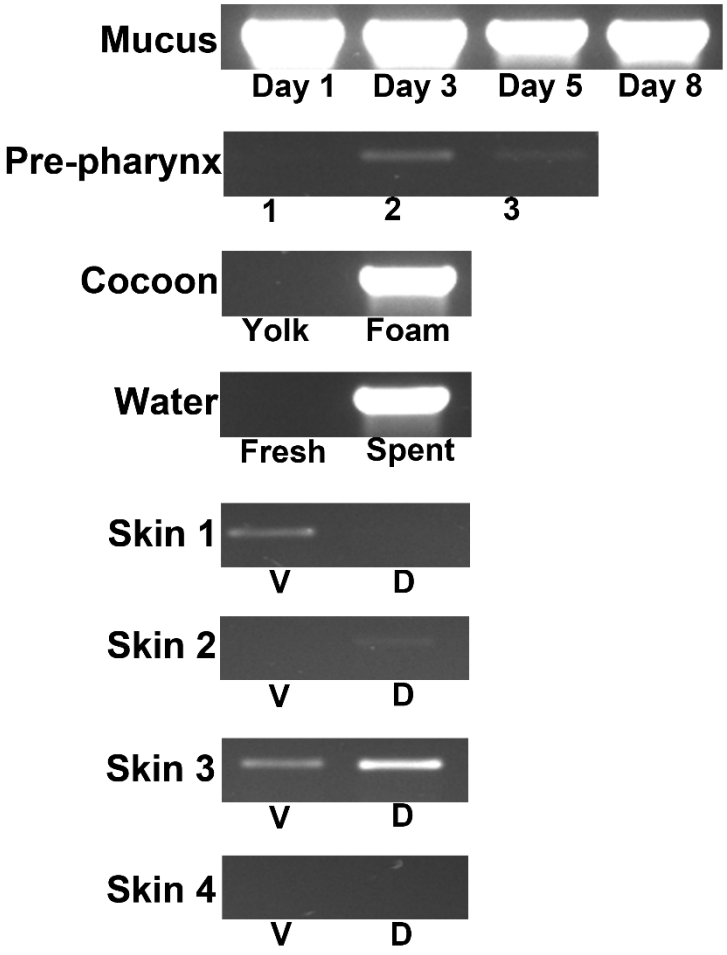

Figure 4C-3. Localization of *Pedobacter* in respective leech-associated niches. Semi-quantitative PCR using *Pedobacter*-specific *gyrB* primers (Table 1) show the presence of *Pedobacter* in leech-associated niches. For skin samples, "V" indicates ventral leech swab, while "D" indicates dorsal leech swab.

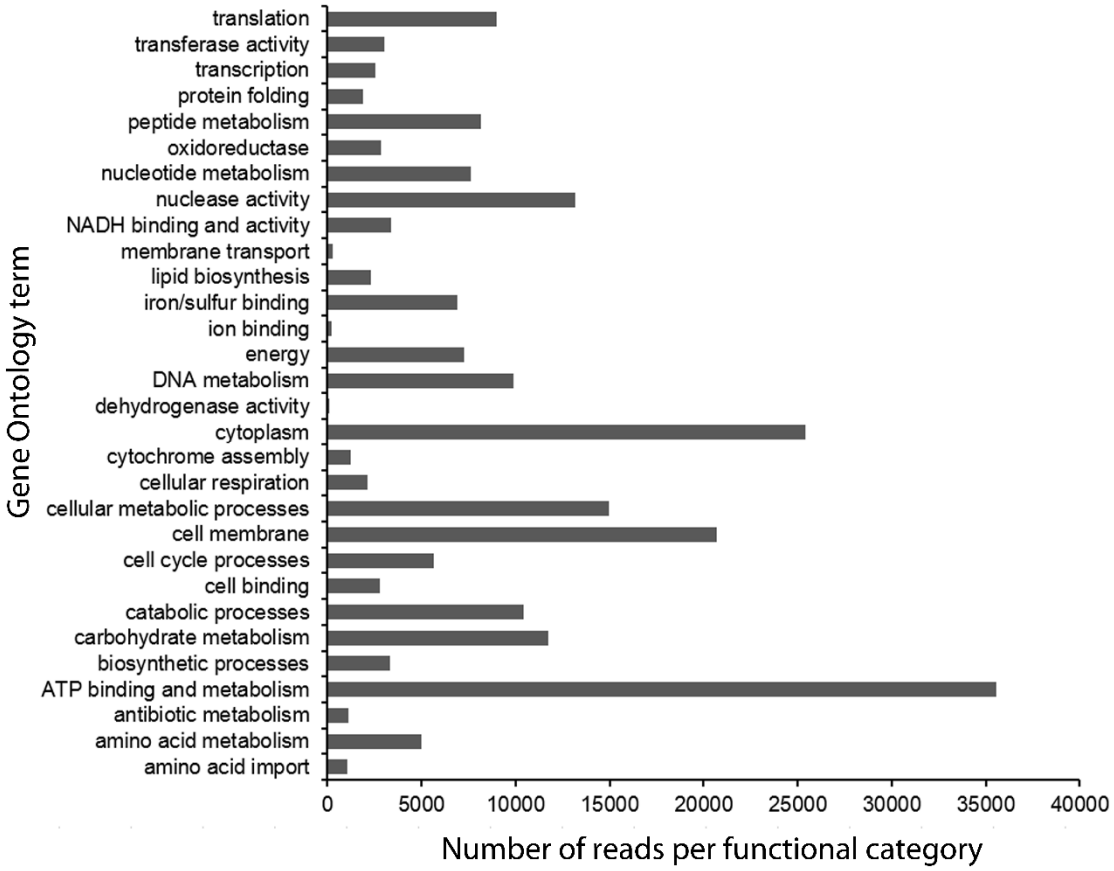

Figure 4C-4. Gene Ontology (GO) terms describing *Pedobacter* activity within leech mucosal secretions.

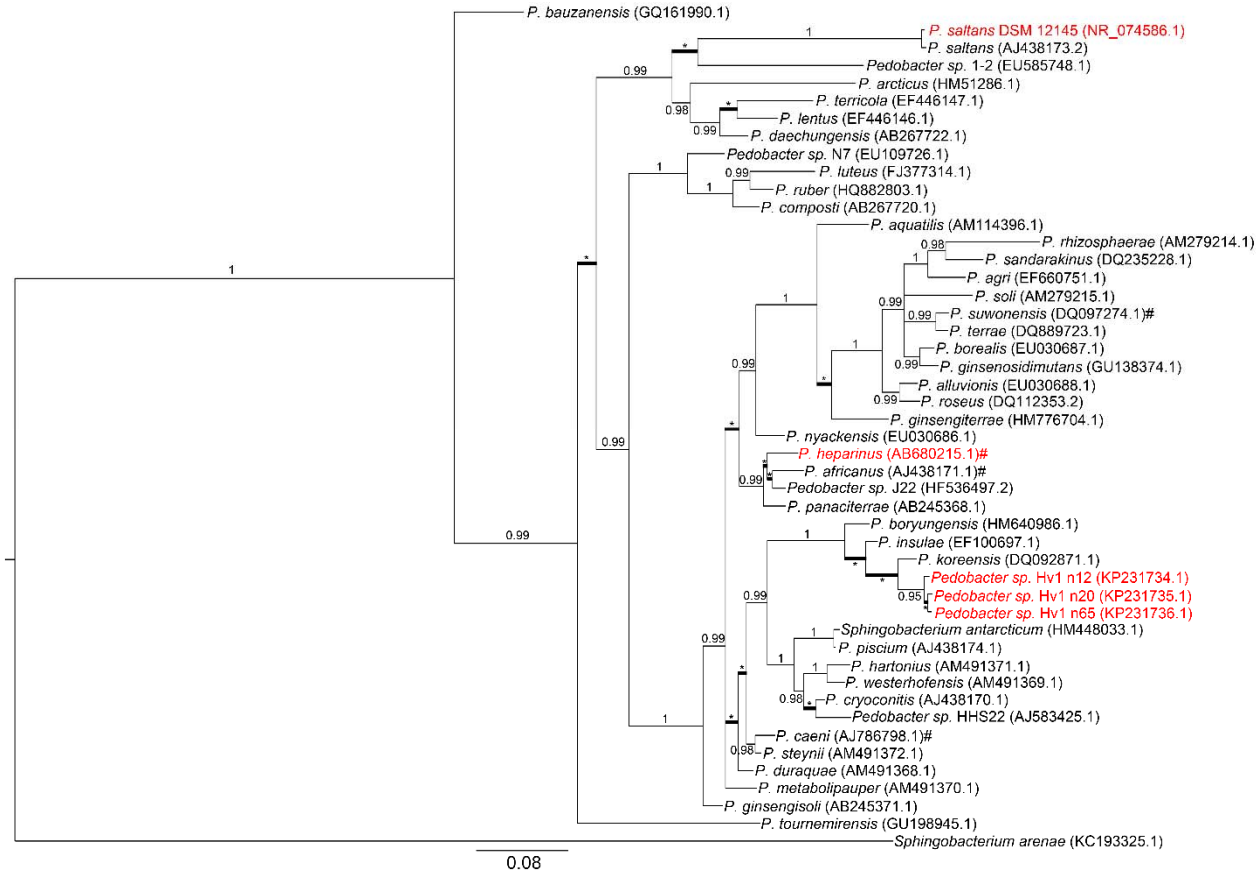

Figure 4C-5. *Pedobacter* 16S phylogenetic diversity. A Bayesian analysis tree created from approximately 1400 aligned nucleotides is shown. Significance values, represented in Bayesian posterior probability (PP), are indicated at respective nodes. Asterisks (*) and bold branches indicate a lack of support (i.e. Bayesian PP < 0.95). Pound sign (#) denotes species known to harbor heparinase-related genes. Species denoted in red were used in further comparative analyses discussed here. Outgroup sequence obtained from NCBI indicated with its accession number.

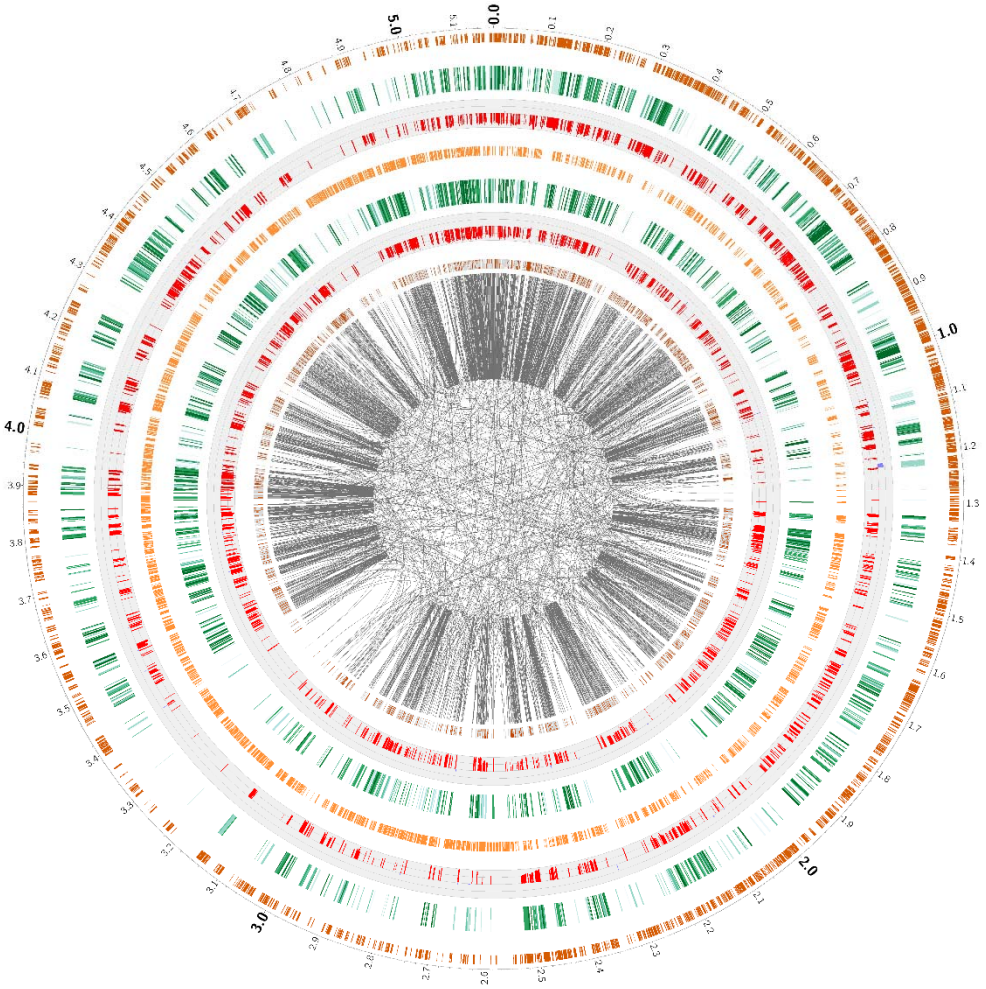

Figure 4C-6. Genome comparison of *Pedobacter heparinus* and *Pedobacter sp*. Hv1. Outer ring depicts forward coding sequences (CDS) of *P. heparinus* (dark orange) and ring 2 depicts forward CDS of *P. sp.* Hv1 (dark green). Ring 3 shows the comparative GC content of each forward CDS of *P. heparinus* and *P. sp.* Hv1, where blue indicates sequences in *P. sp.* Hv1 with a higher GC content and red indicates sequences in *P. heparinus* with a higher GC content. Ring 4 depicts the reverse CDS of *P. heparinus* (light orange) and ring 5 depicts the reverse CDS of *P. sp.* Hv1 (dark green). Ring 6 shows the comparative GC content of each reverse CDS of *P. heparinus* and *P. sp* Hv1 (see above). Ring 7 shows the placement of every *P. heparinus* CDS (forward and reverse) in brown, with the gray "hairball" depicting the corresponding synteny in the *P. sp.* Hv1 genome, where lines connect non-syntenic CDSs.

