The Dissertation Committee for Peter Allan Jorth Certifies that this is the approved version of the following dissertation:

Evolution, Metabolism, and Virulence of the Oral Microbiome

Committee:

Marvin Whiteley, Supervisor

M. Stephen Trent

Andrew Ellington

Christopher Sullivan

Scott Hunicke-Smith

Evolution, Metabolism, and Virulence of the Oral Microbiome

by

Peter Allan Jorth, B.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2013

Dedication

I dedicate my dissertation to my family.

Acknowledgements

My family has provided unwavering support throughout my graduate studies and inspired me to pursue my dreams. For that, I am eternally grateful. I thank my dissertation committee for their thoughtful guidance and discussions. I thank Marvin Whiteley for giving me the opportunities, resources, and freedom to explore any idea I put before him. Most importantly, Holly Huse, thank you for being my rock and keeping me focused.

Evolution, Metabolism, and Virulence of the Oral Microbiome

Peter Allan Jorth, Ph.D.

The University of Texas at Austin, 2013

Supervisor: Marvin Whiteley

The human microbiome has important roles in maintaining health, but dysbiosis of the microbiota can lead to disease. Polymicrobial interactions can result in synergy, producing disease that is worse than the sum of the respective single species infections. Despite this significant impact, synergy is understudied due to the complexity of polymicrobial interactions. Periodontitis is a microbiome-associated disease, and is one of the most common infectious diseases worldwide. Therefore, we have used periodontal disease as a model to study polymicrobial synergy. I have used two complementary approaches to study polymicrobial infections. The opportunistic periodontal pathogen *Aggregatibacter actinomycetemcomitans* exhibits synergy with streptococci in model murine infections. Because polymicrobial interactions are dependent on organisms' abilities to sense their environments, I have examined the genetic regulatory mechanisms used by *A. actinomycetemcomitans* to interact with its environment. Through Northern blot analyses and biochemical approaches, I show that *A. actinomycetemcomitans* uses non-coding RNAs to regulate amino acid transport. Taking a comparative genomics approach, I demonstrate that *A. actinomycetemcomitans* DNA uptake systems are evolutionarily linked to genome defense. To describe host-influenced changes in gene expression, I develop a new technique to transcriptionally profile

A. actinomycetemcomitans in a murine abscess infection, thereby revealing the importance of specific fermentative and anaerobic respiratory genes for *in vivo* survival. The long-term goal is to use these studies as a basis to characterize genetic regulatory mechanisms mediating synergy in polymicrobial *A. actinomycetemcomitans* infections with streptococci and other oral microbes. As a second approach to study polymicrobial infections, I analyze gene expression of healthy and diseased human plaque communities from aggressive periodontitis patients. Profiling ribosome content of healthy and diseased communities, I show that disease communities adopt similar less diverse population structures distinct from healthy populations. In addition to changes in population composition, using community transcriptional profiling I show that a keystone species within diseased communities up-regulates expression of genes involved in making the oral inflammatory molecule butyrate. These studies demonstrate for the first time that microbiome based diseases are marked by gene expression changes in addition to compositional changes.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Overview

The oral microbiome is composed of a complex bacterial population that plays important roles in health and disease [1]. Oral bacteria are important from both health and scientific perspectives. Oral diseases are the most prevalent diseases worldwide and represent significant healthcare burdens [2]. Additionally, the oral microbiome is an excellent system for studying microbial ecology, virulence, and evolution, because many oral bacteria are cultivable, and research has shown that complex interspecies metabolic and physical interactions mediate community structure [3].

Throughout my dissertation research I have used the oral microbiome to address questions related to microbial gene regulation, metabolism, evolution, ecology, and virulence. In this introductory chapter, I will provide a broad background about the oral microbiome, periodontal disease, and a major periodontal pathogen that has been the focus of most of my dissertation research.

1.2 The oral microbiome and periodontitis

1.2.1 The oral microbiome

Bacteria live throughout the human mouth, coating the tongue, buccal epithelium, teeth, and gums. Among these different environments, bacteria that reside in the subgingival crevice are of particular research interest due to their association with periodontal disease [1]. The subgingival crevice is the pocket between the tooth surface

and the gingival epithelium. Traditional culturing and metagenomic surveys of the subgingival crevice have revealed hundreds of species that populate this environment [4-6], and changes in the population composition have been linked to disease [6].

The subgingival crevice is a dynamic environment that supports the growth of a diverse microbiota. Oxygen concentrations decrease with depth within the crevice [7]. This oxygen gradient allows obligate aerobes, facultative anaerobes, and obligate anaerobes to co-populate the crevice. Capillaries bathe the pocket in a nutrient-rich serum exudate, termed crevicular fluid, which contains sugars and proteins that microbes can break down and utilize for growth [8]. While gingival crevicular fluid promotes microbial growth, it also contains antimicrobial factors to prevent bacterial overgrowth [9, 10]. These factors include neutralizing antibodies, bactericidal host complement proteins, and immune phagocytic cells that survey the environment [9, 10]. As a result of these selective pressures, the periodontal microbiome has evolved to utilize nutrients and withstand host defenses present in gingival crevicular fluid.

Highly evolved bacterial "specialists" dominate the subgingival crevice and most species are only found in this environment. Decades of research have revealed complex metabolic and physical interactions among microbes in the subgingival crevice [3]. Within this polymicrobial community, physical interactions promote the ordered development of multispecies surface-attached communities called biofilms [3]. Biofilm communities exhibit enhanced resistance to antimicrobial compounds and protect microbes from phagocytosis [11, 12], likely promoting community survival in the face of crevicular fluid antimicrobials and phagocytic cells. Bacteria are organized spatially

within biofilms [13], which may facilitate interspecies metabolic interactions [14-16]. Ultimately, these complex interactions dictate periodontal health, and imbalances in the population structure are highly correlated with periodontitis.

1.2.2 Periodontal disease

Periodontal disease affects greater than 10% of the world's population at any given time [2], and it is estimated that 90% of people worldwide will experience periodontal disease during their lifetimes [17]. If untreated, periodontitis can lead to bone destruction and eventual tooth loss [18]. Periodontitis can be acute or chronic and is not associated with a single infecting pathogen [18]. Indeed, periodontal infections are polymicrobial diseases and typically associated with disturbances in the indigenous microbiota rather than the invasion of a pathogenic species [1, 6, 18]. Since disease cannot be attributed to any single organism, the microbial contributions to periodontitis have been difficult to study due to the complexity of the system.

In an attempt to better characterize pathogens associated with periodontitis, Socransky et al. grouped bacterial species into complexes on the basis of co-occurrence (Fig. 1.1) [19]. While these complexes provide generalizations about interactions among species, they are not strict rules for infection. The red complex organisms (*Porphyromonas gingivalis*, *Treponema denticola,* and *Tannerella forsythia*) are most commonly associated with adult periodontitis, while disease in adolescents is frequently associated with *Aggregatibacter actinomycetemcomitans* [19]. Of these organisms, *A. actinomycetemcomitans* and *P. gingivalis* have been extensively studied in mono-culture, and numerous virulence factors produced by these bacteria have been identified

that likely contribute to disease at infected sites [20, 21]. However, behavior of these organisms in polymicrobial communities during infection remains understudied.

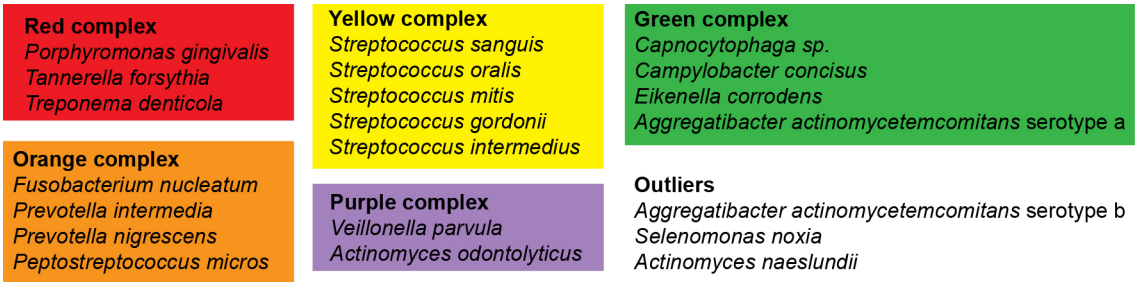


Figure 1.1 Co-occurring complexes of periodontal bacteria.

Complexes as defined by Socransky et al. based on co-occurrence of species in periodontal plaque [19].

1.3 POLYMICROBIAL SYNERGY

1.3.1 Polymicrobial synergy in periodontitis

The microbes residing in the subgingival crevice are almost exclusively found in that environment and are highly adapted to their lifestyles. As such, they have evolved to interact with one another in both healthy and diseased states. These interactions can lead to synergy, thereby exacerbating disease [22, 23]. Synergy occurs when two microbes together produce disease that is worse than the sum of the disease caused by the two organisms in isolation. Synergy between oral microbes is quite common [22-24], but often the molecular mechanisms underlying synergy are unknown.

1.3.2 Synergy in model polymicrobial infections

Recent studies have shown that metabolic interactions and interspecies bacterial communication can lead to synergy [22, 25]. Due to the importance of synergy in periodontitis, our lab has focused on synergistic interactions between *A. actinomycetemcomitans* and oral streptococci, two model organisms that inhabit the subgingival crevice. While we have shown that synergy with *Streptococcus gordonii* is dependent upon hydrogen peroxide sensing and lactate consumption [22], we know little about the global genetic regulatory mechanisms mediating these processes. This pioneering work has primarily been performed *in vitro* and in model infections. Therefore, the mechanisms underlying synergy in human infections remain uncharacterized and largely unexplored.

1.3.3 Bacterial gene regulation

Bacterial gene regulation is extremely complex and undoubtedly plays an essential role in synergistic interactions. Preliminary studies have revealed some regulators that play roles in model interactions. Work in our lab showed that the OxyR transcriptional regulator is used by *A. actinomycetemcomitans* to sense streptococcus-derived peroxide and increase resistance to innate immunity [16]. In otitis media, intercellular signaling was shown to mediate polymicrobial biofilm formation and enhance antibiotic resistance [25]. These studies demonstrate the importance of gene regulation in synergistic interactions. Yet for most infections we know very little about global genetic regulatory processes that mediate polymicrobial synergy.

Bacteria regulate gene expression in response to numerous stimuli, including host- and microbe-derived stresses. Bacteria process information from their environments using membrane-bound sensors that detect environmental perturbations, intracellular sensor proteins, DNA-binding proteins, and non-coding RNA regulators (discussed below). As a result, regulation occurs temporally and hierarchically. Sigma factors, such as sigma 70, are regulatory proteins that bind RNA polymerase and recognize consensus promoter sequences to drive gene expression under set environmental conditions [26]. Transcriptional regulators, like OxyR and Fur, bind DNA directly, controlling transcription from specific promoters in response to very specific stimuli [27, 28]. Genetic regulation is fine-tuned further by regulatory RNAs [29]. Due to the complex nature of these regulatory processes, it is difficult to predict which genes, proteins, and RNAs will affect interspecies interactions among bacteria. If we can define genetic

regulatory networks involved in polymicrobial diseases, we can then develop therapies to disrupt these regulatory schemes.

1.3.4 Bacterial non-coding RNAs

In addition to sigma factors and transcriptional regulators, bacteria use a myriad of non-coding RNAs (ncRNAs) to modulate gene expression. There are several types of regulatory bacterial ncRNAs, including small non-coding regulatory RNAs (sRNAs) and riboswitches. Bacterial sRNAs are post-transcriptional regulators that hybridize with target mRNAs and modulate mRNA stability or translation *in trans* [29]. Typically, RNA chaperone proteins facilitate sRNAs/mRNA hybridization, of which Hfq is the most common and widespread [30]. Due to their mechanism of action, bacterial sRNAs have been likened to eukaryotic microRNAs and regulate important cellular processes including metabolism and virulence. Riboswitches are *cis*-regulatory mRNA elements that act as structured sensory elements [31]. Riboswitches function by binding small intracellular molecules, which leads to a conformational change in the mRNA structure, impacting either transcription or translation of the mRNA. Riboswitches commonly regulate genes involved in amino acid and vitamin metabolism and transport [31]. Among oral bacteria, few regulatory ncRNAs have been characterized, especially in relation to biofilm formation and metabolism.

Bacteria use another class of ncRNAs, termed clustered regularly interspaced short palindromic repeats (CRISPR) RNAs (crRNAs), to defend against parasitic genetic elements [32]. These ncRNAs derive their names from the curious repetitive DNA elements from which they are transcribed. Found in eubacteria and archaea, CRISPRs

consist of anywhere from two up to hundreds of short DNA direct repeats (20-30 bp each) separated by similar length variable spacer sequences [33]. Their role in genome defense was proposed when researchers discovered that the spacer sequences in the CRISPRs were sometimes identical to phage and plasmid DNA sequences [34]. crRNAs are transcripts containing a spacer sequence flanked by half of each adjacent repeat. They function by hybridizing with foreign nucleic acids, which targets the complex for nucleolytic degradation [32]. Surprisingly, CRISPR spacer sequences are identical to bacterial chromosomal DNA, which has led researchers to speculate that some bacteria may have co-opted CRISPRs for bacterial gene regulation [35]. Similar to the regulatory RNAs described above, very little is known about crRNAs in oral bacteria.

1.4 *A. ACTINOMYCETEMCOMITANS* AS A MODEL ORGANISM

1.4.1 Virulence traits

Due to its prominence as a potent oral pathogen and synergistic interactions with other oral bacteria, our group has studied the biology of the opportunistic oral pathogen *A. actinomycetemcomitans*. *A. actinomycetemcomitans* is a member of the *Pasteurellaceae* family and a Gram-negative facultative anaerobe. *A. actinomycetemcomitans* can grow on both sugars and lactate, but lactate consumption is oxygen-dependent [22]. As mentioned above, it is commonly associated with aggressive periodontal infections, particularly in adolescents [19, 20, 36]. Several traits likely contribute to *A. actinomycetemcomitans* success as a pathogen. *A. actinomycetemcomitans* produces both a leukotoxin and cytolethal distending toxin

that kill host cells [37-40]. Tight-adherence fimbriae cause *A. actinomycetemcomitans* strains to auto-aggregate, attach firmly to surfaces, and form biofilms [41, 42]. Biofilms likely contribute to defense against both phagocytosis and antimicrobial proteins in gingival crevicular fluid. Other important self-defense mechanisms of *A. actinomycetemcomitans* include the factor H-binding surface protein, ApiA, and IgA protease, which protect it from innate and adaptive host immune systems, respectively [16, 43]. Despite these common virulence traits, extensive genomic variation is observed among *A. actinomycetemcomitans* strains [44]. Also, while many people carry *A. actinomycetemcomitans*, most carry the bacterium asymptomatically [45]. This begs the question, what triggers *A. actinomycetemcomitans* virulence?

1.4.2 Strain diversity

Diversity among *A. actinomycetemcomitans* strains abound. Several serotypes have been identified based on structural differences in lipopolysaccharide [46]. All *A. actinomycetemcomitans* strains produce a potent leukotoxin that kills eukaryotic cells, but certain highly virulent strains produce elevated levels of the leukotoxin due to a promoter mutation [47]. *A. actinomycetemcomitans* is genetically tractable, and some strains are naturally competent, or able to take up DNA from the environment and undergo natural transformation [48]. Despite these phenotypic differences among strains, the genetic bases for these differences were largely uncharacterized until recently. Whole genome sequencing and phylogenetic analyses showed that *A. actinomycetemcomitans* strains cluster into clades coordinately with serotype, but some traditional virulence factors are not part of the core genome [44]. Interestingly, most strains appeared to contain the genes required for natural transformation, suggesting that the common

ancestor to *A. actinomycetemcomitans* could undergo natural transformation, but some strains lost this ability throughout evolution. Because natural competence and virulence vary among strains, *A. actinomycetemcomitans* evolution with respect to these processes is of particular interest.

1.4.3 Natural transformation and evolution

A. actinomycetemcomitans natural transformation is relatively well characterized since it is closely related to *Haemophilus influenzae*, a model organism used to study the process. However, while nearly all *H. influenzae* strains are readily naturally transformable, or naturally competent, only approximately 25% of *A. actinomycetemcomitans* are competent [48, 49]. The basis for this discrepancy is currently unknown, but many competence genes are implicated in natural transformation and represent a good starting point for determining the basis of competence and non-competence for transformation.

Natural transformation requires genes involved in both DNA up-take and homologous recombination. Because *A. actinomycetemcomitans* is Gram-negative, extracellular DNA has to cross two membranes before it is incorporated into the chromosome via homologous recombination. This process requires genes that encode the machinery for a type IV pilus, which binds DNA, and several inner membrane transport proteins [50-53]. Additionally, several proteins mediate homologous recombination (Figure 1.2) [54, 55]. As in *H. influenzae,* the transcriptional regulator Sxy coordinately regulates competence gene expression [56]. The genes are also subject to catabolite repression: in high glucose competence genes are transcriptionally inactive, while in low

glucose they are activated [57]. These genes are present in all naturally competent species related to *A. actinomycetemcomitans*, including most of the *Pasteurellaceae* [58]. Therefore, parsimony suggests that natural competence is an ancient trait that has been lost throughout the course of *A. actinomycetemcomitans* evolution.

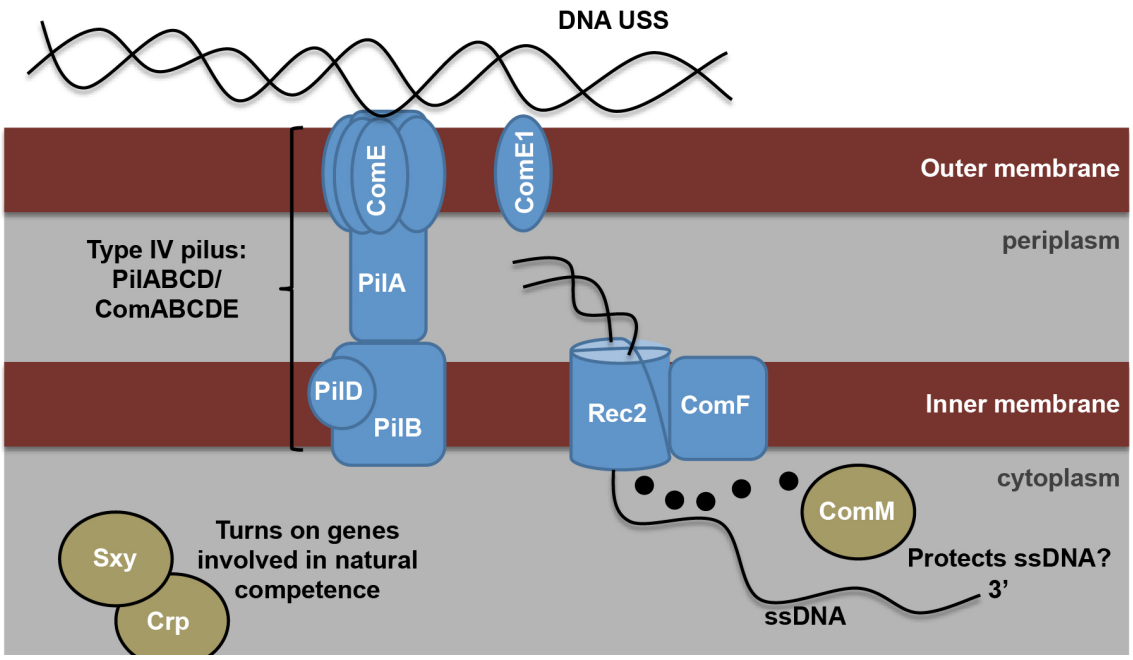


Figure 1.2 Natural transformation of *A. actinomycetemcomitans*.

DNA is bound outside the cell possibly through the ComE and ComE1 proteins, crosses the outer membrane via a type IV pilus consisting of PilABCD, is transported across the inner membrane by Rec2 and ComF, degrading a single strand and leaving single stranded DNA in the cytoplasm. The single stranded DNA may be protected by ComM. Genes encoding these processes are coordinately regulated by Sxy and Crp [50-57].

Natural selection is clearly acting on competence genes since so many bacteria still possess these traits after millennia of evolution. However, this raises the question, is competence loss due to drift or selection for competence gene loss? Because competence genes allow acquisition of new genetic material they have been theorized to accelerate evolution [59]. In addition, alternate theories suggest that competence systems serve to acquire nutrients from environmental macromolecules or aid in DNA repair [60]. Researchers have attempted to test these theories in the lab, but results have been largely inconclusive. Since both competent and non-competent *A. actinomycetemcomitans* strains are commonly isolated, we can use the evolutionary history of this species to address these questions. Specifically, we can determine how natural competence loss occurs and how this affects *A. actinomycetemcomitans* evolution.

1.4.3 Model for studying synergy

A. actinomycetemcomitans is one of many species that naturally populates the subgingival crevice. It has evolved to coexist with a diverse set of microbes, and is therefore an excellent model organism for studying polymicrobial interactions and synergy. *In vitro* studies have demonstrated that *A. actinomycetemcomitans* forms mixed species biofilms with several oral microbes, including Gram-negative and Gram-positive species [61, 62]. Among these, biofilms formed with the Gram-negative bacteria *Veillonella parvula* and *Fusobacterium nucleatum* demonstrated enhanced development over single species biofilms [61]. *A. actinomycetemcomitans* interacts metabolically with streptococci *in vitro*, preferentially consuming waste produced by oral streptococci, even in the presence of higher energy substrates like glucose [14, 22]. These studies demonstrate the utility in studying model interactions *in vitro*, but environmental changes

induced by the host during infection also likely influence microbial interspecies interactions. Thus, it is essential to build on *in vitro* methods and examine interactions in model infections *in vivo*.

While primarily associated with periodontitis, *A. actinomycetemcomitans* has also been detected in extra-oral abscesses and endocardial infections [63-65]. Consequently, several infection models are relevant for studying *A. actinomycetemcomitans in vivo*. Our lab has used a murine abscess model so study synergistic interactions between *A. actinomycetemcomitans* and *S. gordonii* [22]. Major advantages to the abscess model include the controlled inoculation of defined bacterial species and the ability to remove the abscess to assess microbial persistence and disease severity. These studies have allowed our group to determine several individual genes from each bacterium that mediate synergy [22]. The abscess model also presents a means to discover global genetic regulatory mechanisms promoting synergy. One of the goals of this dissertation research is to develop tools to perform global transcriptional profiling of abscess infections. These data will allow me to define genes impacting *A. actinomycetemcomitans* fitness in a single species abscess infection. Specifically, I will identify genes that are induced *in vivo* and likely contribute to persistence. This dataset will then serve as a baseline for studying expression changes in polymicrobial infections.

1.5 DISSERTATION OBJECTIVES

It is important to define molecular mechanisms promoting polymicrobial synergy so that we can target these processes with new therapeutics. I have taken two complementary approaches to begin studying mechanisms of synergy in oral bacteria.

First, I have studied gene regulation in the model polymicrobial pathogen *A. actinomycetemcomitans* in monoculture biofilms and in single species murine abscess infections. The goal of these studies was to better characterize *A. actinomycetemcomitans* gene regulation to serve as a basis for studying *A. actinomycetemcomitans* gene regulation in co-infections with *S. gordonii* and other microbes. Second, I have examined gene regulation in healthy and diseased human periodontal plaque communities to define genetic regulatory pathways mediating synergy *in vivo*.

This dissertation is comprised of 6 chapters. In Chapter 1, I introduced periodontal disease, polymicrobial synergy, and the opportunistic pathogen *A. actinomycetemcomitans*. In Chapter 2, I use a bioinformatics prediction to experimentally discover the first regulatory RNAs in *A. actinomycetemcomitans*. This study resulted in the characterization of the novel riboswitch-regulated lysine transporter LysT. Chapter 3 describes the sequencing of three *A. actinomycetemcomitans* genomes and the evolutionary history of natural competence and CRISPR genetic immunity in *A. actinomycetemcomitans*. In Chapter 4, I use high-throughput RNA sequencing to characterize the *A. actinomycetemcomitans* transcriptome in a murine abscess infection. Using these data, I defined fermentative and anaerobic respiratory genes that impact *A. actinomycetemcomitans* fitness *in vivo*. Chapter 5 builds on the techniques developed in Chapter 4 to analyze gene expression of the human oral microbiome in periodontal health and disease. I demonstrate that not only do bacteria change in abundance during disease, but also differentially regulate gene expression, impacting important disease processes. Chapter 6 examines the implications of these studies and proposes potential future directions for this research.

Chapter 2. Characterization of a Novel Riboswitch-Regulated Lysine Transporter in *Aggregatibacter actinomycetemcomitans*[1]

2.1 INTRODUCTION

Aggregatibacter actinomycetemcomitans is a Gram-negative opportunistic pathogen found exclusively in the mammalian oral cavity and a proposed causative agent of localized aggressive periodontitis [36]. Within the oral cavity, *A. actinomycetemcomitans* resides in the gingival crevice, the area around the tooth bounded by the tooth surface on one side and the gingival epithelium on the other. The gingival crevice is colonized by numerous bacterial species residing in surface-associated biofilm communities and is bathed by a serum exudate called gingival crevicular fluid [7, 66, 67]. Microbes growing in the gingival crevice face an array of host and microbe-derived stresses including host immune factors such as complement and immunoglobulins [66], potentially toxic metabolites such as H_2O_2 and lactic acid [16, 68], intense competition for nutrients, and steep oxygen gradients that may fluctuate continuously [7]. While it is clear that *A. actinomycetemcomitans* senses and responds to precise cues in the gingival crevice that allow it to adapt and proliferate in the presence of these stresses, little is currently known about the regulatory mechanisms that underlie these responses.

Bacteria utilize non-coding regulatory RNAs (ncRNAs) to rapidly respond to environmental stresses and alter gene expression [69]. In contrast to transcription factors, ncRNAs provide a rapid means of fine-tuning gene expression in response to environmental signals [70]. There are two major types of ncRNAs utilized in bacterial

[1] This Chapter was adapted from the following reference (153, © American Society for Microbiology, Journal of Bacteriology, 192: 6240-50).

gene regulation. *Trans*-acting small regulatory RNAs (sRNAs) regulate gene expression by binding target mRNAs, which typically reduces target gene mRNA levels due to degradation of the mRNA-sRNA duplex or reduces translation of a target gene by sequestration of the ribosomal binding site [29]. *Cis*-acting RNAs, or riboswitches, are co-transcribed with the downstream gene(s) they regulate and affect gene expression by binding small molecules such as amino acids and nucleotides [31, 71, 72]. Riboswitches have two functional components: the aptamer domain that mediates binding to a small molecule and the expression platform that controls the downstream gene [31]. Riboswitch expression platforms consist of an RNA stem-loop structure that functions in one of two ways to control the downstream gene. They can affect translation by occluding the ribosomal binding site of the target gene, or they can cause premature transcriptional termination of the downstream gene in a rho-independent manner [31].

Based on their importance in numerous other bacteria, we hypothesized that ncRNAs would play important roles in *A. actinomycetemcomitans* response to external stresses such as nutrient limitation. *A. actinomycetemcomitans* provides an excellent platform for such studies since it seemingly inhabits a single niche, the mammalian oral cavity, where it must respond quickly to changing environmental conditions. Since nutrient availability in the gingival crevice is limited due to competition with host cells and other microbes, ncRNAs involved in nutrient uptake likely play important roles in *A. actinomycetemcomitans* physiology *in vivo*. Using a bioinformatic and experimental approach, we identified three metabolite sensing riboswitches and nine sRNAs in *A. actinomycetemcomitans* during planktonic and biofilm growth. We also characterized the first riboswitch in *A. actinomycetemcomitans*, a lysine riboswitch, and its target, the

novel lysine transport gene, *lysT*. Finally, we demonstrated that this system is conserved in many bacterial species, including the phylogenetically-related pathogen *Haemophilus influenzae*.