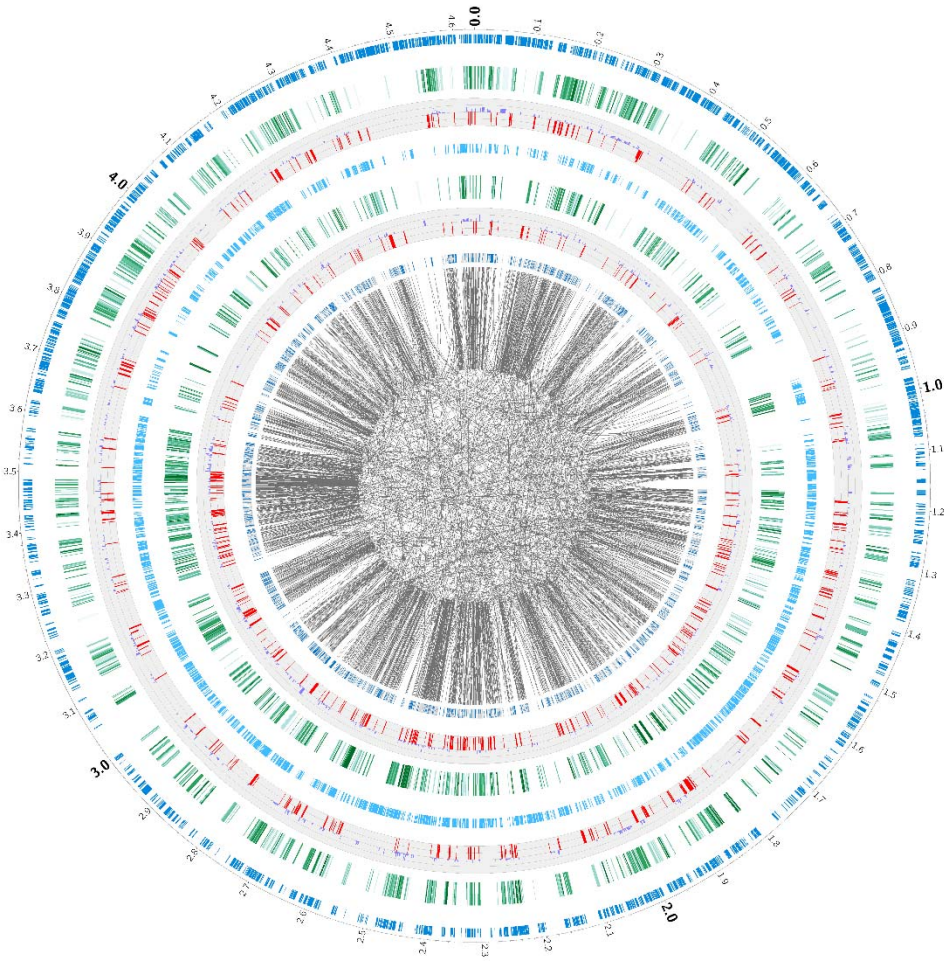

Figure 4C-7. Genome comparison of *Pedobacter saltans* and *Pedobacter sp*. Hv1. Outer ring depicts forward coding sequences (CDS) of *P. saltans* (dark blue) and ring 2 depicts forward CDS of *P. sp.* Hv1 (dark green). Ring 3 shows the comparative GC content of each forward CDS of *P. saltans* and *P. sp.* Hv1, where blue indicates sequences in *P. sp.* Hv1 with a higher GC content and red indicates sequences in *P. saltans* with a higher GC content. Ring 4 depicts the reverse CDS of *P. saltans* (light blue) and ring 5 depicts the reverse CDS of *P. sp.* Hv1 (dark green). Ring 6 shows the comparative GC content of each reverse CDS of *P. saltans* and *P. sp* Hv1 (see above). Ring 7 shows the placement of every *P. saltans* CDS (forward and reverse) in blue, with the gray "hairball" depicting the corresponding synteny in the *P. sp.* Hv1 genome, where lines connect non-syntenic CDSs

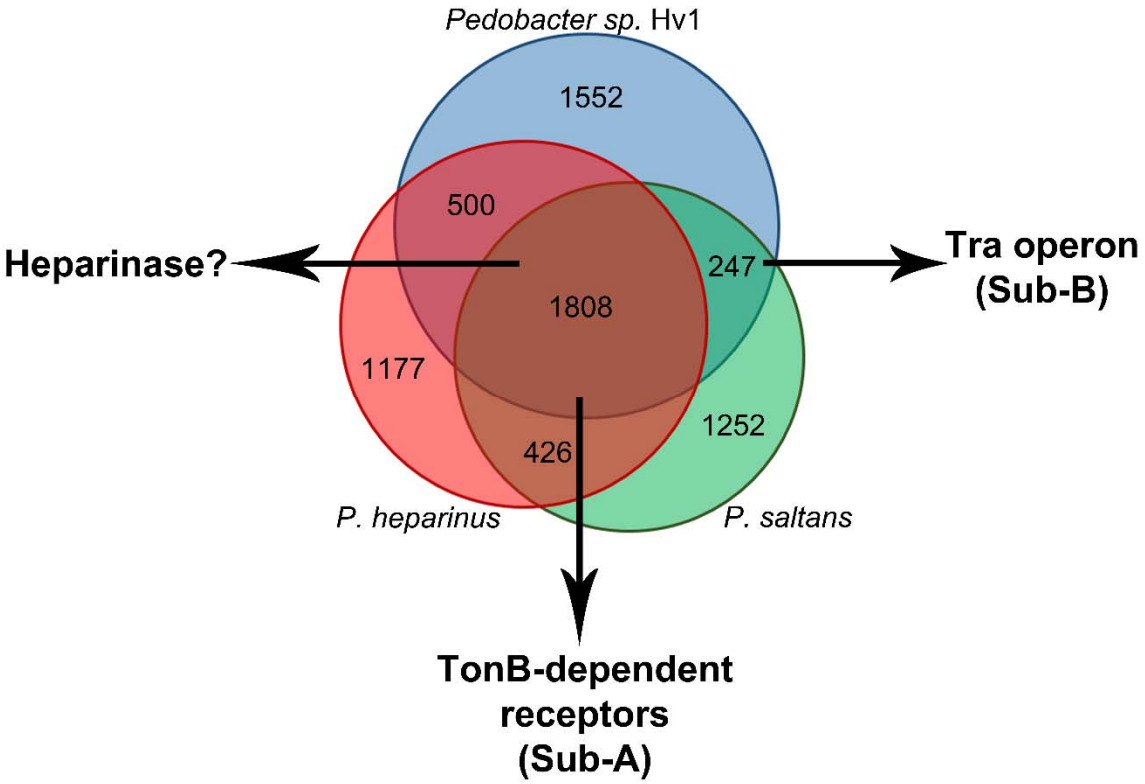

Figure 4C-8. Genomic content comparison of *Pedobacter* species. Three species, *Pedobacter heparinus* (red), *Pedobacter saltans* (green), and *Pedobacter sp*. Hv1 (blue) were compared for their genome content using OrthoMCL groups. Numerical values indicate the number of coding sequences (CDS) unique to each species, shared between two species, or shared between all species. Heparinase-related genes were not found within the *P. sp.* Hv1 genome in this study. See Figure 8(Sub-A) and Figure 8(Sub-B) for the TonB-dependent receptors and Tra operon, respectively.

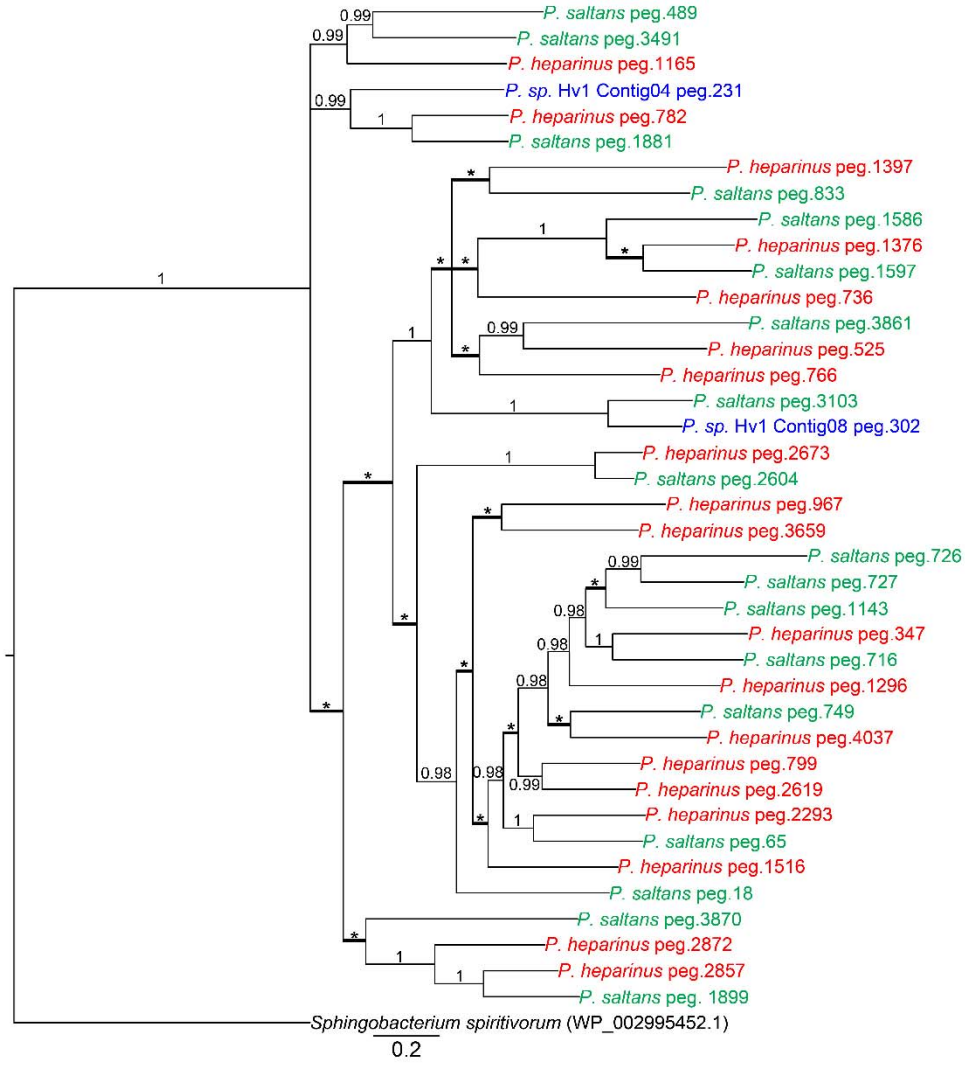

Figure 4C-8(Sub-A). TonB-dependent receptors compared between three *Pedobacter* species. A Bayesian analysis comparing ~1000 aligned amino acids is shown. Significance values, represented in Bayesian posterior probability (PP), are indicated at respective nodes. Asterisks (*) and bold branches indicate a lack of support (i.e. Bayesian PP < 0.95). Sequence identifiers in blue correspond with *Pedobacter sp.* Hv1, red correspond with *Pedobacter heparinus* and green correspond with *Pedobacter saltans*. A *Sphingobacterium* TonB-receptor amino acid sequence was used as an outgroup, depicted in black with corresponding NCBI accession number.