2.2 MATERIALS AND METHODS

2.2.1 Bacterial strains and growth conditions

The bacterial strains used in this study are listed in Table 2.1. *Escherichia coli* was grown aerobically in Luria-Bertani (LB) broth at 37°C, unless otherwise noted. *A. actinomycetemcomitans* was grown at 37°C in Tryptic Soy Broth supplemented with 0.5% yeast extract (TSBYE) or a chemically defined medium (CDM) containing 20 mM glucose [14, 73] in a 5% CO_2 atmosphere shaking at 165 RPM. *H. influenzae* was grown in Brain Heart Infusion broth or CDM supplemented with 10 μg/mL hemin and 10 μg/mL NAD (sBHI and sCDM, respectively) at 37°C in a 5% CO_2 atmosphere shaking at 165 RPM. Bacterial biofilms were grown either as colony biofilms using 0.4 μm nylon filters on top of BHI or TSBYE agar plates [74]. Antibiotic concentrations were: 100 μg/mL ampicillin and 20 μg/mL kanamycin for *E. coli*; 2.5 μg/mL rifampicin for *A. actinomycetemcomitans*; 20 μg/mL kanamycin for *H. influenzae*.

2.2.2 DNA and plasmid manipulations

DNA and plasmid isolations were performed using standard methods [75]. Restriction endonucleases and DNA modification enzymes were purchased from New England Biolabs. Chromosomal DNA was isolated using DNeasy tissue kits (Qiagen),

and plasmid isolations were performed using QIAprep spin miniprep kits (Qiagen). DNA fragments were purified using QIAquick mini-elute PCR purification kits (Qiagen), and PCR was performed using the Expand Long Template PCR system (Roche). DNA sequencing was performed by automated sequencing technology using the University of Texas Institute for Cell and Molecular Biology sequencing core facility.

Table 2.1 Bacterial strains and plasmids used in this study.

Strain or Plasmid	Genotype or morphology	Source
Strains		
A. actinomycetemcomtians VT1169	Smooth strain	[76]
A. actinomycetemcomtians Y4	Smooth strain	[77]
E. coli K12	Wild type	[78]
E. coli DH5α	*F⁻ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG* φ*80*d*lacZ*ΔM15 Δ(*lacZYA-argF*)*U169, hsdR17*(r_K^- m_K^+), (Nal^R)	
E. coli K12 *lysP*	*rrnB3* Δ*lacZ4787 hsdR514* Δ*(araBAD)567* Δ*(rhaBAD)568 rph-1, lysP⁻* (Km^R)	[79]
E. coli K12 *lysP* pPEM-T easy	*rrnB3* Δ*lacZ4787 hsdR514* Δ*(araBAD)567* Δ*(rhaBAD)568 rph-1, lysP⁻* (Km^R Ap^R)	This study
E. coli K12 *lysP* pPJ005	*rrnB3* Δ*lacZ4787 hsdR514* Δ*(araBAD)567* Δ*(rhaBAD)568 rph-1, lysP* (Km^R Ap^R) pPJ005 $lysT_{Aa}$	This study
E. coli K12 *lysP* pPJ018	*rrnB3* Δ*lacZ4787 hsdR514* Δ*(araBAD)567* Δ*(rhaBAD)568 rph-1, lysP* (Km^R Ap^R) pPJ018 $lysT_{H.\ influenzae}$	This study
H. influenzae 86-028NP	Nontypeable	[80]
H. influenzae 86-028NP Δ*lysT*	Nontypeable, *lysT* (Km^R)	This study
Plasmids		
pGEM-T Easy	TA cloning vector	Promega
pPJ005	*A. actinomycetemcomitans lysT* in pGEM-T easy	This study
pPJ018	*H. influenzae lysT* in pGEM-T easy	This study

2.2.3 RNA preparation

Bacterial cultures were mixed 1:1 (vol/vol) with RNAlater (Ambion) and stored up to two weeks at 4°C. Total RNA was purified using RNAbee as described by the manufacturer (tel-test). RNA samples were treated with DNase to remove contaminating chromosomal DNA as described [14].

2.2.4 Northern blot ncRNA screen

Total RNA was prepared from: exponential phase planktonic *A. actinomycetemcomitans* VT1169; stationary phase planktonic *A. actinomycetemcomitans* VT1169; and *A. actinomycetemcomitans* VT1169 growing as a colony biofilm [16, 74]. Gel load dye II (Ambion) was mixed with 10 μg total RNA and separated on 8% polyacrylamide 8 M urea gels. A RNA ladder (Ambion) was transcribed with [α-^{32}P]UTP (Perkin Elmer) from the Century Plus Marker Template (Ambion) using the T7 MaxiScript *in vitro* transcription kit (Ambion) and served as the size standard. Gels were stained with ethidium bromide for imaging using a G:BOX imaging apparatus (Syngene) and transferred to nitrocellulose. RNA was UV crosslinked to the nitrocellulose and blots were pre-hybridized with ExpressHyb Hybridization Solution (ClonTech) for one hour at 55°C in a HL-2000 HybriLinker hybridization oven (UVP). DNA oligonucleotide probes (Table 2.2) were end-labeled with [γ-^{32}P]ATP (Perkin Elmer) using the Kinase Max system (Ambion). Probes were added to the pre-hybridization buffer and incubated 17 hours at 38°C. Blots were washed in 20 mL 2X SSC 0.1% SDS for 20 minutes at 38°C, and the wash was repeated four times. Blots were sealed and exposed to film (Kodak) for 2-17 hours, and developed with an X-OMAT 2000 Processor (Kodak).

Table 2.2 Primers and Northern blot probes used in this study.

Primer or Probe	Sequence (5'-3')
Primers	
Aa-lysT-F	GTCAGTGGTTCTCCCTTAATATTAG
Aa-lysT-R	GTGCTGGATCACGATTTAACG
NTHi-lysT-F	GGCTTAAGGGTGCTACGGTTGTTCTGGTC
NTHi-lysT-R	GGCTTAAGGGCAAGTTTTTGAATGAGATCACAC
NTHi1465KO-P1-F	CCTTCAACCGCACTTATTGG
NTHi1465KO-P1-R	CAGATAGCCCAGTAGCTGACATGTAGTGATTGGGGAGATTACCG
NTHi1465KO-P2-F	GCGGGACTCTGGGGTTCGAAATGTGTGATCTCATTCAAAAACTTG
NTHi1465KO-P2-R	GAAGTGCGGTCAAATTCCAAAC
NTHi1465KO-verify	GTAATGCGCGATTCGATAAAC
Kan-F	ATGTCAGCTGGGCTATCTG
Kan-R	ATTTCGAACCCCAGAGTCCCGC
lysRS-PE	CTTTTAACTGATTTCGGCAAACG
Aa-lysT-RT-F	GTTATTTGCGTTGTTACTCGCCATC
Aa-lysRS-lysT-RT-F	GTTGGTGACGTCTTCGAAAGAACCG
Aa-lysT-RT-R	AAGGCGCGCTTACTGCCGGACATAC
Probes	
Glycine-Probe[a]	AAAGTACCTTGGATCTCTCCAAGGG
FMN-Probe[a]	TTTCACCCTGCCCTGAGAATGCG
aa35-Probe[a]	GCAATACTTAACAGATGTGGCTTATG
aa84-Probe[a]	CTTGGACGGGGCTTGAATGAGGTCA
aa89-Probe[a]	TTACCGCCTTATCTTCACAAAGCGC
aa337-Probe[a]	CTCAAGCCTAAATCCTGAATTTGC
aa575-Probe[a]	GGTTGTGCTTTTTTAAGCACTGGG
aa673-Probe[a]	TCCACCTTTCCAGCCTTCCCAATTA
aa734-Probe[a]	CCGCACTTTAGTTTTTAGTTTCCTCCG
Alpha_RBS-Probe[a]	CAGGATACTCAAACAGAGTGTCAAGG
GcvB-Probe[a]	GGAAACTCTTTAACCAGTAAG
Aa-lysRS-Probe[a]	CACCGTTATCCACTCGGAAAATATAC
Aa-lysRS-Probe-F	GGAATTCCTAATACGACTCACTATAGGGAACAGTAGCGCTCCATAGTTG
Aa-lysRS-Probe-R	CCCAAGCTTGGGCAAAAAACGCGTCGGGGTTTAGCCCGACGACTTGTGTAGAGGTGCAAA
NTHi-lysRS-Probe-F	TAATACGACTCACTATAGGGCAACCGTAGCACTCCATAG
NTHi-lysRS-Probe-R	GTAGAGGCGCAATTATTATAAG
NTHi-lysT-Probe-F	TAATACGACTCACTATAGGGTCGGTGCTGCCGTTGAATC
NTHi-lysT-Probe-R	TCGTGCCTTTGCTGAATGG

[a]DNA oligonucleotide probe.

2.2.5 Primer extension analysis

Primer extension analysis was performed as previously described [81, 82] using a 6-carboxyfluorescein (FAM)-labeled primer, lysRS-PE (Table 2.2), to synthesize cDNA from 20 μg *A. actinomycetemcomitans* VT1169 colony biofilm RNA. Prior to primer extension, RNA integrity was confirmed by agarose gel electrophoresis. 1 μl of a 0.4 μM 5'-FAM labeled primer was added to the RNA (20 μl final volume) and incubated for 10 minutes at 70°C. The mixture was chilled in an ice water bath, incubated 20 minutes at 65°C, and held at 42°C. cDNA synthesis reagents (8 μl 5X buffer, 4 μl 0.1 M DTT, 4 μl 10 mM dNTPs, and 4 μl SuperScript II (Invitrogen)) were added to the mixture, and the reaction was incubated 2 hours at 42°C before ethanol precipitation. cDNA sizing was performed at the University of Oklahoma Health Sciences Center Laboratory for Genomics and Bioinformatics. Fragment analysis was conducted using Peak Scanner software (Applied Biosciences).

2.2.6 RT-PCR for AA02294, *lysT*, and the lysine riboswitch

1 μg RNA from lysine-starved *A. actinomycetemcomitans* Y4 was treated with RQ1 DNase (Promega) and used as a template for random primed cDNA synthesis with SuperScript II reverse transcriptase. A negative control reaction lacking reverse transcriptase was also performed (-RT). Reverse transcriptase reactions were purified with QIAquick spin columns (Qiagen). Three different products were amplified from 10 ng cDNA template that correspond to different regions of the *lysT* mRNA. A portion of the middle of the coding region of *A. actinomycetemcomitans lysT* was amplified using Aa-lysT-RT-F and Aa-lysT-RT-R (Table 2.2). A region spanning the lysine riboswitch and the coding region of *lysT* was PCR amplified using Aa-lysRS-lysT-RT-F and

Aa-lysT-RT-R (Table 2.2). A region corresponding to the 3' end of the *lysT* mRNA was PCR amplified using Aa-lysT-3'-F and Aa-lysT-3'-R (Table 2.2). *A. actinomycetemcomitans* chromosomal DNA served as a positive control template and the -RT products served as a negative control template. PCR products were separated on an agarose gel, stained with ethidium bromide, and visualized with a G:BOX imaging apparatus.

2.2.7 Agarose gel Northern blots

A. actinomycetemcomitans, *H. influenzae,* or *H. influenzae* Δ*lysT* were grown overnight in TSBYE or sBHI, respectively, pelleted by centrifugation, washed three times in CDM or sCDM free of L-lysine, resuspended in the same lysine-free medium and starved for L-lysine for two hours. A fraction of the cells was removed from the culture for RNA extraction, and 10 mM L-lysine was added to the growth medium. Cells were removed from the culture for RNA extraction at 15, 30, 60, and 120 minutes after the addition of L-lysine.

Northern Max glyoxal load dye (Ambion) was mixed 1:1 (vol/vol) with 20 μg total RNA from each sample and incubated at 50°C for 30 min. RNA was immediately separated by gel electrophoresis on a 2% agarose gel using the Northern Max buffer system (Ambion). Ethidium bromide stained RNA gels were imaged using a G:BOX imaging apparatus. Wells were cut away from gels, and RNA was transferred to nitrocellulose by downward transfer for 4 hours using a Whatman Turboblotter. RNA was UV crosslinked to the nitrocellulose and blots were pre-hybridized with UltraHyb buffer (Ambion) for one hour at 68°C in a HL-2000 HybriLinker hybridization oven.

RNA probe templates were PCR amplified from *A. actinomycetemcomitans* chromosomal DNA with primers Aa-lysRS-Probe-F and Aa-lysRS-Probe-R (Table 2.2), or from *H. influenzae* chromosomal DNA using NTHi-lysRS-Probe-F and NTHi-lysRS-Probe-R (probe 1, Fig. 2.7A) or NTHi-lysT-Probe-F and NTHi-lysT-Probe-R (probe 2, Fig. 2.7A) (Table 2.2). RNA probes were *in vitro* transcribed from PCR templates with [α-^{32}P]UTP (Perkin Elmer) using the T7 MaxiScript system. Probes were added to the pre-hybridization buffer and incubated 17 hours at 68°C. Blots were washed in 20 mL 2X SSC 0.1% SDS for 10 minutes at 25°C and washed twice more in 0.1X SSC 0.1% SDS for 15 min each at 68°C. Blots were then sealed and exposed to a phosphor screen (BioRad), imaged using a Personal Molecular Imager (BioRad), and processed with Quantity One software (BioRad).

2.2.8 Lysine riboswitch half-life

A. actinomycetemcomitans Y4 was grown overnight in TSBYE, washed three times in lysine-free CDM, resuspended in lysine-free CDM, and starved for lysine for two hours. After starvation, 10 mM L-lysine was added to the growth medium, and 15 min later rifampicin was added to the culture to inhibit transcription. At 0, 5, 10, 20, and 40 min after rifampicin treatment cells were removed for RNA extraction. RNA samples were subjected to polyacrylamide urea gel Northern blot analysis (as described above) probing for the *A. actinomycetemcomitans* lysine riboswitch with Aa-lysRS-Probe (Table 2.2). Northern blots were exposed to a phosphor screen, imaged on a Personal Molecular Imager, analyzed by densitometry with Quantity One software. The half-life was calculated using the following equation: $H=t((\ln(0.5))/(\ln(S_f/S_0)))$ where H= half-life, t= time , S_f= final signal, and S_0= original signal.