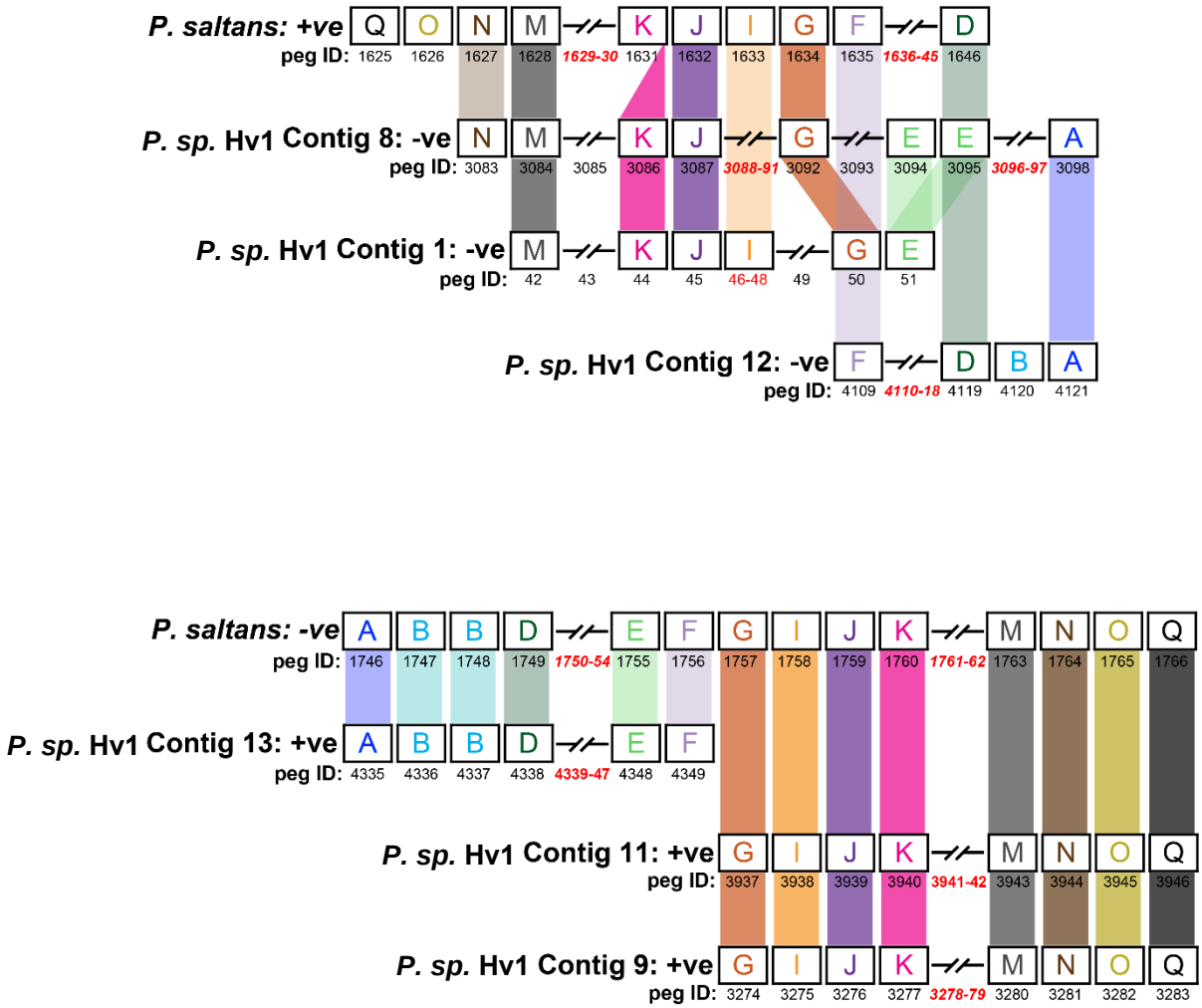

Figure 4C-8(Sub-B). Comparative analyses of Tra genes in *Pedobacter saltans* and *Pedobacter sp*. Hv1. Genes related to the conjugative transposon (Tra) operon were compared in the positive (+ve) and negative (-ve) strands of both *P. saltans* and *P. sp.* Hv1. Each coding sequence (CDS) identified by RAST is represented by a unique peg ID. As the genome of *P. sp.* Hv1 has not been completed, the contigs in which the Tra genes were located are denoted. Each gene is designated by a letter (e.g. *traA*=A) and a color (e.g. *traA*=dark blue). Additionally, respectively colored bands connect genes across contigs with transparent (i.e. fainter) colors indicating coverage between *P. saltans* and 1 *P. sp.* Hv1 contig, while opaque (i.e. darker) colors indicate coverage among *P. saltans* and 2 *P. sp.* Hv1 contigs. Triangles and diagonal lines indicate a shift or merge of Tra genes between species and/or contigs.

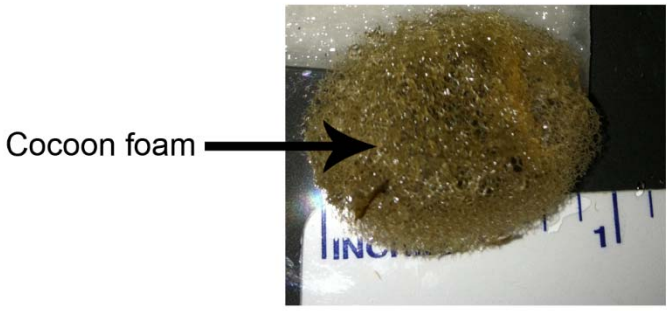

Figure 4C-S1. Medicinal leech cocoon. A ~1 week old cocoon produced by the Rio Lab leech colony. Arrow indicates protective foam produced by the leech "mother". DNA isolated from this foam shows the presence of *Pedobacter*.

Table 4C-1. Primer list and amplification settings.

Primer	Sequence (5'-3')	Amplicon size (bp)	Amplification conditions[a]
Pedobacter sp. Hv1 *gyrB* (PedogyrB)		748	30 sec at 55°C; 1 min at 72°C*
PedogyrB-F	GTT TGC ACG GTG TAG GGC TA		
PedogyrB-R	AAA GCT TCT CCT ACG GCC AC		

[a]Each amplification reaction was initiated with a 3 min denaturation step at 95°C. *40 cycles were used.

Table 4C-2. Next-generation sequencing and *de novo* assembly statistics

Illumina MiSeq data (raw)*	
Number of sequenced bases (Gbp)	1.064
Number of paired end reads	9,356,767
Average length of read (bp)	208 ± 40
Assembly (Mucus metatranscriptome)*	
Number of reads assembled	8,140,387
Total transcriptome size (bp)	77,274,024
Number of contigs generated	128,964
Average contig size (bp)	599 ± 755
N25 contig length (bp)	1,893
N50 contig length (bp)	797
N75 contig length (bp)	382
***Pedobacter*-specific**	
Number of reads assembled	645,328
Total transcriptome size (bp)	1,085,089
Number of contigs assembled	539
Average contig size (bp)	2013 ± 444

*Reprinted from Ott *et al*, (*in review*; Chapter 4A)

Table 4C-3. Functional role of tra-gene encoded proteins.

Functional Role	**Gene name***
Regulation	*finO, finP,* ***traJ***
Pilus biogenesis	***traA***, ***traB***, *traC*, ***traE***, ***traF***, ***traG***, *traH*, ***traK***, *traL*, ***traQ***, *traU, traV, traW, traX, trbC, trbI*
DNA metabolism	***traD***, ***traI***, ***traM***, *traY*
Aggregate stabilization	***traG***, ***traN***
Surface exclusion	*traS, traT*
Intracellular multiplication	***traO, traP***

*Genes shown in red are found in both *Pedobacter sp*. Hv1 and *Pedobacter saltans*, and genes shown in blue are found only in *P. sp*. Hv1.

Table 4C-S1. Genes expressed by *Pedobacter* in leech mucus.

Contig assignment	Gene role (closest Pedobacter)	Read count	Length	Length (kb)	RPKM	E value	Positive Match%
comp23146_c0_seq1	flagellar motor protein MotB [Pedobacter heparinus]	15000	1539	1.539	37940.70884	0	0.83
comp7716_c0_seq1	cold-shock protein [Pedobacter agri]	2158	332	0.332	25302.65873	1E-34	1
comp13564_c0_seq1	30S ribsomal protein S21 [Pedobacter haparinus]	2010	354	0.354	22102.71362	1E-23	0.98
comp13426_c0_seq1	50S ribsomal protein L20 [Pedobacter sp. B…	1934	686	0.686	10974.51042	5E-61	0.99
comp19500_c0_seq3	hypothectical protein [Pedobacter heparinus]	976	381	0.381	9971.893768	9E-10	0.9
comp6235_c0_seq1	TonB family protein [Pedobacter sp. V48]	7184	2919	2.919	9580.430616	1E-99	0.8
comp24497_c0_seq1	30S ribosomal protein S16 [Pedobacter sp. BAL39]	1686	707	0.707	9283.055731	7E-75	0.98
comp6376_c0_seq1	polynucleotide phosphorylase [Pedobacter bor……	5590	2566	2.566	8480.236362	0	0.98
comp13424_c0_seq1	50S ribsomal protein L19 [Pedobacter sp. V48]	1026	542	0.542	7368.869671	1E-70	0.97
comp24782_c0_seq1	glycoside hydrolase [Pedobacter borealis]	1268	684	0.684	7216.322821	6E-96	0.83
comp24468_c0_seq1	30S ribsomal protein S1 [Pedobacter sp. R20-19]	3402	1936	1.936	6840.404083	0	0.96
comp19739_c2_seq5	elongation factor Tu [Pedobacter sp. V48]	2658	1641	1.641	6305.20479	0	0.99
comp23338_c0_seq1	ATP-dependent Clp protease ClpC [Pedobacter…..	2626	1658	1.658	6165.424683	0	0.85
comp24962_c0_seq1	50S ribsomal protein L21 [Pedobacter agri]	1322	869	0.869	5921.946505	6E-48	0.94
comp6264_c0_seq1	RNA polymerase sigma factor rpoD [Pedobacter….	1592	1057	1.057	5863.013273	0	1
comp8309_c0_seq1	ferredoxin [Pedobacter sp. BAL39]	664	452	0.452	5718.504218	3E-55	0.97
comp19739_c2_seq3	elongation factor Tu [Pedobacter sp. V48]	2500	1898	1.898	5127.392949	0	0.99
comp20180_c1_seq1	hypothectical protein [Pedobacter heparinus]	994	873	0.873	4432.257075	6E-151	0.94
comp25597_c0_seq1	50S ribsomal protein L9 [Pedobacter sp. V48]	1178	1150	1.15	3987.495917	1E-81	0.98
comp18574_c0_seq1	DNA-binding protein [Pedobacter sp. R20-19]	478	471	0.471	3950.570266	9E-56	0.94
comp13461_c0_seq1	elongation factor G [Pedobacter agri]	3356	3456	3.456	3780.080248	0	0.98
comp22414_c0_seq1	hypothetical protein N824_03015 [Pedobacter sp. V48]	934	1041	1.041	3492.600791	2E-54	0.79
comp21164_c0_seq1	preprotein translocase subunit SecY [Pedobacter……	6528	7607	7.607	3340.561955	0	0.92
comp21164_c1_seq2	DNA-directed RNA polymerase subunit alpha [Pedobacter…….	3266	3927	3.927	3237.487352	0	0.98
comp17549_c1_seq1	50S ribsomal protein L25 [Pedobacter heparinus]	598	727	0.727	3201.987074	3E-96	0.92
comp19007_c0_seq1	ribnuclease G [Pedobacter sp.V48]	2460	3070	3.07	3119.245325	0	0.97
comp21249_c0_seq1	molecular chaperone DnaK [Pedobacter heparinus]	1608	2016	2.016	3104.905008	0	0.96
comp19714_c0_seq1	50S risomal protein L13 [Pedobacter sp. R20….	992	1253	1.253	3081.863522	8E-91	0.97

comp26823_c0_seq1	chemotaxis protein CheY [Pedobacter sp. V48]	1660	2148	2.148	3008.33788	0	0.97
comp26543_c0_seq1	elongation factor P [Pedobacter heparinus]	554	735	0.735	2934.102132	2E-126	0.98
comp27336_c0_seq1	orotate phosphoribosyltransferase [Pedobacter sp…	884	1197	1.197	2874.82171	1E-141	0.95
comp13755_c0_seq1	hypothetical protein [Pedobacter sp. BAL 39]	764	1061	1.061	2803.049556	2E-69	0.95
comp22152_c0_seq1	DNA polymerase III subunit beta [Pedobacter….	2082	3065	3.065	2644.253255	0	0.99
comp6237_c0_seq1	50S ribsomal protein L2 [Pedobacter heparinus]	2480	3886	3.886	2484.286537	5E-164	0.99
comp13033_c0_seq1	protease Do [Pedobacter heparius]	1124	1802	1.802	2428.087459	3E-150	0.8
comp6016_c0_seq1	cupin [Pedobacter sp. R20-19]	824	1357	1.357	2363.742508	0	0.96
comp13707_c0_seq1	metallo-beta-lactamase [Pedobacter sp.V48]	1244	2051	2.051	2361.062705	8E-135	0.91
comp27115_c0_seq1	2-oxglutarate dehydrogenase E1 [Pedobacter sp. V48]	2598	4299	4.299	2352.472216	0	0.94
comp19691_c0_seq1	hypothetical protein [Pedobacter agri]	506	863	0.863	2282.404014	1E-27	0.76
comp6641_c0_seq1	adenosy lhomocysteinase [Pedobacter borealis]	628	1137	1.137	2150.066935	0	0.98
comp19783_c0_seq1	elongation factor Tu [Pedobacter arctius]	176	323	0.323	2121.108805	4E-38	0.92
comp17314_c0_seq1	ribonucleoside-diphosphate reductase [Pedobacter……	1550	2862	2.862	2108.214859	0	0.86
comp17549_c0_seq1	ribose-phosphate pyrophosphokinase [Pedobacter sp….	644	1214	1.214	2064.999648	0	0.96
comp13753_c0_seq1	acetyl-CoA carboxylase subunit alpha [Pedobacter…..	2584	4952	4.952	2031.256063	0	0.98
comp5956_c0_seq1	preprotein translocase subunit SecA [Pedobacter	1794	3477	3.477	2008.494049	0	0.79
comp22075_c1_seq1	aminopetidase [Pedobacter agri]	3108	6091	6.091	1986.301689	5E-160	0.79
comp21764_c2_seq4	DNA-directed RNA polymerase subunit beta' [Pedobacter	5828	11555	11.555	1963.371102	0	0.98
comp18991_c0_seq1	3-oxyacyl-ACP synthase [Pedobacter heparinus]	2160	4350	4.35	1932.935202	0	0.97
comp17479_c0_seq1	hypothetical protein N824_08775 [Pedobacter sp. V48]	2860	5761	5.761	1932.506481	0	0.81
comp21164_c1_seq1	DNA-directed RNA polymerase subunit alpha [Pedobacter…….	1746	3566	3.566	1905.968426	0	0.98
comp19714_c1_seq1	30S ribsomal protein S2 [Pedobacter borealis]	284	582	0.582	1899.538747	5E-84	0.91
comp17513_c0_seq1	molecular chaperone GroEL [Pedobacter borealis]	1012	2095	2.095	1880.39586	0	0.98
comp27804_c0_seq1	ribnucleoside-diphosphate reductase [Pedobacter…..	2320	4844	4.844	1864.389514	0	0.96
comp6107_c0_seq1	cytochrome oxidase subunit III [Pedobacter h…..	906	1895	1.895	1861.108894	2E-132	0.92
comp21764_c2_seq3	DNA-directed RNA polymerase subunit beta' [Pedobacter	4908	10632	10.632	1796.976457	0	0.98
comp21623_c2_seq1	peptodylprolyl isomerase [Pedobacter heparinus]	1910	4167	4.167	1784.278605	0	0.93
comp16485_c0_seq1	amino acid permase [Pedobacter sp. V48]	818	1788	1.788	1780.896131	0	0.9
comp19748_c0_seq1	F0F1 ATP synthase subunit beta [Pedobacter heparinus]	1148	2543	2.543	1757.309793	0	0.99
comp6482_c0_seq1	chromosal replication intiation protien [Pedobacter…..	698	1573	1.573	1727.346647	0	0.99

comp21730_c0_seq1	molybdopterin oxidoreductase [Pedobacter hep...	6220	14114	14.114	1715.509285	0	0.9
comp18133_c0_seq1	succinate dehydrogenase [Pedobacter heparinus]	530	1240	1.24	1663.822472	0	0.96
comp18965_c0_seq1	sterol desaturase [Pedobacter sp. V48]	1366	3203	3.203	1660.147065	0	0.83
comp23344_c0_seq1	thioredexin [Pedobacter heparinus]	286	684	0.684	1627.656409	4E-121	0.97
comp19506_c0_seq1	F0F1 ATP synthase subunit alpha [Pedobacter sp. V48]	2220	5376	5.376	1607.48347	0	0.99
comp13233_c0_seq1	hypothectical protein [Pedobacter heparinus]	518	1301	1.301	1549.905661	1E-99	0.97
comp19478_c0_seq1	glutamine amidotransferase [Pedobacter sp. B…..	1684	4241	4.241	1545.705015	0	0.97
comp19331_c0_seq1	preprotein translocase subunit SecD [Pedobacter	1428	3633	3.633	1530.085187	0	0.84
comp20643_c0_seq1	hypothetical protein N824_09530 [Pedobacter sp. V48]	420	1086	1.086	1505.470557	1E-105	0.85
comp13652_c1_seq1	adenylosuccinate synthetase [Pedobacter heparinus]	700	1878	1.878	1450.959376	0	0.96
comp13016_c0_seq1	glycine dehydrogenase [Pedobacter sp. V48]	1112	2992	2.992	1446.758356	0	0.95
comp16559_c0_seq1	type II citrate synthase [Pedobacter heparinus]	510	1374	1.374	1444.894855	0	0.95
comp20161_c0_seq1	aconitate hydratase [Pedobacter heparinus]	902	2434	2.434	1442.576207	0	0.97
comp20744_c0_seq1	1-deoxy-D-xylulose-5-phosphate synthase [Pedobacter……	3116	8427	8.427	1439.385941	0	0.94
comp6531_c1_seq1	translation intiatiom factor IF-2 [Pedobacter….	1832	4978	4.978	1432.594826	0	0.84
comp19496_c0_seq2	thioredoxin [Pedobacter heparinus]	158	433	0.433	1420.437051	5E-64	0.99
comp19719_c0_seq1	phosphopantothenoylcysteine decarboxylase [Pedobacter…	1832	5055	5.055	1410.772907	0	0.88
comp27500_c0_seq1	dehydrogenase [Pedobacter sp. BAL39]	962	2669	2.669	1403.069873	0	0.91
comp22096_c0_seq1	ribonuclease G [Pedobacter boralis]	1242	3517	3.517	1374.681312	0	0.77
comp19496_c0_seq1	thioredoxin [Pedobacter heparinus]	336	969	0.969	1349.796512	9E-65	0.99
comp36549_c0_seq1	thymidylate synthase [Pedobacter heparinus]	286	840	0.84	1325.377362	0	0.97
comp21421_c0_seq1	DNA polymerase III subunit gamma/tau [Pedobacter…..	792	2358	2.358	1307.477374	0	0.81
comp13426_c1_seq1	translation intiatiom factor IF-3 [Pedobacter….	202	617	0.617	1274.438864	2E-91	1
comp6459_c0_seq1	glycyl-tRNA sythetease [Pedobacter sp.V48]	520	1598	1.598	1266.716332	0	0.97
comp13652_c0_seq1	MFS transporter [Pedobacter heparinus]	782	2422	2.422	1256.85569	0	0.97
comp13010_c0_seq1	MultiSpecies: decarboxylase [Pedobacter]	598	1865	1.865	1248.17405	0	0.99
comp19739_c2_seq2	elongation factor Tu [Pedobacter borealis]	900	2814	2.814	1245.00535	0	0.9
comp6550_c0_seq1	beta-lactamase [Pedobacter herparinus]	508	1609	1.609	1229.024299	4E-180	0.87
comp18400_c0_seq1	chromosome partitioning protein ParA [Pedobacter…..	1040	3318	3.318	1220.140264	8E-171	0.98
comp25659_c0_seq1	seryl-tRNA synthetase [Pedobacter borealis]	408	1303	1.303	1218.901324	0	0.93
comp18365_c0_seq1	leucine dehydrogenase [Pedobacter heparinus]	922	2970	2.97	1208.446068	0	0.94

comp26589_c0_seq1	protease [Pedobacter heparinus]	276	937	0.937	1146.627339	0	0.96
comp24784_c0_seq1	pyruvate dehydrogenase [Pedobacter borealis]	786	2691	2.691	1137.003102	0	0.83
comp16819_c0_seq1	methionyl-tRNA formyltransferase [Pedobacter……	502	1737	1.737	1125.010822	2E-173	0.89
comp6312_c0_seq1	isoleucyl-tRNA synthase [Pedobacter sp. BAL3……	144	499	0.499	1123.349116	6E-92	0.98
comp6316_c0_seq1	transketolase [Pedobacter sp. V48]	704	2475	2.475	1107.261647	0	0.92
comp14150_c0_seq1	GTP-binding protein TypA [Pedobacter sp. R20…..	520	1847	1.847	1095.946236	0	0.98
comp21040_c0_seq1	ABC transporter [Pedobacter sp. V48]	630	2238	2.238	1095.804977	4E-160	0.96
comp20737_c0_seq1	glutamate dehydrogenase [Pedobacter sp. V48]	944	3386	3.386	1085.270109	0	0.95
comp5533_c0_seq1	glutamine synthetase [Pedobacter sp. BAL39]	642	2308	2.308	1082.809419	0	0.95
comp17542_c0_seq1	pyruvate kinase [Pedobacter heparinus]	476	1715	1.715	1080.4275	0	0.96
comp13100_c0_seq2	ATP phosoribosyltransferase [Pedobacter sp. V48]	282	1023	1.023	1073.065608	3E-170	0.92
comp21040_c1_seq1	S-adensosylmethionine tRNA ribosyltransferase [Pedobacter…….	604	2211	2.211	1063.41063	0	0.97
comp20441_c0_seq1	electron transfer flavoprotein subunit alpha [Pedobacter…..	1034	3790	3.79	1062.023508	2E-176	0.94
comp20550_c1_seq1	ribsomal protein S12 methylthiothransferase [Pedobacter…..	1710	6308	6.308	1055.254534	0	0.98
comp33597_c0_seq1	hypothetical protein [Pedobacter agri]	390	1445	1.445	1050.629428	0	0.97
comp6655_c0_seq1	RNA polymerase sigma54 factor [Pedobacter he…….	552	2046	2.046	1050.234425	0	0.95
comp21186_c0_seq1	carbamoyl phosphate synthase large subunit [Pedobacter……	818	3100	3.1	1027.17493	0	0.94
comp25479_c0_seq1	DNA gyrase subunit A [Pedobacter sp. BAL39]	588	2274	2.274	1006.559998	0	0.96
comp16682_c0_seq1	hypothetical protein [Pedobacter sp. R20-19]	1018	3946	3.946	1004.253834	0	0.89
comp20981_c0_seq1	protease 2 [Pedobacter borealis]	1600	6238	6.238	998.4525109	0	0.93
comp21040_c1_seq2	S-adensosylmethionine tRNA ribosyltransferase [Pedobacter…….	550	2164	2.164	989.3688539	0	0.97
comp5420_c0_seq1	Cell divison protein FtsA [Pedobacter hepar…	1628	6467	6.467	979.9509559	0	0.97
comp21555_c0_seq1	AMP-binding protein [Pedobacter borealis]	1204	4789	4.789	978.6658884	0	0.92
comp20216_c0_seq1	malate dehydrogenase [Pedobacter heparinus	858	3416	3.416	977.7373981	0	0.94
comp14206_c0_seq1	DNA helicase [Pedobacter sp. V48]	526	2123	2.123	964.4696177	0	0.94
comp5301_c0_seq1	molecular chaperone DnaK [Pedobacter gucosi…	484	2016	2.016	934.5609603	0	0.88
comp36943_c0_seq1	2-amino-3-ketobutyrate CoA ligase [Pedobacter……	338	1416	1.416	929.193682	0	0.98
comp21764_c2_seq7	DNA-directed RNA polymerase subunit beta' [Pedobacter	1090	4588	4.588	924.8171823	0	0.85
comp24920_c0_seq1	glyceraldehyde-3-phosphate dehydrogenase [Pedobacter…….	606	2552	2.552	924.3676867	0	0.95