2.2.9 Expression of *A. actinomycetemcomitans lysT* and *H. influenzae lysT* in the *E. coli lysP* mutant

A. actinomycetemcomitans lysT was PCR amplified from *A. actinomycetemcomitans* Y4 chromosomal DNA using the primers Aa-LysT-For and Aa-LysT-Rev (Table 2.2). The PCR product was ligated into pGEM-T Easy TA cloning vector (Applied Biosystems) to create pPJ005, which was introduced into *E. coli* DH5α and the *lysP* mutant obtained from the Keio collection [79]. Similarly, *lysT* was PCR amplified from *H. influenzae* 86-028NP chromosomal DNA with NTHi-LysT-For and NTHi-LysT-Rev (Table 2.2). The PCR product was ligated into pGEM-T Easy to construct pPJ018, which was introduced into *E. coli* DH5α, strain K12, and the *lysP* mutant obtained from the Keio collection [79]. Plasmid constructs were confirmed by restriction enzyme digest and DNA sequencing.

2.2.10 Constructing a *H. influnezae* Δ*lysT* mutant

Allelic replacement of *H. influenzae lysT* was performed by natural transformation of *H. influenzae* strain 86-028NP as previously described [83, 84]. The regions flanking *lysT* on the *H. influenzae* chromosome were PCR amplified with NTHi1465KO-P1-F, NTHi1465KO-P1-R, NTHi1465KO-P2-F, and NTHi1465KO-P1-R and a kanamycin resistance gene (*aphA*) was amplified from pBBR1MCS-2 [85] using Kan-F and Kan-R (Table 2.2). The kanamycin resistance gene was placed between the two flanking regions by overlap extension PCR [86]. The resulting ~4 kb PCR product was then gel purified with a QIAquick gel purification spin column (Qiagen), and 2 μg product was used to transform *H. influenzae* 86-028NP by natural transformation [83, 84]. *H. influenzae* Δ*lysT* mutants were selected on sBHI agar + 20 μg/mL kanamycin.

Chromosomal DNA was extracted from mutants and allelic replacement was confirmed by PCR amplification with NTHi1465KO-verify and KanR primers (Table 2.2).

2.2.11 Lysine transport assays

Bacteria were grown overnight in either LB or sBHI for *E. coli* and *H. influenzae*, respectively. Cultures were diluted 1:1 in the same medium and incubated 1 hour shaking at 37°C. *E. coli* was washed three times in lysine-free MOPS 20 mM glucose 5 μg/mL Fe_2SO_4 700 μg/mL $MgSO_4$ 100 μg/mL $CaCl_2$ (MOPS) and *H. influenzae* was washed in lysine-free sCDM, and resuspended in 500 μL of the same medium to OD_{600} 0.40. Bacteria were incubated for 5 min at 37°C and 4 μCi (100 nM) [^{14}C]L-lysine was added and incubated at 37°C. At 10 and 20 min, 250 μL was removed from each transport reaction and quenched with 2.5 mL ice cold MOPS or sCDM with 10 mM L-lysine. The quenched transport reactions were filtered through a 0.4 μm filter, the filters were washed with 3 mL ice cold MOPS or sCDM with 10 mM L-lysine, and the filters were placed in 4 mL Ecolite scintillation fluid and counted on a Beckman-Coulter LS6500 liquid scintillation counter. Endpoint cold competition transport assays were carried out for 20 minutes, as described above, however the medium used for the transport assay contained 20 mM of the competing amino acid (i.e. L-arginine, L-glycine, or L-lysine).

2.2.12 Lysine riboswitch RNA secondary structure analyses

The *A. actinomycetemcomitans* lysine riboswitch nucleotide sequence was analyzed with Mfold to predict the RNA secondary structure [87]. The homologous sequence upstream of NTHi1465 was also analyzed with Mfold [72, 87]. Both riboswitch

structures were compared to lysine riboswitch crystal structures from *Thermotoga maritima* to identify conserved lysine-binding nucleotides and structural elements [88, 89].

2.2.13 Identifying LysT homologs and putative lysine riboswitches

BLASTp analysis was used to find proteins homologous to *A. actinomycetemcomitans* LysT, with an e-value cutoff of 10^{-100} [90]. ClustalW (version 1.83) was used to align amino acid sequences of proteins from 64 species and conduct phylogenetic analysis with 1000 bootstrap replicates [91]. Phylogenetic trees were drawn with Tree View software [92]. All loci encoding BLASTp LysT homologs were analyzed for putative lysine riboswitches in the 400 bp region upstream of their start codons. The 400 bp region was analyzed with RiboSW software which scanned for any riboswitches in the sequence. When a putative lysine riboswitch was identified by RiboSW, the predicted riboswitch start location relative to the start codon was recorded, along with the free energy associated with the secondary structure, and the corresponding HMM e-value [93].

2.2.14 Glycine riboswitch RNA secondary structure analyses

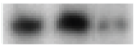
The intergenic region upstream of AA00167 was analyzed with RiboSW to identify the putative glycine riboswitch sequence [93]. Mfold software was used to generate a secondary structure, which was then compared to the consensus glycine riboswitch structure determined by Kwon and Strobel [72, 87].

2.3 RESULTS

2.3.1 Discovery of *A. actinomycetemcomitans* regulatory RNAs

While there has not been an investigation of ncRNAs in *A. actinomycetemcomitans*, a recent *in silico* analysis predicted 35 ncRNAs encoded in intergenic regions of the *A. actinomycetemcomitans* genome (http://www.oralgen.lanl.gov/). To determine if these predicted ncRNAs are produced by *A. actinomycetemcomitans*, RNA was isolated from logarithmic, stationary phase, and colony biofilm cells, and Northern blot analysis was used to probe for 23 of the predicted ncRNAs. Twelve ncRNAs were detected, including 3 putative riboswitches and 9 sRNAs ranging in size from ~110-500 nt (Table 2.3). Of the 12 ncRNAs identified, 8 were identified under all growth conditions, and 4 were only observed in colony biofilm-grown *A. actinomycetemcomitans*. Additionally, several ncRNAs (putative glycine riboswitch, putative lysine riboswitch, Aa84, Alpha_RBS, Aa337, GcvB, and Aa575) were expressed at higher levels during biofilm growth. We chose to further characterize the putative lysine riboswitch since it was only observed in colony biofilms, it possessed high homology to characterized lysine riboswitches, and it was positioned upstream of a gene of unknown function that is conserved in a large number of bacteria (see below).

Table 2.3 Riboswitch and sRNA discovery in *A. actinomycetemcomitans*.

Riboswitch or sRNA	*Aa* exp[a]	*Aa* stat[b]	*Aa* bio[c]	ORF up[d]	IGR start[e]	IGR stop[f]	ORF down[g]	downstream ORF description	size (nt)[h]
Riboswitches									
glycine				AA00166	118129	118272	AA00167	possible amino acid transporter	250
FMN				AA00470	328126	328309	AA00471	*ribA*, riboflavin biosynthesis	500
lysine				AA02292	1563630	1563804	AA02294	hypothetical gene	200
sRNAs									
Aa35				AA00139	100597	100785	AA00140		110
Aa84				AA00349	233588	233760	AA00350		450
Aa89				AA00370	249069	249320	AA00371		500
Alpha_RBS				AA01238	844559	844675	AA01239		350
Aa337				AA01321	894176	894386	AA01323		175
GcvB				AA01794	1210848	1210988	AA01796		175
Aa575				AA02262	1538988	1539231	AA02263		210
Aa673				AA02609	1820245	1820510	AA02610		400
Aa734				AA02879	2020650	2020896	AA02880		500

[a]*A. actinomycetemcomitans* VT1169 grown planktonically to exponential phase.
[b]*A. actinomycetemcomitans* VT1169 grown planktonically to stationary phase.
[c]*A. actinomycetemcomitans* VT1169 grown as a colony biofilm
[d]ORF predicted upstream of riboswitch or sRNA.
[e]Beginning coordinates from intergenic region (IGR) from *A. actinomycetemcomitans* strain HK1651 genome (http://oralgen.lanl.gov/cgi-bin/coordinate.cgi?dbname=aact).
[e]Ending coordinates from intergenic region (IGR) from *A. actinomycetemcomitans* strain HK1651 genome (http://oralgen.lanl.gov/cgi-bin/coordinate.cgi?dbname=aact).
[g]ORF predicted downstream of riboswitch or sRNA.
[h]Approximate size of riboswitch or sRNA, as determined by Northern blot analysis.

2.3.2 Mapping the transcriptional start site and promoter of the putative lysine riboswitch

To determine a transcriptional start site and map a putative promoter region for the putative lysine riboswitch, primer extension was used. The major primer extension product was 93 nt in length (Fig. 2.1A), corresponding to an adenine transcriptional start site (Fig. 2.1B) positioned 7 bp upstream of the predicted riboswitch aptamer domain and 267 bp upstream of the hypothetical gene AA02294 (Fig. 2.1D). A promoter with a -10 site and a -35 site was identified upstream of the transcriptional start site (Fig. 2.1B). Little is known about *A. actinomycetemcomitans* promoters, but this promoter varies considerably from the canonical *E. coli* sequence for these sites. Upstream and in the opposite orientation of the putative riboswitch is *hns*, which encodes a global transcriptional regulator [94]. A canonical rho-independent terminator was identified 4 bp downstream of the aptamer domain (Fig. 2.1C).

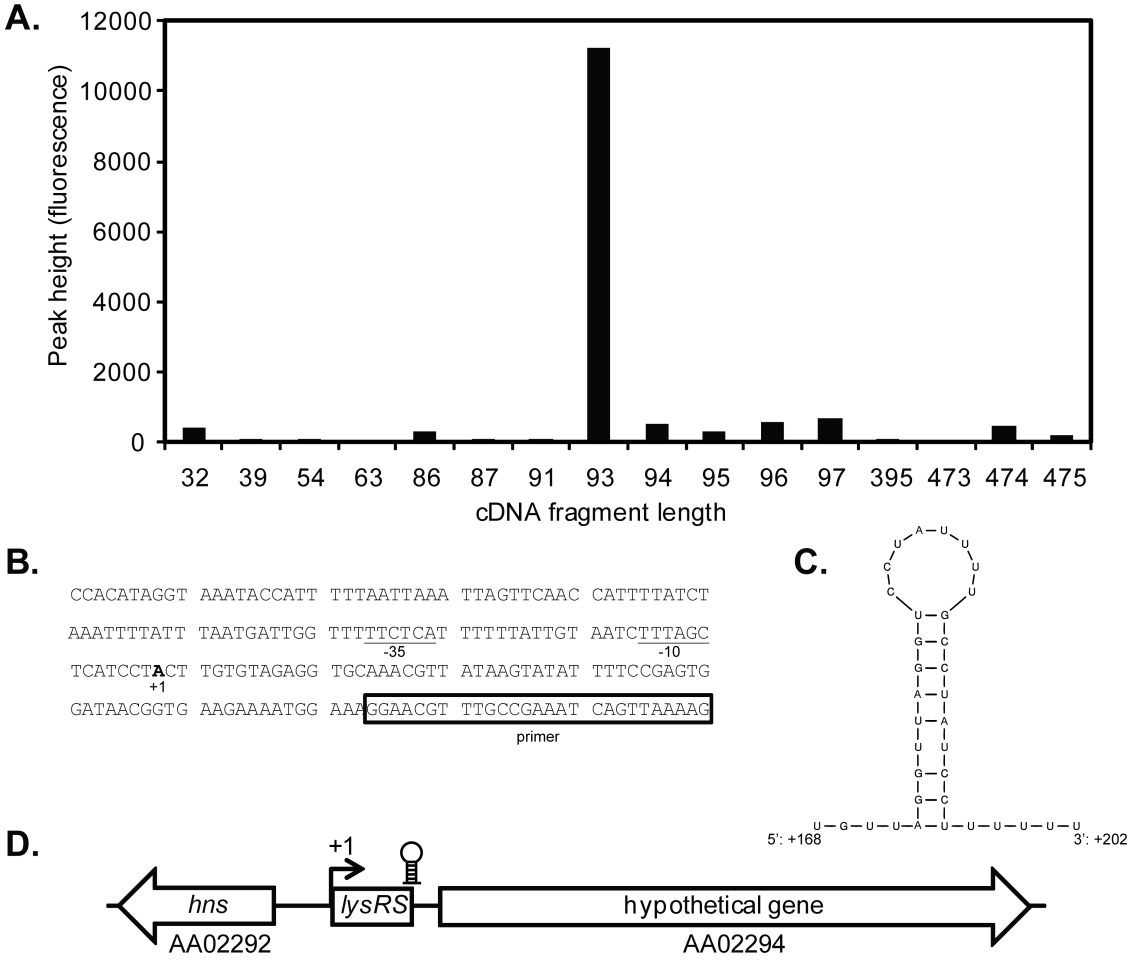


Figure 2.1 Mapping the transcriptional start site of the *A. actinomycetemcomitans* AA02294 lysine riboswitch with primer extension analysis.

(A) Fluorescent intensities of primer extension products synthesized from *A. actinomycetemcomitans* colony biofilm RNA. Fluorescent peak height corresponds to cDNA levels. **(B)** Transcriptional start site (+1) and promoter sequence for the A. actinomycetemcomitans AA02294 lysine riboswitch. The sequence complementary to the primer used for primer extension is boxed, the transcriptional start site is bold-faced, and the putative -10 and -35 promoter sequences are underlined. Sequence begins at 1,563,524 and ends at 1,563,723 on the *A. actinomycetemcomitans* chromosome. **(C)** The AA02294 lysine riboswitch rho-independent terminator. The terminator begins at +168 and terminates at +202 relative to the transcriptional start site. **(D)** Schematic of the *A. actinomycetemcomitans* AA02294 lysine riboswitch genomic context. The arrow indicates the transcriptional start site determined by primer extension analysis and the rho-independent terminator is shown as a hairpin. Image is drawn to scale.