comp12568_c0_seq1	hypothectical protein [Pedobacter heparinus]	814	3428	3.428	924.3498879	0	0.78
comp18625_c0_seq1	GTP-binding protein Lep4 [Pedobacter agri]	452	1905	1.905	923.626226	0	0.99
comp27600_c0_seq1	Fe-S oxidoreductase [Pedobacter borealis]	338	1425	1.425	923.3250903	0	0.91
comp21022_c0_seq1	CTP synthetase [Pedobacter sp. V48]	826	3551	3.551	905.4869097	0	0.96
comp22309_c0_seq9	preprotein translocase subunit SecA [Pedobacter s…..	904	3906	3.906	900.9257351	0	0.94
comp19662_c0_seq1	hypothectical protein [Pedobacter heparinus]	720	3163	3.163	886.1068743	0	0.87
comp12464_c0_seq2	isocitrate dehydrogenase [Pedobacter heparinus]	262	1175	1.175	867.9930064	0	0.94
comp16963_c0_seq2	diapophytoene dehydrogenase [Pedobacter borealis]	400	1800	1.8	865.0481616	0	0.85
comp31408_c0_seq1	IMP cyclohydrolase [Pedobacter sp. V48]	374	1709	1.709	851.8876863	0	0.92
comp20659_c0_seq1	cytochrome C biogensis protein [Pedobactersp. V48]	1232	5693	5.693	842.407695	0	0.91
comp17500_c0_seq1	oxidoreductase [Pedobacter sp. BAL39]	584	2756	2.756	824.8717593	4E-173	0.97
comp14059_c1_seq1	malate--CoA ligase subunit beta [Pedobacter…..	192	907	0.907	824.037058	0	0.98
comp19506_c0_seq2	F0F1 ATP synthase subunit gamma [Pedobacter….	338	1618	1.618	813.1880431	0	0.96
comp5589_c0_seq1	transcription elongation factor GreA [Pedbacter….	216	1035	1.035	812.3930561	1E-91	0.96
comp5557_c0_seq1	ATPase AAA [Pedobacter sp.V48]	1060	5099	5.099	809.2331301	0	0.95
comp18133_c1_seq1	fumarate reductase [Pedobacter sp. R20-19]	132	638	0.638	805.3896677	7E-103	0.97
comp23538_c0_seq1	sugar kinase [Pedobacter sp. R20-19]	412	1995	1.995	803.9094193	0	0.95
comp16505_c0_seq1	transcription termination factor Rho [Pedobacter…..	624	3062	3.062	793.2904107	0	0.98
comp18519_c0_seq1	sodium:solute symporter [Pedobacter borealis	456	2275	2.275	780.2544297	0	0.92
comp16836_c0_seq2	DEAD/DEAH box helicase [Pedobacter heparinus]	314	1568	1.568	779.5363854	0	0.97
comp19004_c0_seq1	methlmalonyl-CoA mutase [Pedobacter agri]	1092	5519	5.519	770.2204504	0	0.95
comp6013_c0_seq1	peptidase M41 [Pedobacter saltans]	310	1567	1.567	770.0971189	0	0.96
comp16690_c0_seq1	transaldolase [Pedobacter heparinus]	294	1489	1.489	768.6089441	9E-126	0.95
comp18981_c0_seq1	ABC transporter ATP-binding protein [Pedobacter….	374	1902	1.902	765.4448243	0	0.96
comp31124_c0_seq1	amidophosphoribosyltransferase [Pedobacter sp. V48]	462	2387	2.387	753.4290439	0	0.95
comp24311_c0_seq1	threonyl-tRNA synthetase [Pedobacter sp. R20-19]	382	1997	1.997	744.6258336	0	0.97
comp16473_c1_seq1	Lon protease [Pedobacter sp. V48]	688	3605	3.605	742.9096	0	0.96
comp19208_c0_seq1	fructose-biphosphate aldolase [Pedobacter agri]	682	3665	3.665	724.3745724	0	0.97
comp18001_c0_seq1	heat shock protein Hsp90 [Pedobacter heparinus]	442	2382	2.382	722.3261097	0	0.92
comp21553_c0_seq1	preprotein translocase subunit SecY [Pedobacter……	1436	7853	7.853	711.8223889	0	0.97
comp21553_c0_seq2	preprotein translocase subunit SecY [Pedobacter……	1428	7815	7.815	711.2987186	4E-166	0.77

comp22103_c0_seq1	UDP-glucose 6-dehydrogenase [Pedobacter sp.R20-19]	460	2528	2.528	708.3266196	0	0.83
comp5284_c0_seq1	NADH dehydrogenase subunit D [Pedobacter hep..	712	3943	3.943	702.9201901	0	0.97
comp17474_c0_seq1	hypothetical protein [Pedobacter sp. BAL 39]	1080	6020	6.02	698.3611404	0	0.86
comp20730_c0_seq2	penicillin-binding protein [Pedobacter sp. R…..	1046	5840	5.84	697.2228932	0	0.8
comp21235_c0_seq1	NADH:ubiquinone oxidoreductase subunit L [Pedobacter….	2342	13186	13.186	691.3956146	0	0.9
comp21616_c0_seq1	GMC family oxidoreductase [Pedobacter sp. BAL…….	1892	10697	10.697	688.5126715	0	0.94
comp34107_c0_seq1	cystahionine beta-synthase [Pedobacter heparinus]	354	2005	2.005	687.2926291	0	0.96
comp5552_c0_seq1	thrioredexin reductase [Pedobacter sp. V48]	278	1579	1.579	685.3548132	0	0.94
comp32131_c0_seq1	gliding motility protein [Pedobacter heparinus]	678	3948	3.948	668.5060641	0	0.88
comp16579_c0_seq1	hypothetical protein N824_07735 [Pedobacter sp. V48]	254	1489	1.489	664.0362986	0	0.77
comp14526_c0_seq1	hypothectical protein [Pedobacter heparinus]	668	3925	3.925	662.5056748	0	0.82
comp18780_c0_seq1	DNA mismatch protein MutL [Pedobacter sp….	2100	12343	12.343	662.2948332	0	0.85
comp20163_c0_seq1	transporter [Pedobacter sp. BAL39]	1228	7310	7.31	653.9338086	0	0.91
comp19152_c0_seq1	ATP-dependent DNA helicase [Pedobacter sp. V48]	906	5493	5.493	642.0537693	0	0.94
comp22075_c0_seq1	membrane protein [Pedobacter agri]	498	3114	3.114	622.5346596	0	0.84
comp8354_c0_seq1	transmembrane HD family protein [Pedobacter….	354	2232	2.232	617.3932443	0	0.93
comp14393_c0_seq1	aspartyl-tRNA synthetase [Pedobacter sp. V48]	316	1997	1.997	615.9732027	0	0.96
comp21339_c0_seq1	acetolactate synthase [Pedobacter sp. V48]	854	5410	5.41	614.4880009	0	0.92
comp6155_c0_seq1	GTP-binding protein Der [Pedobacter sp. BAL3…..	402	2610	2.61	599.5678637	0	0.97
comp21335_c0_seq1	GTP-binding protein TypA [Pedobacter sp. R20…..	142	927	0.927	596.2953347	6E-161	0.85
comp7676_c0_seq1	transcription-repair coupling factor [Pedobacter….	420	2756	2.756	593.2296899	0	0.96
comp23759_c0_seq1	radical SAM protein [Pedobacter sp. V48]	222	1459	1.459	592.3119352	0	0.95
comp14331_c0_seq1	TonB-dependent receptor [Pedobacter heparinus]	794	5231	5.231	590.8654332	0	0.88
comp5650_c0_seq1	cell division protein FtsK [Pedobacter hepar….	320	2112	2.112	589.8055647	0	0.9
comp24049_c0_seq1	aspartate aminotransferase [Pedobacter borealis]	188	1260	1.26	580.8180513	0	0.95
comp35680_c0_seq1	membrane protein [Pedobacter borealis]	1174	7950	7.95	574.8489858	0	0.86
comp18533_c0_seq1	hypothetical protein N824_03605 [Pedobacter sp. V48]	162	1104	1.104	571.2138675	9E-60	0.72
comp6245_c0_seq1	hypothetical protein [Pedobacter borealis]	182	1246	1.246	568.5990725	3E-60	0.96
comp20350_c0_seq1	membrane protein [Pedobacter heparinus]	728	5005	5.005	566.2133421	0	0.9
comp32377_c0_seq1	sterol desaturase [Pedobacter heparinus]	214	1488	1.488	559.8396368	0	0.87
comp18908_c0_seq1	23S rRNA pseudourine synthase F [Pedobacter sp…	296	2063	2.063	558.52843	3E-161	0.79

comp20350_c0_seq4	membrane protein [Pedobacter heparinus]	1072	7683	7.683	543.14621	0	0.9
comp13080_c1_seq1	dTDP-4-dehydrorhamnose reductase	374	2702	2.702	538.8142324	2E-170	0.89
comp28466_c0_seq1	argininosuccinate synthase [Pedobacter agri]	486	3541	3.541	534.2728973	0	0.98
comp14523_c0_seq1	glucose-6-phosphate dehydrogenase [Pedobacter…….	1064	7842	7.842	528.1625348	0	0.91
comp18573_c0_seq1	phenylalanyl-tRNA sythenase subunit beta [Pedobacter……	802	5963	5.963	523.5550587	0	0.9
comp16664_c1_seq1	2-methylthioadenine synthetase [Pedobacter h…	650	4849	4.849	521.8118937	0	0.96
comp18848_c0_seq3	oxidoreductase [Pedobacter borealis]	354	2643	2.643	521.3854413	0	0.97
comp16505_c0_seq2	transcription termination factor Rho [Pedobacter…	264	1977	1.977	519.8164977	2E-167	0.98
comp19102_c0_seq1	signal recoginition particle [Pedobacter sp. V48.]	616	4614	4.614	519.703837	0	0.97
comp17653_c1_seq1	tryptophan 2, 3-dioxygenase [Pedobacter agri]	162	1227	1.227	513.9528197	0	0.95
comp41239_c0_seq1	ABC transporter [Pedobacter sp. V48]	222	1716	1.716	503.6032129	8E-118	0.86
comp16369_c0_seq3	phosphoribosyltransferase [Pedobacter borealis]	124	965	0.965	500.2040147	2E-81	0.95
comp16939_c0_seq1	methinoyl-tRNA synthetase [Pedobacter sp. BA….	154	1201	1.201	499.1493555	7E-162	0.9
comp51530_c0_seq1	ATP synthase subunit beta [Pedobacter borealis]	74	579	0.579	497.5147458	3E-76	0.87
comp16729_c0_seq1	7-carboxy-7-dezaguanine synthase [Pedobacter sp….	112	877	0.877	497.1314406	3E-139	0.97
comp21616_c0_seq2	GMC family oxidoreductase [Pedobacter sp. BAL…….	714	5753	5.753	483.1218048	0	0.94
comp38550_c0_seq1	glucose-1-phosphate thymidylyltransferase [Pedobacter……	378	3180	3.18	462.7191581	0	0.82
comp19605_c0_seq3	flagellar motor protein MotB [Pedobacter heparinus]	86	734	0.734	456.0948754	6E-103	0.79
comp19792_c0_seq1	phosphoribosylaminoimidazole carboxylase [Pedobacter……..	290	2489	2.489	453.5507637	0	0.95
comp19066_c0_seq1	hypothectical protein [Pedobacter heparinus]	868	7550	7.55	447.5335257	0	0.78
comp13792_c0_seq1	peptide transporter [Pedobacter borealis]	644	5624	5.624	445.7520576	0	0.92
comp6253_c0_seq1	patatin [Pedobacter borealis]	754	6693	6.693	438.5340523	0	0.88
comp42249_c0_seq1	serine protease [Pedobacter sp. BAL39]	176	1565	1.565	437.7751719	0	0.84
comp19323_c0_seq1	hypothetical protein [Pedobacter agri]	570	5136	5.136	432.0187956	0	0.86
comp20313_c0_seq2	cell division protein [Pedobacter heparinus]	662	6037	6.037	426.8640837	0	0.89
comp8474_c0_seq1	30S ribsomal protein S10 [Pedobacter ba…	28	256	0.256	425.765892	1E-24	1
comp20207_c0_seq1	GH3 auxin-responsive promoter [Pedobacter heparinus]	856	7870	7.87	423.4009553	0	0.93
comp50474_c0_seq1	peptide chain release factor 2 [Pedobacter heparinus]	238	2198	2.198	421.504359	5E-126	0.94
comp29149_c0_seq1	nucleoside permease [Pedobacter agri]	268	2497	2.497	417.8005938	0	0.95
comp102682_c0_seq1	FeS cluster assembly scaffold IscU [Pedobacter……	26	243	0.243	416.5046704	7E-47	0.99
comp6984_c0_seq1	N-acetylmuramoyl-L-alanine amidase [Pedobacter sp……	102	956	0.956	415.331701	2E-158	0.9