2.3.3 Structural predictions demonstrate features characteristic of lysine riboswitches

Structures of riboswitch aptamers are critical for their function. Structural studies of lysine riboswitch aptamers show two characteristic elements: an "RNA kissing loop" that binds L-lysine at the base of the two loops; and a loop E motif that enables RNA twisting [88, 89, 95]. The secondary structure prediction program Mfold was used to generate an RNA secondary structure of the putative *A. actinomycetemcomitans* lysine riboswitch based on free energies of internal base-pairing [87] (Fig. 2.2A). The predicted structure has the characteristic kissing loops and loop E motif. Conserved nucleotides known to be important for coordinating L-lysine at the base of the kissing loops were also present [89] (Fig. 2.2A), providing strong *in silico* support that this ncRNA is a lysine riboswitch.

2.3.4 The riboswitch accumulates in the presence of lysine

Riboswitches are co-transcribed with the genes they regulate [96]. To determine whether this putative riboswitch formed a contiguous transcript with the downstream gene AA02294, RT-PCR was used. Primers were designed to amplify a region spanning the riboswitch and the AA02294 coding region (Fig. 2.2B). Our results reveal that *A. actinomycetemcomitans* does produce a contiguous mRNA containing both the putative riboswitch and AA02294. Based on the location of this mRNA, we hypothesize that the putative riboswitch serves as a *cis*-regulatory element in the 5' untranslated region of AA02294.

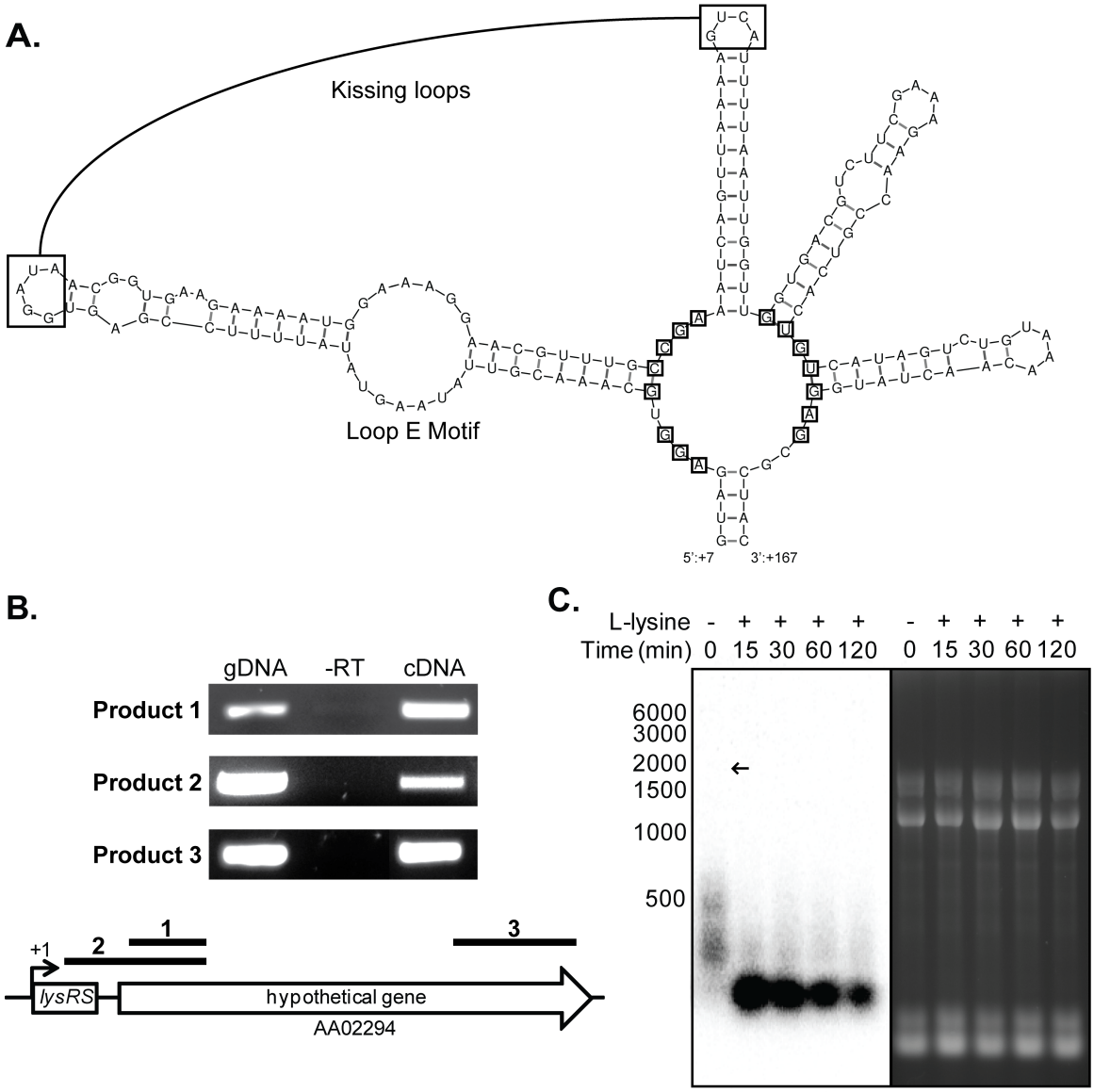


Figure 2.2 The *A. actinomycetemcomitans* AA02294 lysine riboswitch.

(A) Predicted structure of the *A. actinomycetemcomitans* lysine riboswitch aptamer. The structure was predicted using Mfold software [87]. Structural motifs conserved in all lysine riboswitches, kissing loops and loop E motif, are identified. Boxed nucleotides are predicted to be involved in binding lysine based on sequence and structural homology to the crystal structure of the lysine riboswitch from *T. maritima* [88, 89]. **(B)** RT-PCR detecting the AA02294 mRNA (Products 1 and 3) and the read through transcript containing both the lysine riboswitch and AA02294 (Product 2). The positive control

reactions (gDNA) show PCR amplification from *A. actinomycetemcomitans* chromosomal DNA, while the negative control (-RT) represents PCR amplification from cDNA synthesis reactions without reverse transcriptase added, and the final lane (cDNA) depicts PCR amplification from cDNA synthesized from RNA harvested lysine-starved *A. actinomycetemcomitans*. **(C)** The AA02294 lysine riboswitch accumulates in the presence of lysine. *A. actinomycetemcomitans* lysine-starved cells (- lys, 0 min) were exposed to lysine (+ lys) and samples removed at 15, 30, 60 and 120 min post-addition. The arrow indicates where the full-length *lysT* transcript should be detected. RNA from each timepoint was subjected to Northern blot analysis using a probe for the lysine riboswitch.

Upon binding lysine, most lysine riboswitch expression platforms prematurely terminate transcription of the downstream gene [71, 89, 95-98]. Because we identified a potential rho-independent terminator downstream of the lysine aptamer domain and upstream of AA02294 (Fig. 2.1C), we hypothesized that *A. actinomycetemcomitans* grown in the presence of lysine would produce high levels of the prematurely terminated riboswitch transcript compared to lysine-starved cells. To test this hypothesis, Northern blot analysis was carried out to probe for the riboswitch in the absence and presence of lysine. As anticipated, the terminated riboswitch transcript is only present after the addition of lysine to the media (Fig. 2.2C), consistent with previous studies of lysine riboswitches [71, 96]. The full-length mRNA (approximately 2000 nt) was not observed by Northern blot, although in the absence of lysine, an RNA larger than the riboswitch was detected (0 min, Fig. 2.2C). Thus, it appears that the full length transcript is unstable or there is a technical issue in detecting the riboswitch when present in the full length transcript. Importantly, the full length transcript could be detected by RT-PCR (Fig. 2.2B), and the full-length transcript along with the riboswitch were detected in the closely related bacterium *H. influenzae* (see below).

2.3.5 Riboswitch half-life

Since the prematurely terminated transcript accumulated to high levels upon exposure to lysine (Fig. 2.2C), we used Northern blot analysis to calculate the lysine riboswitch half-life. For these experiments, *A. actinomycetemcomitans* grown in the absence of lysine were exposed to lysine to induce production of the riboswitch. After 15 min, the transcription inhibitor rifamipicin was added and degradation of the riboswitch over time was monitored using Northern blots and densitometry (Fig. 2.3). The results

indicate that the riboswitch has a half-life of 17 ± 4 min, which is nearly five times the average half-life of an *E. coli* transcript (3.69 min) [99].

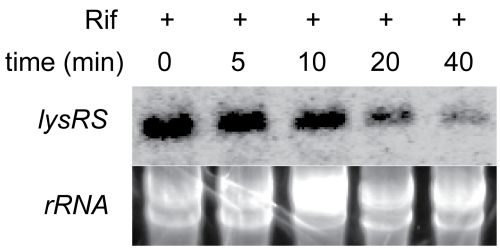


Figure 2.3 The *A. actinomycetemcomitans* lysine riboswitch has a half-life of 17 ± 4 min.

A. actinomycetemcomitans lysine-starved cells were exposed to lysine and treated with rifampicin (Rif) to stop transcription. Lysine riboswitch levels were monitored by Northern blot analysis and half-life was calculated as described in Materials and Methods. The ethidium bromide stained gel was used to ensure equal RNA loading.

2.3.6 *A. actinomycetemcomitans lysT* encodes a novel lysine transporter

Lysine riboswitches have been shown to regulate lysine biosynthetic genes in *B. subtilis* and the lysine transport gene *lysP* in *E. coli* [71, 100]. The gene downstream of the *A. actinomycetemcomitans* lysine riboswitch, AA02294, is a hypothetical gene of unknown function. A previous study hypothesized that genes downstream of several lysine riboswitches encoded novel lysine transporters, although no empirical evidence was provided to test this hypothesis [100]. There are several well characterized bacterial lysine transporters, including the lysine-specific LysP transporter in *E. coli* [101, 102], the lysine-arginine-ornithine (LAO) transporter in *E. coli* [103], and LysI in *Corynebacterium glutamicum* [104]. While LysI and LysP uniquely transport lysine, the LAO transporter preferentially transports arginine and ornithine [105]. Despite the prevalence of several bacterial lysine transporters, no lysine transport genes are annotated in the *A. actinomycetemcomitans* genome. Bioinformatic analysis with Phyre protein prediction software predicted that AA02294 encodes an inner membrane transport protein [106] with an NhaC domain [107-109]. Despite having no homology to other lysine transporters, these observations led us to hypothesize that AA02294 encodes a novel lysine transporter.

To determine whether *A. actinomycetemcomitans* AA02294 encodes a lysine transporter, we tested its ability to heterologously restore lysine transport in an *E. coli* lysine transport mutant. A plasmid carrying *A. actinomycetemcomitans* AA02294 (pPJ005) was transformed into an *E. coli lysP* mutant that is unable to take up L-lysine, and the ability of this strain to transport [^{14}C]L-lysine was assessed. Expression of AA02294 *in trans* restored L-lysine transport to the *E. coli lysP* mutant (Fig. 2.4A),

indicating that AA02294 encodes a novel lysine transporter (referred to as LysT). It was also important to determine the specificity of LysT since some lysine transporters have been shown to also transport L-arginine [102, 105]. To test the specificity of *lysT*, the ability of unlabeled L-lysine, L-arginine, and L-glycine to inhibit uptake of [^{14}C]L-lysine by the *E. coli lysP* mutant expressing *A. actinomycetemcomitans lysT* was assessed. The addition of unlabeled L-lysine reduced uptake of [^{14}C]L-lysine to background levels while addition of unlabeled L-arginine inhibited [^{14}C]L-lysine uptake approximately 50% (Fig. 2.4B). The addition of unlabeled glycine had no effect on lysine uptake.