comp50823_c0_seq1	MarR family transcriptional regulator [Pedobacter……	90	846	0.846	414.1188007	1E-122	0.9
comp15872_c0_seq1	ATPase AAA [Pedobacter heparinus]	142	1347	1.347	410.3680588	0	0.81
comp57198_c0_seq1	glucose-6-phosphate isomerase [Pedobacter heparinus]	294	2822	2.822	405.5488015	0	0.91
comp20731_c0_seq1	glyceraldehyde-3-phosphate dehydrogenase [Pedobacter…….	416	4086	4.086	396.3216247	5E-127	0.88
comp20207_c0_seq2	ketol-acid reductoisomerase [Pedobacter borealis]	526	5218	5.218	392.4049441	0	0.98
comp20350_c0_seq3	membrane protein [Pedobacter heparinus]	792	7940	7.94	388.2911395	0	0.9
comp18191_c0_seq1	elongation factor Tu [Pedobacter sp. BAL39]	92	926	0.926	386.749394	0	0.97
comp25830_c0_seq1	beta-carotene 15,15-monooxygenase [Pedobacter sp..	116	1170	1.17	385.9445644	9E-155	0.89
comp7292_c0_seq1	phosphorine phosphatase [Pedobacter heparinus]	240	2446	2.446	381.9509462	0	0.94
comp15000_c0_seq1	ATP-dependent Clp protease ClpC [Pedobacter…	30	306	0.306	381.6388948	2E-21	1
comp13477_c0_seq1	C4-dicarboxylate ABC transporter [Pedobacter sp….	238	2443	2.443	379.2331482	0	0.91
comp6067_c0_seq1	NUDIX hydrolase [Pedobacter heparius]	70	729	0.729	373.7862426	2E-140	0.94
comp21820_c0_seq1	UDP-N-acetylmuramate--alanine ligase [Pedobacter..	564	5876	5.876	373.637208	6E-143	0.73
comp26513_c0_seq1	DNA polymerase III subunit alpha [Pedobacter…….	454	4733	4.733	373.3981395	0	0.92
comp7049_c0_seq1	hypothectical protein [Pedobacter heparinus]	362	3786	3.786	372.2037652	0	0.73
comp6043_c0_seq1	pseudoridine synthase [Pedobacter agri]	166	1760	1.76	367.153964	1E-135	0.93
comp19386_c1_seq4	hypothetical protein [Pedobacter borealis]	104	1107	1.107	365.7114179	5E-108	0.9
comp19594_c1_seq1	glycyl-tRNA sythetease [Pedobacter heparinus]	86	926	0.926	361.5266075	3E-162	0.98
comp17352_c0_seq2	acyl-CoA dehydrogenase [Pedobacter heparinus]	184	1995	1.995	359.0275077	0	0.96
comp57759_c0_seq1	alkyl hydroperoxide reductase [Pedobacter heparinus]	222	2413	2.413	358.1363918	2E-135	0.98
comp15464_c0_seq1	molecular chaperone GroEL [Pedobacter glucos...	92	1018	1.018	351.7975824	0	0.92
comp66309_c0_seq1	glycosyl transferase family 2 [Pedobacter sp. V48]	128	1427	1.427	349.1715074	3E-143	0.9
comp48994_c0_seq1	50S ribsomal protein L25 [Pedobacter heparinus]	58	661	0.661	341.5696977	5E-69	0.89
comp21301_c0_seq3	ATP-depemdent DNA helicase RecQ [Pedobacter….	642	7320	7.32	341.4104015	0	0.96
comp67976_c0_seq1	acetylornithine aminotransferase [Pedobacter……	104	1194	1.194	339.0641035	0	0.96
comp13792_c0_seq2	acetyl-CoA acetyltransferase [Pedobacter sp….	200	2299	2.299	338.6443434	0	0.93
comp21301_c0_seq2	ATP-depemdent DNA helicase RecQ [Pedobacter….	626	7237	7.237	336.7197279	0	0.96
comp21301_c0_seq1	ATP-depemdent DNA helicase RecQ [Pedobacter….	366	4256	4.256	334.7590042	0	0.96
comp21509_c1_seq1	ABC transporter [Pedobacter sp. V48]	872	10148	10.148	334.4943817	0	0.95
comp24357_c0_seq1	nifR3 family TIM-barrel protein [Pedobacter sp. V48]	88	1029	1.029	332.9048319	0	0.94
comp37820_c0_seq1	glucuronate isomerase [Pedobacter heparinus]	114	1345	1.345	329.9403025	0	0.91

comp33903_c0_seq1	RNA procession exonuclease-like protein [Pedobacter	212	2527	2.527	326.5753645	3E-165	0.85
comp7168_c0_seq1	beta-lactamase [Pedobacter herparinus]	140	1671	1.671	326.1402404	3E-101	0.87
comp10767_c1_seq1	chemotaxis protein CheY [Pedobacter salatans]	108	1301	1.301	323.146354	1E-100	0.92
comp19130_c0_seq1	ATP synthase F0 subunit A [Pedobacter heparinus]	116	1407	1.407	320.9347124	4E-107	0.96
comp16053_c0_seq1	glycosyl transerferase family 2 [Pedobacter sp. V48]	312	3795	3.795	320.0336281	6E-160	0.82
comp54605_c0_seq1	transcriptional regulator [Pedobacter sp. R20-19]	104	1266	1.266	319.780837	2E-137	0.97
comp19387_c0_seq1	elongation factor G [Pedobacter oryzae]	52	634	0.634	319.2764508	3E-101	0.84
comp62832_c0_seq1	leucyl-tRNA synthetase [Pedobacter heparinus]	234	2853	2.853	319.2764508	0	0.92
comp21305_c1_seq1	homoserine 0-acetyltransferase [Pedobacter sp. V48]	244	2989	2.989	317.772794	0	0.9
comp20731_c0_seq2	glyceraldehyde-3-phosphate dehydrogenase [Pedobacter…….	266	3313	3.313	312.5453213	6E-129	0.88
comp18188_c1_seq2	histidine kinase [Pedobacter heparinus]	416	5205	5.205	311.1181861	1E-167	0.91
comp68802_c0_seq1	nitrite reductase [Pedobacter borealis]	50	629	0.629	309.4369417	2E-107	0.95
comp3954_c0_seq1	hypothetical protein N824_09520 [Pedobacter sp. V48]	58	731	0.731	308.8612451	1E-53	0.86
comp12997_c0_seq1	ABC transporter ATP-binding protein [Pedobacter….	238	3012	3.012	307.5918264	0	0.92
comp7779_c0_seq1	NmrA family protein [Pedobacter heparinus]	112	1429	1.429	305.0974622	2E-108	0.71
comp62059_c0_seq1	UDP-2,3-diacylglucosamine hydrolase [Pedobacter…..	262	3351	3.351	304.3544561	8E-134	0.88
comp20350_c0_seq2	membrane protein [Pedobacter heparinus]	408	5262	5.262	301.8298032	0	0.9
comp71823_c0_seq1	glutamyl-tRNA synthetase [Pedobacter sp.V48]	114	1473	1.473	301.269319	0	0.88
comp27894_c0_seq1	hypothetical protein N824_11010 [Pedobacter sp. V48]	148	2011	2.011	286.4853683	0	0.82
comp18573_c0_seq2	phenylalanyl-tRNA sythenase subunit beta [Pedobacter……	358	4870	4.87	286.1586424	0	0.9
comp3843_c0_seq1	metallophosphoesterase [Pedobacter sp. V48]	174	2375	2.375	285.1927202	1E-94	0.7
comp27963_c0_seq1	short-chain dehydrogenase [Pedobacter sp. V48]	32	446	0.446	279.2980611	9E-94	0.97
comp66142_c0_seq1	hypothetical protein N824_23655 [Pedobacter sp. V48]	68	948	0.948	279.2244066	7E-159	0.87
comp3614_co_seq1	hemolysin D [Pedobacter borealis]	208	2914	2.914	277.8603566	0	0.89
comp10942_c0_seq1	methionine synthase [Pedobacter agri]	354	5018	5.018	274.6157277	0	0.93
comp72378_c0_seq1	peptidase M16 [Pedobacter sp. BAL39]	256	3665	3.665	271.9059978	0	0.84
comp85567_c0_seq1	dehydrogenase [Pedobacter oryzae]	72	1035	1.035	270.7976854	2E-164	0.8
comp66203_c0_seq1	cyctochrome oxidase subunit III [Pedobacter g……..	74	1080	1.08	266.7231831	1E-61	0.77
comp21509_c1_seq4	ABC transporter [Pedobacter sp. V48]	520	7617	7.617	265.7493368	0	0.93
comp60731_c0_seq1	biotin synthase [Pedobacter heparinus]	104	1536	1.536	263.5693617	0	0.96
comp88974_c0_seq1	agrinase [Pedobacter heparinus]	122	1813	1.813	261.9478437	0	0.92

comp19714_c0_seq2	50s risomal protein L13 [Pedobacter sp. R20….	36	545	0.545	257.133582	5E-91	0.95
comp21461_c2_seq5	DNA mismatch repair protein MutS [Pedobacter….	404	6238	6.238	252.109259	0	0.92
comp6044_c0_seq1	helicase SNF2 [Pedobacter sp. BAL 39]	220	3414	3.414	250.8487639	0	0.91
comp5078_c0_seq1	2-methylthioadenine synthetase [Pedobacter b.	96	1504	1.504	248.4712804	0	0.83
comp23924_c0_seq1	hydrolase [Pedobacter sp. V48]	58	913	0.913	247.2919717	4E-141	0.89
comp60897_c0_seq1	ATP F0F1 synthase subunit alpha [Pedobacter……..	140	2213	2.213	246.2631459	0	0.94
comp19386_c1_seq3	hypothetical protein [Pedobacter borealis]	70	1107	1.107	246.1519159	5E-108	0.9
comp77726_c0_seq1	cell divison protein MraZ [Pedobacter sp. BAL39]	18	290	0.29	241.6169003	2E-53	0.98
comp20237_c0_seq3	methylmalonyl-CoA mutase [Pedobacter borealis]	148	2390	2.39	241.0552618	0	0.85
comp16836_c0_seq1	hypothetical protein N824_07065 [Pedobacter sp. V48]	150	2467	2.467	236.6872757	0	0.96
comp21461_c2_seq1	hypothetical protein [Pedobacter heparinus]	352	5874	5.874	233.2714144	0	0.97
comp14910_c1_seq1	glycogen branching protein [Pedobacter heparinus]	40	672	0.672	231.709329	4E-34	0.73
comp16963_c0_seq1	damage-inducible protein CinA [Pedobacter sp. V48]	26	437	0.437	231.6032835	3E-22	0.95
comp19605_c0_seq2	flagellar motor protein MotB [Pedobacter sp. V48]	50	842	0.842	231.1589505	2E-164	0.9
comp79186_c0_seq1	membrane protein [Pedobacter agri]	76	1289	1.289	229.5162694	0	0.98
comp20138_c0_seq4	transcriptional regulator [Pedobacter sp. V48]	72	1229	1.229	228.0517529	1E-114	0.78
comp65534_c0_seq1	beta-lactamase [Pedobacter herparinus]	50	857	0.857	227.1129946	7E-161	0.93
comp61986_c0_seq1	3-octaprenyl-4-hydroxybenzoate carboxy-lyase [Pedobacter….	38	667	0.667	221.7739665	9E-113	0.98
comp16265_c0_seq1	elongation factor P [Pedobacter heparinus]	36	637	0.637	219.9965497	2E-114	0.94
comp106324_c0_seq1	NADH dehydrogenase subunit G [Pedobacter sal…….	46	834	0.834	214.7061984	2E-124	0.9
comp21461_c2_seq4	hypothetical protein [Pedobacter heparinus]	416	7618	7.618	212.5715619	0	0.97
comp21461_c2_seq2	hypothetical protein [Pedobacter heparinus]	274	5041	5.041	211.5858725	0	0.97
comp17258_c1_seq1	aconitate hydratase [Pedobacter saltans]	38	708	0.708	208.9311238	5E-133	0.93
comp5754_c0_seq2	methylmalonyl-CoA carboxyltransferase [Pedobacter….	190	3563	3.563	207.5824244	0	0.95
comp85012_c0_seq1	cold-shock protein [Pedobacter heparinus]	16	301	0.301	206.9218194	8E-26	0.9
comp5142_c0_seq1	MultiSpecies: 50S ribosmal protein L11 [Pedobacter heparinus]	36	678	0.678	206.6929236	2E-65	0.97
comp145324_c0_seq1	superoxide dismutase [Pedobacter sp. V48]	20	378	0.378	205.963848	4E-84	0.99
comp21219_c0_seq6	1-deoxy-D-xylulose 5-phosphate reductoisomerase [Pedobacter sp	210	4060	4.06	201.3474169	0	0.87
comp64997_c0_seq1	transcriptional regulator [Pedobacter heparinus]	18	351	0.351	199.6264988	5E-74	1
comp16781_c0_seq1	ribonucleoside-diphosphate reductase [Pedobacter……	78	1527	1.527	198.8421118	0	0.93