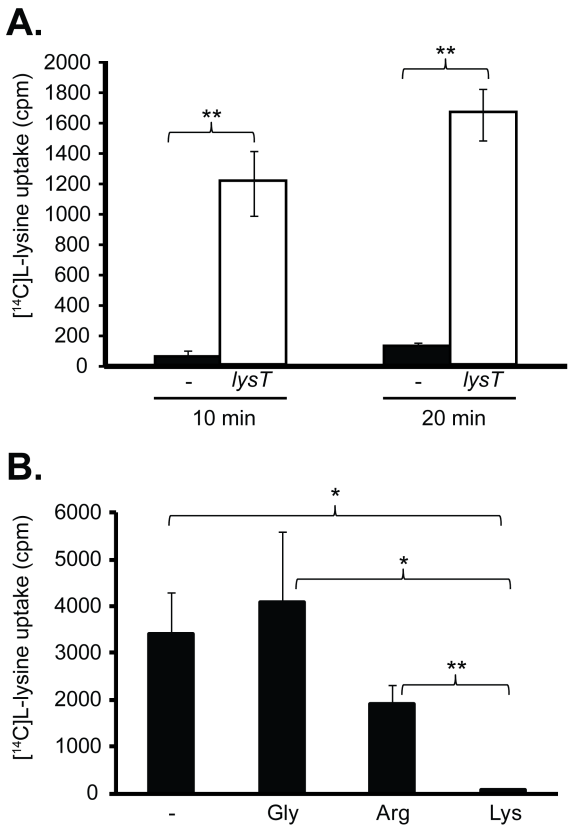


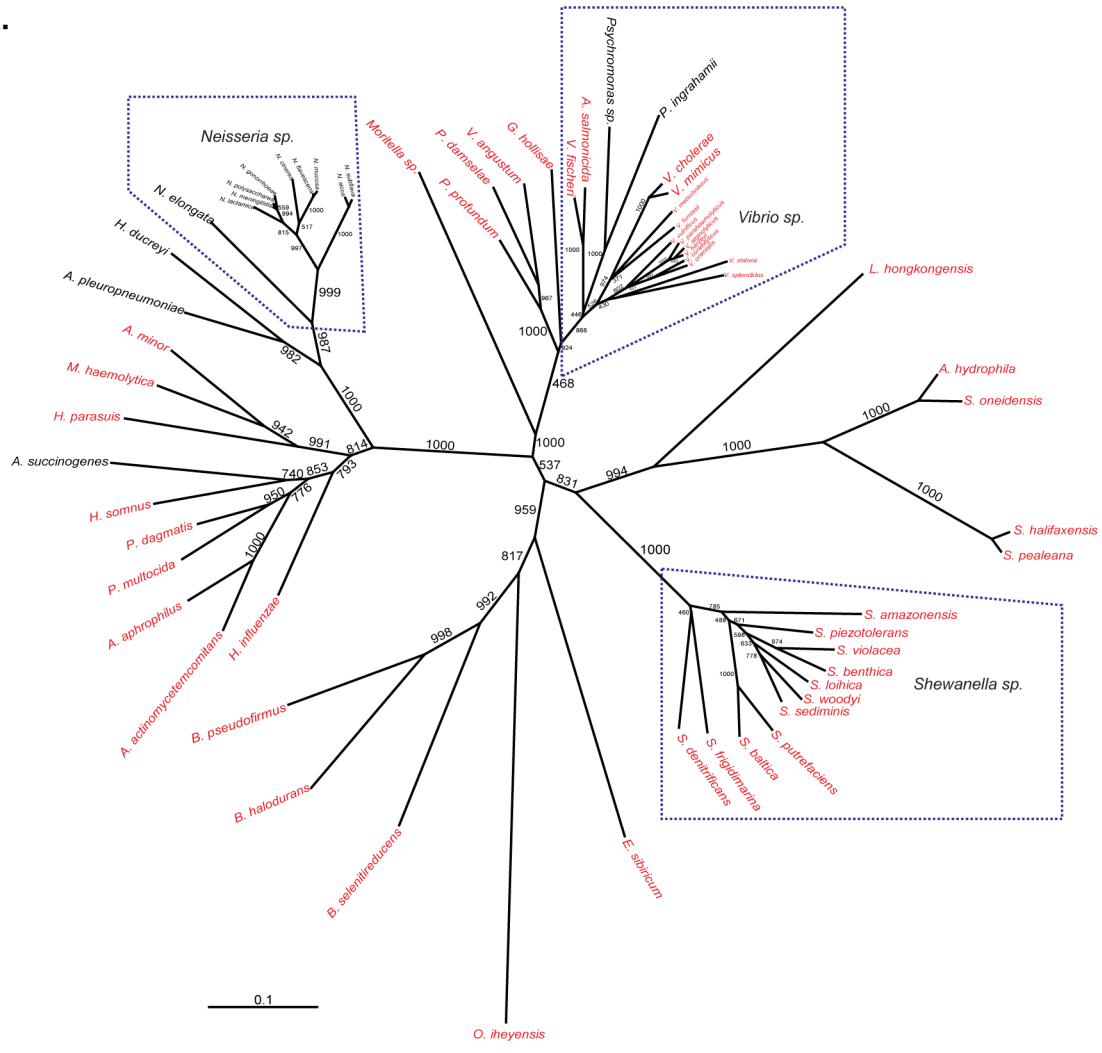
Figure 2.4 *A. actinomycetemcomitans lysT* restores lysine transport in an *E. coli* lysine transport mutant.

(A) Transport of [^{14}C]L-lysine by the *E. coli lysP* mutant with an empty vector (-, black) or a vector expressing *A. actinomycetemcomitans lysT* from pPJ005 (*lysT*, white). Transport was measured in counts per minute at 10 and 20 minutes after the addition of [^{14}C]L-lysine to the assay. **(B)** Unlabeled L-lysine inhibits up-take of [^{14}C]L-lysine by *A. actinomycetemcomitans* LysT. Lysine transport assays were carried out in the presence of excess unlabeled glycine, L-arginine, or L-lysine. The "-" represents a negative control assay where no competing amino acid was added. The data in both graphs represents the mean of three independent experiments, error bars in all assays represent the standard error, and statistical significance was determined by an unpaired two-tailed Student's t-test: $*P < 0.05$, $**P < 0.005$.

2.3.7 *Haemophilus influenzae* LysT is a lysine transporter

Although LysT lacks homology to known lysine transporters such as LysP in *E. coli* [101, 102], a BLASTp homology search revealed LysT homologs from 64 species of bacteria, including *Gamma-proteobacteria, Beta-proteobacteria*, and *Bacilli* (e-value < 10^{-100}, Fig. 2.5) [90, 110]. Of the genes encoding the LysT homologs, 48 are predicted to have lysine riboswitches within 400 bp of their start codons [93]. Included within this group containing riboswitches is the LysT homolog (ORF NTHi1465) from the human pathogen *Haemophilus influenzae* which causes otitis media, bacterial meningitis, and pulmonary infections [111]. The *H. influenzae* LysT is a hypothetical inner membrane protein with 75% identity and 92% similarity to *A. actinomycetemcomitans* LysT over 249 amino acids in the predicted loop regions. To determine if *H. influenzae* NTHi1465 encoded a lysine transporter, NTHi1465 was expressed in the *E. coli lysP* mutant from pPJ108 and tested for its ability to restore lysine transport. As expected, heterologous expression of NTHi1465 restored the ability of the *E. coli lysP* mutant to transport lysine (Fig. 2.6A), indicating that *H. influenzae* NTHi1465 encodes a L-lysine transporter. In addition, NTHi1465 appears to be the only lysine transporter in *H. influenzae* as deletion of *lysT* eliminated the ability of *H. influenzae* to transport lysine (Fig. 2.6B).

A.


Neisseria sp.
N. elongata
H. ducreyi
A. pleuropneumoniae
A. minor
M. haemolytica
H. parasuis
A. succinogenes
H. somnus
P. dagmatis
P. multocida
A. aphrophilus
A. actinomycetemcomitans
H. influenzae
B. pseudofirmus
B. halodurans
B. selenitireducens
O. iheyensis
E. sibiricum
Moritella sp.
P. profundum
P. damselae
V. angustum
G. hollisae
V. fischeri
A. salmonicida
Psychromonas sp.
P. ingrahamii
Vibrio sp.
V. cholerae
V. mimicus
V. shilonii
V. splendidus
L. hongkongensis
A. hydrophila
S. oneidensis
S. halifaxensis
S. pealeana
Shewanella sp.
S. amazonensis
S. piezotolerans
S. violacea
S. benthica
S. loihica
S. woodyi
S. sediminis
S. putrefaciens
S. baltica
S. frigidimarina
S. denitrificans
0.1

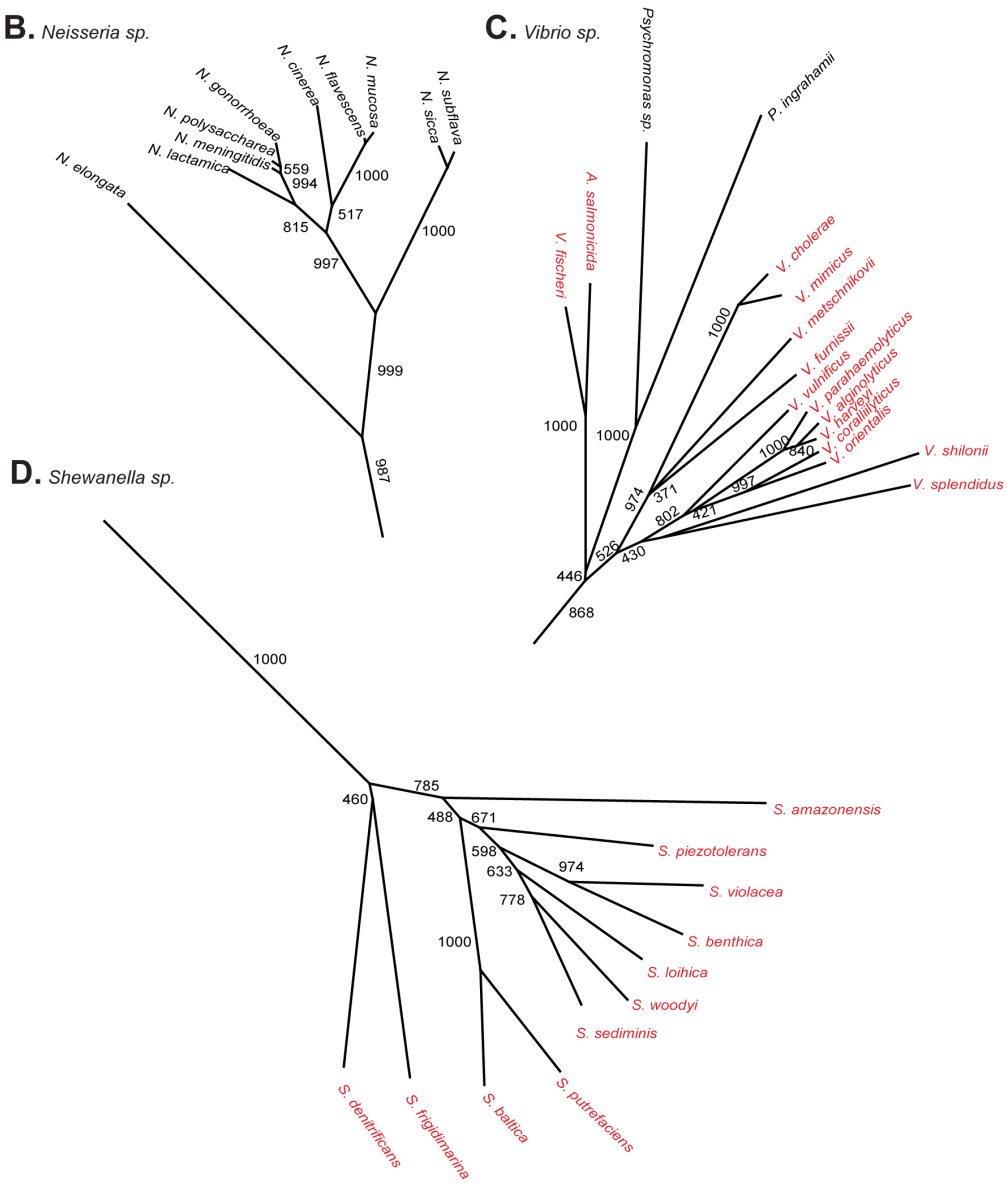


Figure 2.5 LysT phylogenetic tree.

(A) LysT homologs were identified by BLASTP analysis (e value cutoff < 10^{-100}), aligned, subjected to bootstrap analysis from a plurality of 1000 trees with ClustalW (version 1.83) [91]. The phylogenetic tree was drawn with Tree View software [92]. The bootstrap values are shown at branch points and the scale bar indicates the distance for 1 amino acid change in 10 residues. Species with predicted lysine riboswitches in the region 400 bp upstream of *lysT* are noted in red text. **(B)** Two-fold enlargement of the *Neisseria sp*. clade, **(C)** the *Vibrio sp*. clade, **(D)** the *Shewanella sp*. clade.

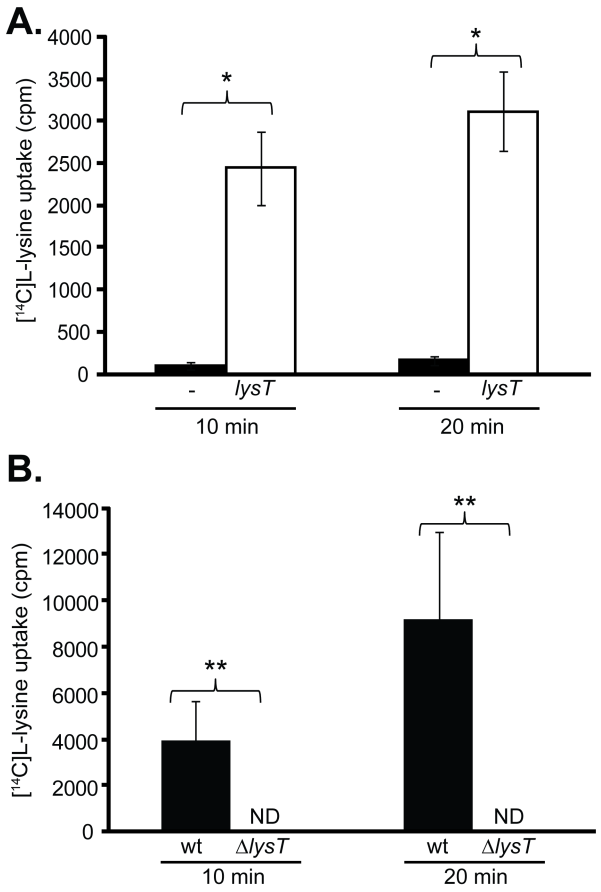


Figure 2.6 The *H. influenzae lysT* homolog NTHi1465 encodes a L-lysine transporter.

(A) *H. influenzae lysT* restores lysine transport in an *E. coli* lysine transport mutant. Transport of [^{14}C]L-lysine by the *E. coli lysP* mutant with an empty vector (-, black) or a vector expressing *H. influenzae lysT* from pPJ018, (*lysT*, white) was measured in counts per minute at 10 and 20 minutes. **(B)** NTHi1465 encodes the only lysine transporter in *H. influenzae*. Transport of [^{14}C]L-lysine by wild type *H. influenzae* (wt, black) and the *H. influenzae* Δ*lysT* (Δ*lysT*, white) mutant was measured in counts per minute at 10 and 20 min after the addition of [^{14}C]L-lysine to the assay. "ND" indicates that no lysine uptake was detectable. The data are the mean of three independent experiments, error bars in all assays represent the standard error, and statistical significance was determined by an unpaired two-tailed Student's t-test: *$P < 0.05$ and **$P < 0.005$.

2.3.8 *H. influenzae* lysine riboswitch

As previously discussed, bioinformatics also predicted a putative lysine riboswitch upstream and contiguous with *H. influenzae lysT*. Similar to *A. actinomycetemcomitans*, *hns* was located upstream of *lysT* and in the opposite orientation (Fig. 2.7A). Structural predictions of the *H. influenzae* lysine riboswitch aptamer domain using Mfold revealed conserved structural elements, as well as complete conservation of the nucleotides required for lysine binding (Fig. 2.7B) [87]. Although predicted to be structurally similar and contiguous with their respective *lysT* genes, the *A. actinomycetemcomitans* and *H. influenzae* riboswitches are localized to different positions in regards to the *lysT* start codons. In *A. actinomycetemcomitans*, the riboswitch is predicted to terminate approximately 50 bp upstream of the start codon for *lysT*; however, in *H. influenzae* the riboswitch is predicted to overlap the start codon of *lysT*.

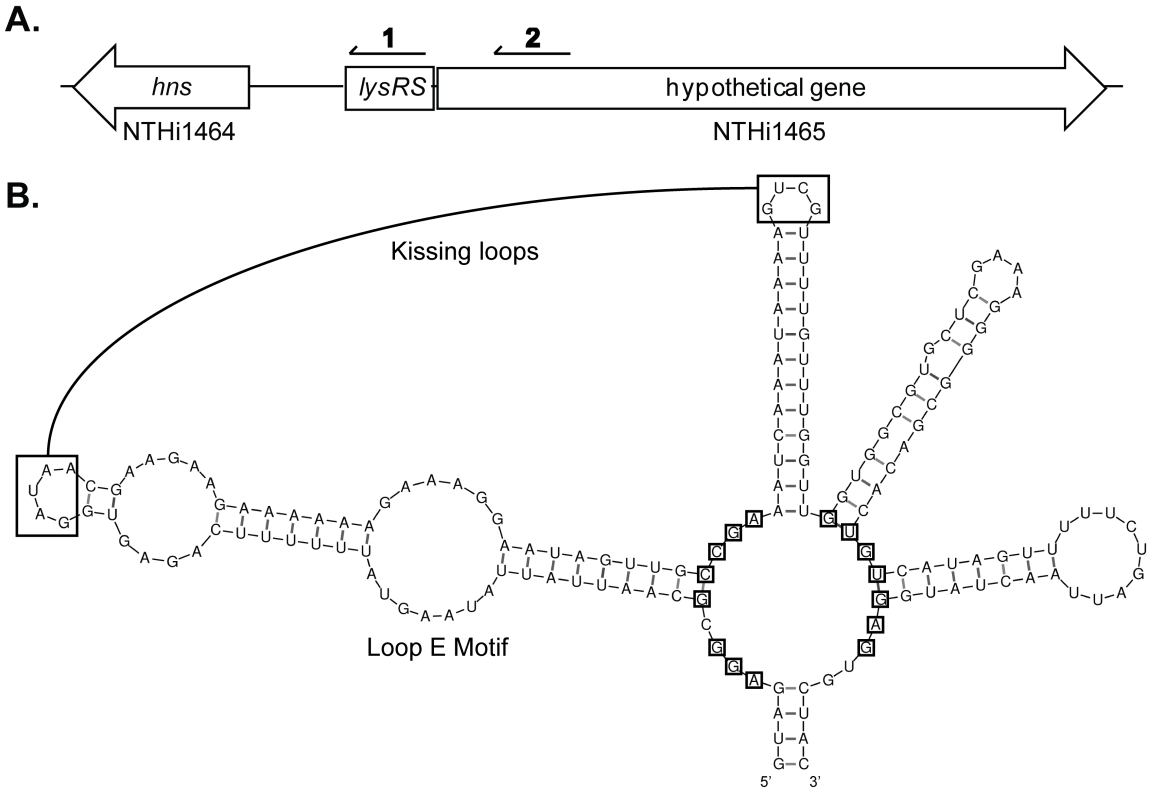


Figure 2.7 Lysine riboswitch and *lysT* homolog in *H. influenzae*.

(A) Genomic context of the *H. influenzae lysT* homolog (NTHi1465) and its lysine riboswitch. RNA probes for Northern blot analysis are depicted as arrows 1 and 2. **(B)** Predicted secondary structure of the *H. influenzae* lysine riboswitch aptamer. Mfold was used to predict RNA structure [87]. Structural motifs are noted and predicted lysine-binding nucleotides are boxed.

To determine if the *H. influenzae* riboswitch acts at the transcriptional level, *H. influenzae* were grown with and without L-lysine, and Northern blot analysis was used to probe for both the riboswitch and *lysT*. Similar to *A. actinomycetemcomitans*, two transcripts were present when using the riboswitch-specific probe: a small transcript corresponding to the lysine riboswitch and a larger transcript likely corresponding to the degraded riboswitch-*lysT* transcript (Fig. 2.8A). As expected, the lysine riboswitch accumulated upon addition of L-lysine, while the larger transcript disappeared. Since we were not able to definitively detect the full length transcript with the riboswitch probe, we also utilized a probe specific for *lysT*. This probe detected a specific mRNA at ~2000 nt, corresponding to the full-length *lysT* mRNA. Also, the probe only hybridized to the full-length *lysT* transcript in the lysine-starved samples, and did not detect any transcripts in cells exposed to lysine. However, as expected from Northern blots with riboswitch-specific probes (Fig. 2.2C and Fig. 2.8A), it appears that the full length transcript is highly unstable as a number of smaller transcripts were also detected (Fig. 2.8B). Importantly, no hybridization was observed when RNA from the *H. influenzae lysT* deletion strain was used (Fig. 2.8A&B), indicating that the transcripts identified in the wild-type RNA samples are specific for the riboswitch and *lysT*. Together these results demonstrate that the lysine riboswitch is conserved in *H. influenzae* and regulates *lysT* at the transcriptional level.

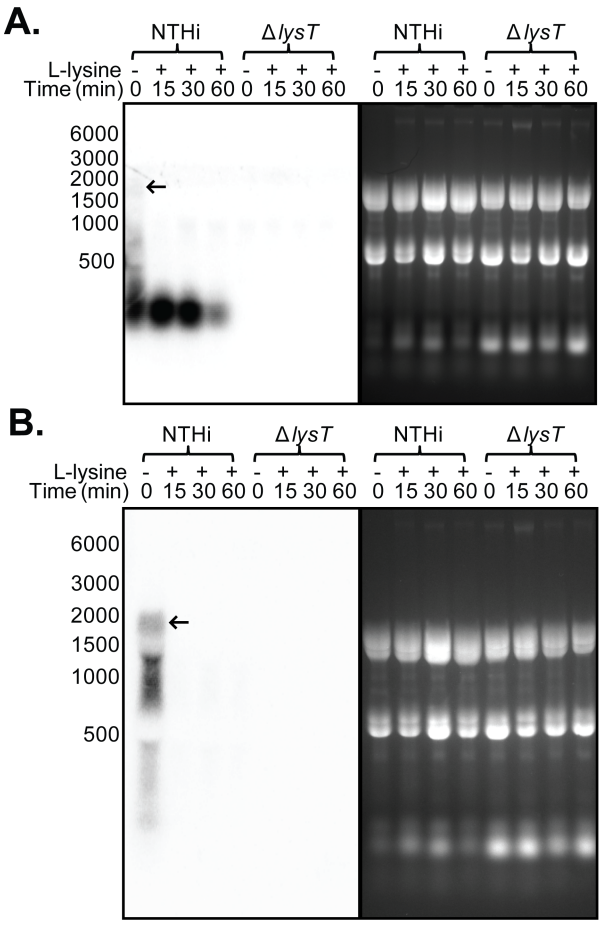


Figure 2.8 The lysine riboswitch is conserved in *H. influenzae*.

(A) The NTHi1465 lysine riboswitch accumulates in the presence of lysine. *H. influenzae* wild-type and Δ*lysT* lysine-starved cells (- lys, 0 min) were exposed to lysine (+ lys) and samples removed at 15, 30, and 60 min post-addition. RNA from each timepoint was subjected to Northern blot analysis using a probe for the NTHi1465 L-lysine riboswitch. **(B)** *H. influenzae lysT* diminishes upon addition of lysine. The RNA samples from the lysine starvation experiment describe above were subjected to Northern blot analysis and probed for *lysT*. Arrows indicate the full-length *lysT* transcript in both panels.

2.4 DISCUSSION

The study describes identification of 12 new ncRNAs in *A. actinomycetemcomitans*. While most of these ncRNAs were produced by planktonic and biofilm bacteria, several appear to be produced at higher levels in biofilm grown bacteria. Interestingly, three ncRNAs that are induced in biofilm bacteria have predicted functional overlap, specifically they appear to regulate amino acid transport. Along with the lysine riboswitch characterized in this study, a putative glycine riboswitch was identified upstream of AA00167 (Table 2.3, Fig. 2.9). The riboswitch has structural and sequence homology to other glycine riboswitches [72], and AA00167 belongs to the sodium/alanine symporter superfamily [112] which includes a known glycine transporter [113]. We also observed high levels of an sRNA homologous to GcvB in *A. actinomycetemcomitans* biofilms. In *Salmonella enterica* serovar Typhimurium and *E. coli*, GcvB regulates more than 10 genes involved in amino acid transport [114-117]. Since all three of these ncRNAs are proposed to down-regulate amino acid transport in the presence of high levels of intracellular amino acids, these data suggest that within the colony biofilm, *A. actinomycetemcomitans* possesses high intracellular amino acid concentrations. In the gingival crevice, *A. actinomycetemcomitans* and other bacteria are known to release proteases that would increase free amino acid pools in the surrounding environment [118, 119]. However, free amino acid levels have not been directly measured in gingival crevicular fluid. Further studies will be needed to determine if these ncRNAs are important for biofilm formation and biofilm-specific phenotypes.

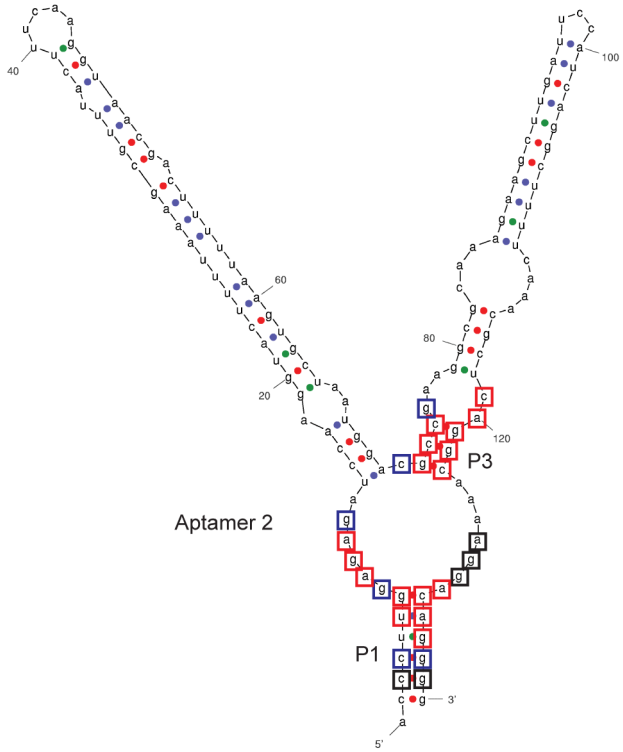


Figure 2.9 Glycine riboswitch structural prediction.

The RNA secondary structure prediction for the *A. actinomycetemcomitans* glycine riboswitch aptamer determined by RiboSW and Mfold analyses [87, 93]. Nucleotides conserved in 75% (black boxes), 90% (blue boxes), and 97% (red boxes) of identified glycine riboswitches are indicated on the structure [72].

While it is not surprising that a lysine riboswitch controls transcription of a gene encoding a lysine transporter, the characterization of LysT describes a new family of lysine transporters. LysT homologs were identified in 64 species of bacteria, including Gram-negative and Gram-positive species (Fig. 2.5), and our studies of *H. influenzae* LysT suggest (Fig. 2.6) that many of the LysT homologs will be involved in lysine transport. Interestingly, as in *A. actinomycetemcomitans* and *H. influenzae*, riboswitches are predicted to regulate transcription of 75% of the LysT homologs. This latter observation suggests that riboswitch regulation is a common mechanism for regulating transcription of lysine transporters in bacteria. This mode of regulation allows transcription to be linked to intracellular lysine concentrations; thus allowing bacteria to fine-tune expression in real-time in response to changing lysine levels.

While not examined in this study, lysine riboswitches could also play important regulatory roles aside from regulating genes involved in lysine transport and biosynthesis. Because our Northern blot analyses showed that the lysine riboswitch has a relatively long half-life of approximately 17 min (Fig. 2.3), the lysine riboswitch could act as a *trans*-regulatory sRNA in a lysine-dependent signaling process. Interestingly, Phan and Schumann noted a similar result in *Bacillus subtilis* [96], and also postulated that the prematurely terminated lysine riboswitch might serve additional functions within the cell. This is not unreasonable as the S-adenosyl methionine riboswitch in *Listeria monocytogenes* is known to regulate an important virulence regulator *in trans* [120].

This study used bioinformatic and experimental approaches to identify ncRNAs in *A. actinomycetemcomitans,* some of which appear to be highly expressed in biofilms and

others which have no known homologs (Aa35, Aa84, Aa337, and Aa575). These results provide the first insight into ncRNAs in *A. actinomycetemcomitans,* providing a basis for future functional studies. There is no doubt that more ncRNAs are produced by *A. actinomycetemcomitans,* many of which may be specific for other growth environments such as the oral cavity or in polymicrobial communities. Future studies will address the overall expression of ncRNAs in *A. actinomycetemcomitans* and determine specific regulatory functions that ncRNAs serve, not only within a single bacterium, but within complex polymicrobial communities.

Chapter 3. An Evolutionary Link between Natural Transformation and CRISPR Adaptive Immunity[2]

3.1 INTRODUCTION

Natural transformation is a primary means of bacterial horizontal gene transfer (HGT). Natural transformation occurs when environmental cues trigger the expression of competence genes, which allow the bacterium to take up DNA from its environment and incorporate new genes into its genome. Bacterial competence systems can vary among species, but all systems involve DNA transport proteins and machinery that incorporates the new DNA into the chromosome via homologous recombination [50-57]. Due to their genetic complexities, competence systems are thought to be ancestral traits in many bacterial lineages; however, competence loss within bacterial families is common [49, 121]. Evolutionary models predict that HGT and homologous recombination by natural transformation can accelerate adaptation to new environments and allow organisms to more rapidly reach fitness equilibria [59, 122]. Despite predictions that competence is a beneficial trait, many bacteria are non-competent and, among competent bacteria, competence genes are sometimes lost by strains within a species [48, 49]. This leads to the hypothesis that throughout evolution naturally competent bacterial populations are constantly giving rise to non-competent siblings, and at certain points, non-competent genomes are selected for over the dynamic genomes of competent bacteria.