comp104447_c0_seq1	hypothetical protein [Pedobacter borealis]	44	878	0.878	195.0791982	6E-79	0.75
comp3882_c0_seq1	peptidase M16 [Pedobacter heparinus]	104	2084	2.084	194.2622551	0	0.8
comp78740_c0_seq1	ABC transporter ATP-binding protein [Pedobacter….	80	1606	1.606	193.9086788	0	0.75
comp120757_c0_seq1	30S ribsomal protein S4 [Pedobacter sp. V48]	56	1136	1.136	191.8944865	2E-114	0.93
comp120577_c0_seq1	chromosome partitioning protein ParA [Pedobacter…..	16	325	0.325	191.6414389	1E-53	0.94
comp21219_c0_seq5	1-deoxy-D-xylulose-5-phosphate reductoisomerase [Pedobacter……	76	1605	1.605	184.3280195	0	0.87
comp15851_c0_seq1	dihydrolipoamide dehygrogenase [Pedobacter s…	20	428	0.428	181.9026508	3E-59	0.9
comp136965_c0_seq1	transcriptional regulator [Pedobacter oryzae]	22	471	0.471	181.8254098	3E-84	0.92
comp101658_c0_seq1	ABC transporter ATP-binding protein [Pedobacter….	42	905	0.905	180.6564669	8E-120	0.96
comp93548_c0_seq1	acetyl-CoA synthetase [Pedobacter oryzae]	52	1124	1.124	180.0900977	0	0.89
comp93909_c0_seq1	hypothetical protein N824_02200 [Pedobacter sp. V48]	26	562	0.562	180.0900977	2E-47	0.86
comp13208_c0_seq1	DNA helicase UvrD [Pedobacter borealis]	148	3265	3.265	176.4539282	0	0.93
comp106410_c0_seq1	MFS transporter [Pedobacter heparinus]	112	2478	2.478	175.941999	0	0.9
comp87029_c0_seq1	molecular chaperone DnaK [Pedobacter agri]	16	360	0.36	173.0096323	3E-44	0.86
comp7332_c0_seq1	30S ribsomal protein S10 [Pedobacter haparinus]	24	559	0.559	167.1291618	3E-51	1
comp18296_c0_seq1	elongation factor Tu [Pedobacter sp. BAL39]	14	335	0.335	162.680699	6E-68	0.98
comp12464_c0_seq1	isocitrate dehydrogenase [Pedobacter borealis]	148	3678	3.678	156.6400423	0	0.94
comp142093_c0_seq1	Fur family transcriptional regualtor [Pedobacter……	10	250	0.25	155.7086691	1E-52	1
comp81842_c0_seq1	deoxycytidine triphosphate deaminase [Pedobacter…..	28	733	0.733	148.6985926	6E-95	0.84
comp21599_c0_seq9	PAS sensor protein [Pedobacter heparinus]	94	2485	2.485	147.2496468	3E-158	0.7
comp131652_c0_seq1	succinate dehydrogenase [Pedobacter sp. R20-19]	34	901	0.901	146.8949708	7E-153	0.92
comp16265_c0_seq2	elongation factor P [Pedobacter heparinus]	22	585	0.585	146.3927658	9E-110	0.94
comp19605_c0_seq1	flagellar motor protein MotB [Pedobacter sp. V48]	16	428	0.428	145.5221206	5E-88	0.99
comp198959_c0_seq1	acyl-CoA dehydrogenase [Pedobacter heparinus]	16	446	0.446	139.6490306	2E-84	0.96
comp16676_c0_seq1	glucose dehydrogenase [Pedobacter oryzae]	52	1450	1.45	139.6008757	6E-165	0.75
comp115649_c0_seq1	transposon excision protein [Pedobacter sp. V48]	68	1918	1.918	138.010812	0	0.93
comp122247_c0_seq1	phytoene dehydrogenase [Pedobacter heparinus]	34	961	0.961	137.7235887	1E-173	0.92
comp13992_c0_seq1	phosphodiesterase [Pedobacter sp. V48]	40	1145	1.145	135.990104	2E-165	0.84
comp151335_c0_seq1	patatin [Pedobacter heparinus]	22	631	0.631	135.72071	7E-128	0.94
comp5760_c0_seq1	glutamine amidotransferase [Pedobacter sp. B…..	48	1379	1.379	135.4970289	0	0.92

comp11656_c1_seq1	glucuronyl hydrolase [Pedobacter heparinus]	18	524	0.524	133.7192769	4E-42	0.8
comp129359_c0_seq1	phytoene dehydrogenase [Pedobacter heparinus]	12	351	0.351	133.0843325	3E-62	0.96
comp21599_c0_seq2	hypothetical protein [Pedobacter arcticus]	94	2797	2.797	130.8242304	4E-173	0.73
comp9799_c0_seq2	phosphoribosylformylglycinamidine synthase [Pedobacter…..	36	1156	1.156	121.2264725	0	0.95
comp120279_c0_seq1	acriflavin resistance protein [Pedobacter borealis]	92	2957	2.957	121.1125935	0	0.97
comp21599_c0_seq4	hypothetical protein [Pedobacter arcticus]	90	2980	2.98	117.5652703	2E-172	0.73
comp12464_c0_seq3	isocitrate dehydrogenase [Pedobacter borealis]	110	3711	3.711	115.3863756	0	0.93
comp132293_c0_seq1	hypothetical protein N824_06280 [Pedobacter s	34	1161	1.161	113.9985949	7E-116	0.76
comp5754_c0_seq1	porin [Pedobacter heparinus]	68	2324	2.324	113.9004894	0	0.82
comp172863_c0_seq1	ATP synthase subunit epsilon [Pedobacter heparinus]	8	274	0.274	113.6559628	3E-33	0.92
comp103795_c0_seq1	hypothetical protein N824_02845 [Pedobacter sp. V48]	26	912	0.912	110.9765734	9E-102	0.85
comp256002_c0_seq1	translation intiatiom factor IF-2 [Pedobacter….	10	357	0.357	109.0396842	2E-58	0.98
comp219007_c0_seq1	3-ispropylmalate dehydratase large subunit [Pedobacter……	8	289	0.289	107.7568644	2E-50	0.94
comp185753_c0_seq1	amidophosphoribosyltransferase [Pedobacter sp. V48]	12	435	0.435	107.385289	6E-49	0.75
comp11478_c0_seq2	S-adensoyl-L-homocysteine hydrolase [Pedobacter…	10	369	0.369	105.4936782	1E-73	0.97
comp109135_c0_seq1	histone H1 [Pedobacter sp. BAL39]	8	300	0.3	103.8057794	8E-30	1
comp212790_c0_seq1	hypothetical protein N824_22195 [Pedobacter sp. V48]	16	602	0.602	103.4609097	7E-105	0.98
comp158265_c0_seq1	hypothetical protein [Pedobacter heparinus]	28	1063	1.063	102.5362826	2E-62	0.81
comp19521_c0_seq3	hydrolase TatD [Pedobacter oryae]	38	1457	1.457	101.5258995	2E-83	0.73
comp21219_c0_seq4	1-deoxy-D-xylulose-5-phosphate reductoisomerase [Pedobacter……	34	1342	1.342	98.62322557	0	0.87
comp28612_c0_seq1	ABC transporter permase [Pedobacter sp. V48]	22	869	0.869	98.54979056	3E-165	0.86
comp19714_c0_seq3	50S risomal protein L13 [Pedobacter sp. R20….	6	239	0.239	97.72510612	6E-44	0.94
comp187125_c0_seq1	hypothetical protein N824_02630 [Pedobacter sp. V48]	14	563	0.563	96.79935023	2E-83	0.99
comp259708_c0_seq1	hypothetical protein [Pedobacter heparinus]	8	333	0.333	93.51872017	7E-17	0.98
comp197182_c0_seq1	glycoside hydrolase family 3 [Pedobacter sp. V48]	6	250	0.25	93.42520145	2E-45	0.95
comp186602_c0_seq1	RNA polymerase subunit sigma-24 [Pedobacter…….	8	334	0.334	93.238724	2E-63	0.96
comp256945_c0_seq1	S-adensosylmethionine tRNA ribosyltransferase [Pedobacter…….	8	342	0.342	91.05770122	9E-66	0.95
comp27161_c0_seq1	hypothetical protein [Pedobacter sp. BAL39]	8	344	0.344	90.52829598	1E-46	0.94
comp211150_c0_seq1	30S ribsomal protein S19 [Pedobacter glucos…….	6	262	0.262	89.14618459	1E-45	0.99

comp262612_c0_seq1	glycyl-tRNA sythetease [Pedobacter heparinus]	6	265	0.265	88.1369825	2E-25	0.73
comp211261_c0_seq1	multidrug transporter AcrB [Pedobacter agri]	6	272	0.272	85.86875133	3E-65	0.94
comp306496_c0_seq1	transcriptional regulator [Pedobacter sp. BAL39]	8	367	0.367	84.85486053	7E-69	0.93
comp304620_c0_seq1	hypothetical protein N824_12855 [Pedobacter sp. V48]	10	468	0.468	83.17770784	4E-18	0.96
comp368850_c0_seq1	recombinase RecA [Pedobacter sp. V48]	8	381	0.381	81.73683416	2E-57	0.94
comp190709_c0_seq1	transcriptional regulator [Pedobacter heparinus]	10	479	0.479	81.26757259	2E-90	1
comp195198_c0_seq1	iron-sulfur binding protein [Pedobacter heparinus]	6	292	0.292	79.98733001	1E-61	1
comp19521_c0_seq4	hydrolase TatD [Pedobacter oryae]	22	1081	1.081	79.2227271	5E-85	0.73
comp9799_c0_seq1	phosphoribosylformylglycinamidine synthase [Pedobacter….	18	912	0.912	76.8299354	0	0.95
comp16676_c0_seq2	glucose dehydrogenase [Pedobacter oryzae]	24	1223	1.223	76.39018925	1E-83	0.75
comp332802_c0_seq1	30S ribsomal protein S13 [Pedobacter heparinus]	8	408	0.408	76.32777896	2E-42	1
comp189624_c0_seq1	thioredexin [Pedobacter heparinus]	6	306	0.306	76.32777896	4E-62	1
comp411158_c0_seq1	50s ribsomal protein L6 [Pedobacter agri]	4	206	0.206	75.58673256	3E-36	1
comp21623_c0_seq1	ATP-dependent Clp protease ClpP [Pedobacter….	10	519	0.519	75.00417586	3E-64	0.96
comp246517_c0_seq1	UDP-3-0-(3-hydroxymyristoyl0 glucosamine N-acyltransferase [Pedobacter sp. V48]	8	419	0.419	74.32394705	5E-45	0.92
comp386402_c0_seq1	ribsome hibernation protein yhbh [Pedobacter sp……	4	213	0.213	73.10266154	3E-33	1
comp371822_c0_seq1	cysteine desulfurase [pedobacter agri]	6	329	0.329	70.99179441	1E-67	0.99
comp437079_c0_seq1	ATP-dependent Clp protease ClpP [Pedobacter sp. V48]	4	230	0.23	67.69942134	4E-56	1
comp441940_c0_seq1	cyctochrome C oxidase [Pedobacter heparinus]	4	235	0.235	66.25900812	2E-20	0.86
comp16939_c0_seq2	methinoyl-tRNA synthetase [Pedobacter agri]	6	356	0.356	65.60758529	5E-28	0.91
comp398101_c0_seq1	major facilitator transporter [Pedobacter heparinus]	6	358	0.358	65.24106246	3E-25	0.89
comp340845_c0_seq1	biopolymer transporter ExbD [Pedobacter sp. V48]	4	240	0.24	64.87861212	6E-35	1
comp363913_c0_seq1	cyctochrome C [Pedobacter heparinus]	4	240	0.24	64.87861212	7E-22	0.84
comp291195_c0_seq1	hypothetical protein [Pedobacter heparinus]	12	721	0.721	64.78862791	1E-39	1
comp292636_c0_seq1	peptidyprolyl isomerase [Pedobacter sp. V48]	10	605	0.605	64.34242524	1E-100	0.98
comp419823_c0_seq1	DNA polymerase III subunit beta [Pedobacter sp. V48]	4	247	0.247	63.039947	1E-12	1
comp257890_c0_seq1	molecular chaperone ClpB [Pedobacter heparinus]	4	249	0.249	62.53360204	1E-48	1
comp380232_c0_seq1	hypothetical protein [Pedobacter borealis]	4	251	0.251	62.03532633	2E-22	0.99
comp368865_c0_seq1	Fis family transcriptional regulator [Pedobacter…….	4	256	0.256	60.82369886	1E-46	0.96
comp19521_c0_seq5	methylcrotonoyl-CoA carboxylase [Pedobacter…	22	1432	1.432	59.80430726	0	0.89

comp194577_c0_seq1	molecular chaperone GroES [Pedobacter heparinus]	6	396	0.396	58.98055647	8E-55	0.98
comp214228_c0_seq1	elongation factor G [Pedobacter sp. V48]	6	406	0.406	57.5278334	5E-88	0.99
comp19521_c0_seq1	methylcrotonoyl-CoA carboxylase [Pedobacter…	26	1808	1.808	55.97933346	0	0.89
comp367172_c0_seq1	histidyl-tRNA synthase [Pedobacter saltans]	4	279	0.279	55.80955881	7E-42	0.87
comp492518_c0_seq1	frumarate hydratase [Pedobacter heparinus]	4	288	0.288	54.0655101	2E-57	0.99
comp381947_c0_seq1	preprotein translocase subunit SecD [Pedobacter	4	292	0.292	53.32488667	5E-44	0.93
comp374627_c0_seq1	RNA polymerase subunit sigma-24 [Pedobacter sp. V48]	6	438	0.438	53.32488667	6E-102	1
comp252836_c0_seq1	30S ribsomal protein S8 [Pedobacter sp. BAL39]	4	296	0.296	52.60428009	9E-57	0.95
comp379948_c0_seq1	TonB family protein [Pedobacter sp. V48]	4	307	0.307	50.71943618	4E-64	1
comp525112_c0_seq1	hypothetical protein N824_10155 [Pedobacter sp. V48]	4	311	0.311	50.06709617	6E-33	0.97
comp541670_c0_seq1	cell divison protein ZqpA [Pedobacter heparinus]	4	316	0.316	49.27489528	4E-57	0.98
comp501407_c0_seq1	DNA polymerase III subunit beta [Pedobacter….	4	318	0.318	48.96499028	5E-61	1
comp185464_c0_seq1	hypothetical protein [Pedobacter heparinus]	4	320	0.32	48.65895909	5E-35	0.96
comp255010_c0_seq1	chromosal replication intiation protien [Pedobacter…..	4	321	0.321	48.50737355	1E-67	1
comp506226_c0_seq1	NADH dehydrogenase subunit N [Pedobacter sp. V48]	4	322	0.322	48.35672953	1E-12	0.92
comp490110_c0_seq1	MFS transporter [Pedobacter heparinus]	4	323	0.323	48.20701829	6E-59	0.94
comp491790_c0_seq1	hypothetical protein N824_09085 [Pedobacter sp. V48]	4	330	0.33	47.18444518	3E-64	0.98
comp308934_c0_seq1	AraC family transcriptional regualtor [Pedobacter……..	6	502	0.502	46.52649475	1E-64	0.94
comp259567_c0_seq1	ribose-phosphate pyrophosphokinase [Pedobacter sp….	4	342	0.342	45.52885061	1E-71	1
comp188534_c0_seq1	beta-N-acetylhexosaminidase [Pedobacter sp. …….	4	351	0.351	44.36144418	4E-42	0.74
comp201795_c0_seq1	glycogen synthase [Pedobacter sp. V48]	4	399	0.399	39.02472909	1E-84	0.99
comp397750_c0_seq1	hypothetical protein [Pedobacter sp. BAL39]	2	205	0.205	37.97772417	6E-15	1
comp504935_c0_seq1	sugar isomerase [Pedobacter arcticus]	2	208	0.208	37.42996853	6E-35	0.96
comp401136_c0_seq1	hypoxanthine phosphoribosyltransferase [Pedobacter	4	421	0.421	36.98543209	4E-34	0.88
comp388184_c0_seq1	short-chain dehydrogenase [Pedobacter heparinus]	2	211	0.211	36.89778888	3E-32	0.95
comp495014_c0_seq1	elongation factor Tu [Pedobacter sp. V48]	2	217	0.217	35.87757352	4E-29	1
comp636397_c0_seq1	elongation factor G [Pedobacter sp. R20-19]	2	221	0.221	35.22820567	8E-39	0.97
comp513079_c0_seq1	nucleotidyltransferase [Pedobacter borealis]	2	224	0.224	34.75639935	6E-15	0.97
comp540094_c0_seq1	thioredoxin [Pedobacter sp. V48]	2	225	0.225	34.60192646	2E-28	1
comp505728_c0_seq1	auxin-regulated protein [Pedobacter sp. BAL39]	2	227	0.227	34.29706367	3E-24	0.82
comp455647_c0_seq1	hypothetical protein [Pedobacter heparinus]	2	228	0.228	34.14663796	4E-15	0.89

comp302170_c0_seq1	translation intiatiom factor IF-2 [Pedobacter….	2	235	0.235	33.12950406	4E-27	0.92
comp283902_c0_seq1	amino acid transporter [Pedobacter salatans]	2	239	0.239	32.57503537	3E-22	0.84
comp117356_c0_seq1	AraC family transcriptional regualtor [Pedobacter……..	2	243	0.243	32.0388208	1E-28	0.89
comp417713_c0_seq1	30S ribsomal protein S4 [Pedobacter sp. V48]	2	250	0.25	31.14173382	2E-31	1
comp554048_c0_seq1	elongation factor Ts [Pedobacter sp. V48]	2	257	0.257	30.29351539	1E-48	0.99
comp277126_c0_seq1	succinate dehydrogenase [Pedobacter sp. V48]	2	264	0.264	29.49027823	5E-47	0.99
comp304100_c0_seq1	molecular chaperone GroEL [Pedobacter sp. V48]	2	283	0.283	27.51036556	2E-36	0.99
comp388924_c0_seq1	transketolase [Pedobacter heparinus]	2	290	0.29	26.84632226	2E-56	0.96
comp554062_c0_seq1	alanyl-tRNA synthetase [Pedobacter saltans]	2	300	0.3	25.95144485	8E-54	0.88
comp542246_c0_seq1	F0F1 Atp synthase subunit alpha [Pedobacter……	4	606	0.606	25.69449985	1E-130	0.99
comp290066_c0_seq1	ATP synthase F0 subunit C [Pedobacter glucos…….	2	320	0.32	24.32947954	8E-10	0.96

Chapter 5: Concluding Remarks

Our world is undoubtedly a microbial world, where microbiomes (i.e. groups of microorganisms co-existing within the same general space; [1]) are associated with every known environmental ecosystem. Additionally, host-associated microbiomes are widespread, with no known complex organism naturally existing in the absence of a symbiotic microbial flora. As these symbioses are critical for the majority of organisms, microbial acquisition is a vital aspect in the biology and ecology of all partners. However, as many symbioses are very complex, with multiple microbial species cohabitating and interacting simultaneously, it is difficult to determine the acquisition mechanism and subsequent impact of transmission on the symbiosis. Therefore, I utilized a simple model system composed of the medicinal leech host (*Hirudo verbana*) and its gut symbiont, *Aeromonas veronii*, to further investigate the impact of transmission. As determined by previous research, *A. veronii* is vertically transmitted to a majority of juvenile leeches through the maternally deposited cocoon (~80%) [2], leaving the remainder of the population to obtain this symbiont through the environment (i.e. horizontal transmission). My work expands on this knowledge, encompassing the discovery of a vehicle composed of host-derived mucosal secretions used for *A. veronii* horizontal transmission, how *A. veronii* adapts to this novel environment, and the presence of a previously uncharacterized host-associated microbiome within the mucus.

First, I examined the potential for *A. veronii* transfer to and subsequent survival within these mucosal secretions, which are constantly produced by host globose glands [3, 4] and are shed at regular intervals. I determined that the digestive tract seeds these mucosal secretions with *A. veronii* capable of proliferating within this environment, and that mucosal administration of this symbiont is sufficient for microbial transmission between leech individuals. Hosts also exhibit attraction to the mucus of conspecifics, which would allow for symbiont transmission to

occur naturally. Although aposymbiotic hosts (i.e. lacking the *A. veronii* symbiont) engage in more exploratory behavior, which may indicate the search for its missing symbiont while also risking predation, the host shows a general attraction to mucus, regardless of symbiont presence. As such, future work should investigate mucosal castings using gas chromatography/mass spectrometry to identify what molecular compounds comprise this substrate. Additionally, components which are capable of diffusing into the surrounding water within a designated time frame (i.e. 10 minutes; Chapter 2) should be identified by calculating the corresponding rate of diffusion [5].

As the leech host is attracted to factors independent of *A. veronii* presence, this implies that this symbiont hitchhikes on a previously established host biological process, which may be involved in conspecific mate attraction. This hitchhiking hypothesis is supported by an examination of the *A. veronii* transcriptome. By comparing mucosal gene expression with that found in the gut (crop) or *in vitro* monoculture environments [6], I showed that *A. veronii* experiences high levels of stress within mucus. The upregulation of genes involved in combating oxidative stress and environmental attack, in addition to an increase in energy spent to maintain cell integrity, indicates that the leech host has not expended the effort to make this harsh environment more hospitable for *A. veronii*.

Although *A. veronii* is capable of surviving in this environment, despite the stress it encounters, the mechanisms used for this survival was not entirely clear. However, further investigation into the *A. veronii* transcriptome revealed a potential metabolic overlap between the mucus and the leech crop, as indicated by the high expression of isocitrate lyase *aceA*, an enzyme critical to the TCA cycle, in both environments [7]. In the leech gut, isocitrate lyase converts short chain fatty acids (SCFAs), produced by *Mucinovorans hirudinis*, formerly known

as a *Rikenella*-like bacterium [8], another crop symbiont, to malate and succinate. It is feasible that this enzyme functions for a similar purpose within the mucus, as *M. hirudinis* is not only present, but also exhibits similar population dynamics to *A. veronii*, indicating a continuation of this syntrophic relationship.

In addition to the metabolism of SCFAs, I was also interested in examining genes potentially utilized for *A. veronii* proliferation that were dependent on the mucosal environment. As mucus consists of glycosaminoglycans [3], the major component of which is the monosaccharide N-acetylglucosamine (NAG) [9], I searched the transcriptome and discovered the presence of a pathway that converts NAG to N-acetylmuramic acid (MurNAc). As both of these compounds are critical for the synthesis of peptidoglycan, a major component of the cell wall, I confirmed increasing levels of expression of these genes in the presence of corresponding increases in environmental NAG. While these results showed the genes to be constitutively expressed, it is still possible that *A. veronii* scavenges NAG from the surrounding mucosal environment to assist in proliferation.

Although examination of the transcriptome revealed potential mechanisms by which *A. veronii* is able to survive and proliferate within the mucus, this does not tell the whole story. As it is hypothesized that many microbes are transmitted as species assemblages [10], investigation into the mucosal microbial community could further our understanding of *A. veronii* transmission through the mucus. Examination of the mucosal microbiome revealed not only a relatively diverse microbial community consisting of previously-described leech symbionts as well as novel leech-associated microbes, but also the expression of communication-related (i.e. quorum sensing) genes. The presence of this microbial community could indicate the broad use of mucus as a horizontal transmission mechanism, and certainly acknowledges a gap in our

understanding of leech-associated symbionts due to the number of novel microbes discovered. One of these novel microbes, and coincidentally the most abundant microbe within the mucosal environment, is a *Pedobacter* species. A brief glance into this microbe provided information not only on the presence of this microbe within the leech-associated niches, but activity within the mucosal secretions and its potentially relevant genomic repertoire, proving that further investigation into the microbes found in the mucosal environment will broaden our understanding of their relationships with the leech.

Additionally, interaction with this mucosal microbial community could provide a significant opportunity for *A. veronii* to engage in horizontal gene transfer (HGT) before relocating either to another host or into a free-living lifestyle. In fact, deeper investigation into the activity of the mucosal microbial community shows expression of genes related to type IV secretion systems, known to be involved in DNA transfer between cells [11].

The acquisition of microbial symbionts is critical for the survival of numerous organisms. As such, understanding the mechanisms used for transmission and their subsequent impact on both partners is crucial for understanding partner biology and ecology. This work confirms the use of a host biological process, i.e. mucosal secretions, as a mechanism for the horizontal transmission of a bacterial symbiont, i.e. *A. veronii*. As this vehicle can be used in conjunction with the vertical transmission of *A. veronii*, this suggests that a mixed transmission mode (MMT; i.e. combining both vertical and environmental mechanisms) supports the continuation of this symbiosis. Whether MMT represents a snapshot in the transition from horizontal to vertical transmission or if this is actually a stable strategy remains to be determined. It is possible that MMT benefits both the host and symbiont sufficiently to negate antagonistic selective pressures between vertical and horizontal transmission. As such, future investigations should examine the

role of species assemblages in microbial transmission, as this could elucidate the importance of mucus in the continuation and stability of the host microbiota.

Additionally, further inquiry into the impact of transmission on both partners would be beneficial. Comparing the *A. veronii* transcriptome with the other members of the microbial community, as well as examining the frequency of HGT events, could reveal interspecies interactions and further determine how mucosal transmission impacts the *A. veronii* symbiont. Additionally, as the absence of *A. veronii* was shown to alter host behavior, resulting in more exploratory behavior which could lead to predation, future work should examine how *A. veronii* colonization of the leech gut impacts this behavior, further elucidating the importance of symbiont acquisition.

This dual mode of symbiont transmission may prove evolutionarily advantageous, as it ensures the acquisition of beneficial symbionts by the leech host and provides accessibility to a higher genetic diversity of symbionts, while also ensuring biome lifestyle options to *A. veronii* beyond that of mutualism. While the impact of mucosal transmission on the biology of both partners is still not entirely understood, this work begins to reveal the inherent nature of this system. As such, understanding the features enabling mixed transmission, particularly if mixed species assemblages are involved, may prove instrumental for developing strategies to promote beneficial symbioses.

REFERENCES

1. Lederberg J, McCray AT. *Ome Sweet 'Omics-a genealogical treasury of words.*. Scientist, 2001. **15**:p. 8.
2. Rio R V et al. *Symbiont succession during embryonic development of the European medicinal leech, Hirudo verbana*. Appl Env Microbiol, 2009., 2009/08/04 ed. **75**:p. 6890–6895.
3. Michalsen A et al. *Medicinal Leech Therapy* 2007. Thieme Medical Publishers, New York, New York.
4. Sawyer RT. *Leech Biology and Behaviour* 1986. Clarendon Press, Oxford, United Kingdom.
5. Keyes JL. *Fluid, Electrolyte, and Acid-base Regulation* 1990. Jones and Bartlett Learning, Burlington, MA.
6. Bomar L, Graf J. *Investigation into the Physiologies of Aeromonas veronii in vitro and Inside the Digestive Tract of the Medicinal Leech Using RNA-seq*. Biol Bull, 2012. **223**:p. 155–166.
7. Bomar L et al. *Directed culturing of microorganisms using metatranscriptomics*. MBio, 2011., 2011/04/07 ed. **2**:p. e00012–11.
8. Nelson M et al. *Mucinivorans hirudinis gen. nov., sp. nov., an anaerobic, mucin-degrading bacterium isolated from the digestive tract of the medicinal leech, Hirudo verbana.*. Int J Syst Evol Microbiol, 2015.
9. Esko JD et al. *Proteoglycans and Sulfated Glycosaminoglycans* 2009. Cold Spring Harbor Laboratory Press.
10. Ebert D. *The Epidemiology and Evolution of Symbionts with Mixed-Mode Transmission*. Annu Rev Ecol Evol Syst, 2013. **44**:p. 623–643.
11. Babic A et al. *Direct visualization of horizontal gene transfer.*. Science, 2008. **319**:p. 1533–6.