To test this hypothesis, we focused on the evolutionary history of the opportunistic human periodontal pathogen *Aggregatibacter actinomycetemcomitans*, in which ~30% of isolated strains are naturally competent [48]. *A. actinomycetemcomitans*

[2]This Chapter was adapted from the following reference (185, © American Society for Microbiology, mBio, 3: e00309-12).

is a member of the *Pasteurellaceae*, a family of bacteria predicted to have descended from a competent ancestor [121]. An experimental survey testing natural transformation demonstrated that *A. actinomycetemcomitans* competence loss is clonal [48], suggesting that non-competent strains arose and radiated into non-competent lineages. The persistence of non-competent *A. actinomycetemcomitans* strains during evolution allows us to utilize genome sequence information to delineate the evolutionary history of competence loss and genome stability. Using comparative genomics we identify that competence loss has occurred multiple times during *A. actinomycetemcomitans* evolution in parallel. Compared to competent strains, non-competent *A. actinomycetemcomitans* have smaller genomes with fewer rearrangements, as demonstrated by whole genome alignments. Pairwise strain average nucleotide identity (ANI) calculations reveal that competence loss is tied to evolutionary divergence among *A. actinomycetemcomitans* strains. Moreover, the loss of competence is directly correlated with the loss of *A. actinomycetemcomitans* CRISPR-*cas* adaptive immune systems, resulting in non-competent strains with more parasitic genetic elements than competent siblings. In non-competent strains, remnants of CRISPR-*cas* systems that remain appear to have been co-opted for *A. actinomycetemcomitans* self-gene regulation, similar to eukaryotic microRNAs. Together these results support a model for *A. actinomycetemcomitans* evolution whereby competence and non-competence are means to genetic diversity and stability, respectively.

3.2 RESULTS AND DISCUSSION

3.2.1 Parallel loss of competence by common mechanisms

Because the *Pasteurellaceae* common ancestor is predicted to be competent, we sought to determine the genetic nature of competence loss throughout *A. actinomycetemcomitans* evolution. Therefore it was important to generate an accurate evolutionary history for the species and map competence loss to this phylogeny. The alignment of concatenated core genes can be used to accurately determine phylogenetic relationships among bacteria [123]. Core genes were aligned from 17 sequenced *A. actinomycetemcomitans* genomes, including 14 previously sequenced strains [44, 124-126] and 3 genomes sequenced in this study (Table 3.1). The concatenated core gene super alignment was subjected to maximum likelihood phylogenetic analysis to determine evolutionary relationships between strains, revealing the presence of three major *A. actinomycetemcomitans* lineages (Fig. 3.1). Among these 17 strains, 7 have previously been subjected to natural transformation assays and 2 were found to be competent [48, 127]. Subsequent alignment of multiple individual competence genes from the 17 sequenced *A. actinomycetemcomitans* genomes reveals that this bacterial lineage derived from a competent ancestor, and throughout evolutionary history several non-competent lineages have arisen via inactivation or deletion of genes critical for DNA uptake and incorporation on the chromosome (Fig. 3.1). The results indicate the presence of 3 competent and 14 non-competent sequenced strains, which reflects the prevalence of non-competent strains in larger *A. actinomycetemcomitans* populations [48]. Importantly,

the bioinformatics predictions for competent and non-competent strains correlate perfectly with results of previously published natural transformation assays. Overall, non-competence has arisen on at least six independent occasions (Fig. 3.1) with the most common mutation (3 occasions) occurring via a site-specific mobile DNA element inserting into *comM*, whose gene product is important for efficient incorporation of DNA onto the chromosome following uptake [55]. This pattern of *A. actinomycetemcomitans* competence loss contrasts to other naturally competent bacteria like *Haemophilus influenzae*, in which competence loss is common but occurs randomly throughout evolution and typically by many unique mechanisms [49, 58]. Because *A. actinomycetemcomitans* competence loss occurred multiple times during evolution in more than one lineage, these data support a model for the parallel evolution of competence loss. While parallel evolution is indicative of natural selection, this does not mean that competence loss is adaptive; in fact it may be neutral, but the genome content of the stably maintained non-competent strains is presumably beneficial.

Table 3.1 ***A. actinomycetemcomitans*** **genome sequencing statistics for** ***de novo*** **genome assembly and read-mapping to complete reference genomes.**

de novo	***Aa* 624**	***Aa* VT1169**	***Aa* Y4**
454 reads (bp)	81,529,600	10,992,296	33,150,533
454 mate-paired reads	257,891	16,220	63,629
454 coverage	35.8X	5.1X	15.5X
454 contigs	121	1,250	145
454 N50 contig size	68,926	1,514	40,197
454 scaffolds	19	80	8
454 N50 scaffold size	330,558	32,073	527,427
454 scaffold total bp	2,279,748	1,873,150	2,142,003
SOLiD reads (bp)	438,962,055	320,303,340	499,721,075
SOLiD coverage	191X	147X	233X
read mapping			
best reference	D7S-1	HK1651	HK1651
% coverage	99	97	97
fold-coverage	25.6X	21.9X	30.4X

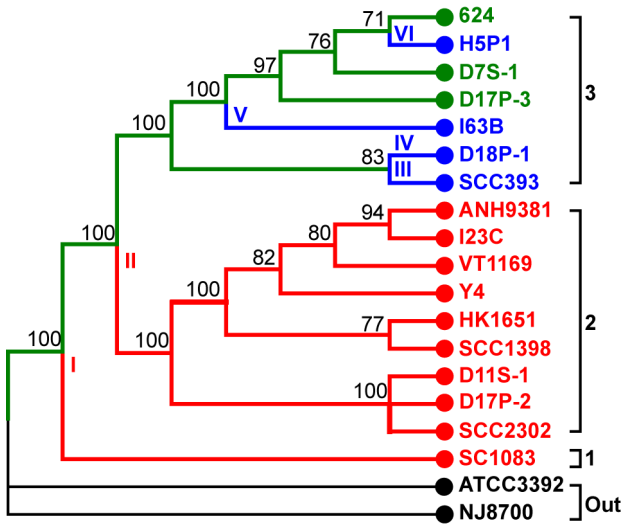


Figure 3.1 Competence loss occurs throughout *A. actinomycetemcomitans* evolution.

The evolutionary history of 17 *A. actinomycetemcomitans* strains was determined from the alignment of 30 concatenated core genes by the Maximum Likelihood method based on the Tamura-Nei model with 100 bootstrap replicates, using *A. aphrophilus* ATCC3392 and NJ8700 as the out group (Out) [128-131]. Competence is traced with green branches. Red branches represent ancestral non-competent lineages while blue branches are recent non-competent strains. Roman numerals indicate non-competence branch points. Bootstrap values are above the branches and branches reproduced in less than 70% of bootstrap replicates are collapsed.

3.2.2 Divergence and speciation of non-competent strains

Competence is a primary mechanism of HGT and DNA acquisition in bacteria [132] and as expected, the genomes of competent *A. actinomycetemcomitans* strains are on average 200,000 bp larger than non-competent strains (2.3 Mb vs. 2.1 Mb, respectively; $P = 9 \times 10^{-5}$, two-tailed unpaired t-test). A major consequence of competence loss is the inability to acquire new traits via genomic exchange with related species, leading to the hypothesis that competence loss leads to genetic isolation, which could ultimately result in speciation. To test this hypothesis, we analyzed non-competent and competent genomes for genomic rearrangements and average nucleotide identity (ANI) across whole genomes. Our results revealed that non-competent strains have similar syntenic genomes (i.e. similar gene arrangement), while the competent strains have dynamic genomes with multiple rearrangements (Fig. 3.2). Further differences were revealed by ANI. Based on a proposed modern species definition (ANI > 95%) [133] the non-competent strain with the most ancient branch point (lineage 1, Fig. 3.1) is a different species than strains within lineage 2 and lineage 3 (Fig. 3.3). Interestingly, ANI comparisons of strains within lineage 2 and within lineage 3 show that strains within these lineages are ~99% identical; however comparisons of strains between lineages 2 and 3 reveal that they differ significantly (ANI ~97%, $P < 0.001$). Thus lineage 2 and 3 strains have clearly begun to diverge into distinct populations. This supports a model in which competence loss leads to genetic isolation and ultimately divergence in *A. actinomycetemcomitans*.

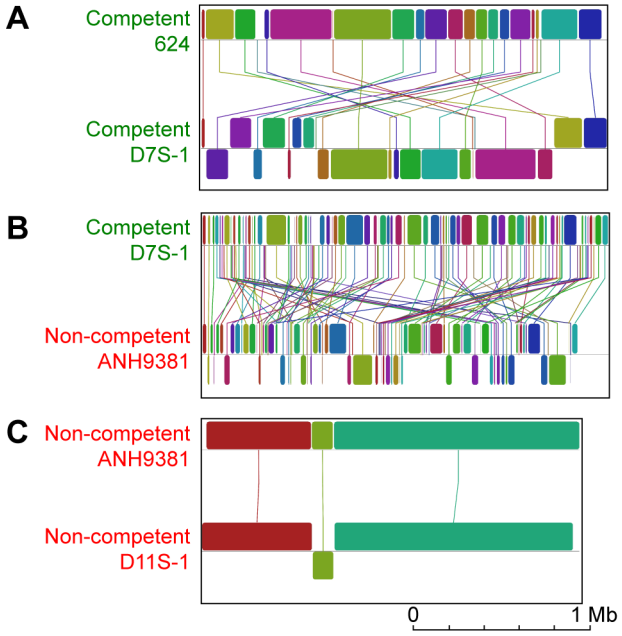


Figure 3.2 Non-competent strains have highly syntenic genomes and competent genomes have multiple rearrangements.

Whole genome alignments of two competent strains **(A)**, a competent and non-competent strain **(B)**, and two non-competent strains **(C)**. Blocks represent collinear genomic segments and lines indicate rearrangements.

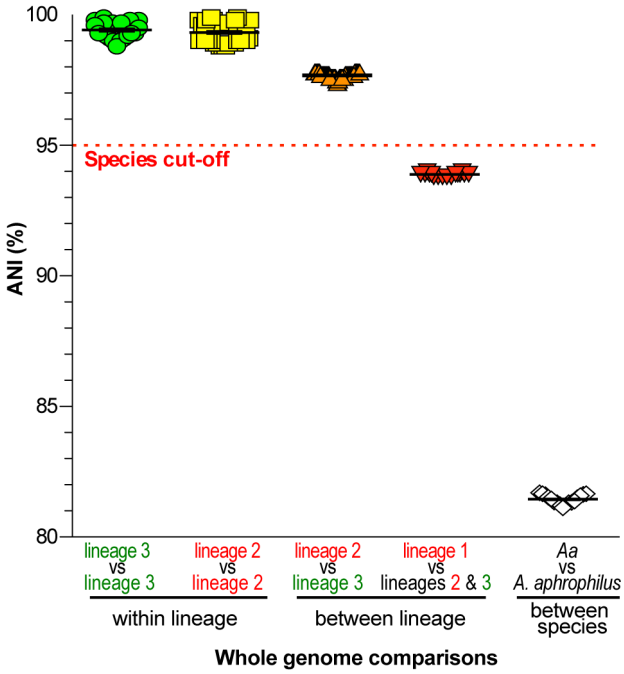


Figure 3.3 Speciation of non-competent *A. actinomycetemcomitans*.

Pairwise whole genome average nucleotide identities (ANI) within *A. actinomycetemcomitans* lineages (lineage 3 vs lineage 3, $n = 21$; lineage 2 vs lineage 2, $n = 36$), between *A. actinomycetemcomitans* lineages (lineage 2 vs lineage 3, $n = 63$; lineage 1 vs lineages 2 & 3, $n = 16$), and between *A. actinomycetemcomitans* and *Aggregatibacter aphrophilus* (*Aa* vs *A. aphrophilus*, $n = 17$). Each point represents a pairwise genome comparison. Bars indicate the mean, and error bars are standard error of the mean. ANI between lineages are significantly different (ANOVA test, $P < 0.001$), while within lineage comparisons are not.

3.2.3 Non-competent strain diversity due to mobile genetic elements

Despite the fact that non-competent genomes had high synteny and ANI, some differences among non-competent genomes were observed. We hypothesized that non-competent strains of *A. actinomycetemcomitans*, having eliminated a predominant pathway for HGT, would utilize alternative mechanisms for HGT to maintain evolutionary fitness. Indeed, mathematical modeling suggests that in the absence of HGT and homologous recombination, genetic drift resulting from the accumulation of random neutral mutations will ultimately lead to decreased fitness [59], indicating that at least some nominal level of HGT is required for a strain to remain fit. Transducing phage and conjugative plasmids are also sources of HGT, and on average there were ~5 unique plasmid and phage related DNA elements in each non-competent genome, significantly more than that observed in competent lineages which had ~2 (two-tailed unpaired t-test, $P = 0.02$). The unique DNA elements in the non-competent strains include 2 self-replicating plasmids, 14 integrated plasmid elements, and 22 prophage elements. Thus, while non-competent strains possess increased synteny and smaller genomes compared to competent strains (Fig. 3.2), the observed differences in DNA content in these strains are almost entirely due to incorporation of plasmid and prophage DNA.

3.2.4 Non-competent strains have compromised CRISPR adaptive immune systems

The presence of plasmids and prophage in the non-competent strains suggest that their defenses against foreign nucleic acids were potentially compromised. Bacterial adaptive immune systems called clustered regularly interspaced short palindromic repeats (CRISPRs) have evolved to protect against phage and plasmid-mediated HGT [32, 134, 135]. Many variations of CRISPRs exist, and they have been extensively reviewed [136,

137]. Briefly, CRISPRs are genetic loci consisting of direct repeats and spacers [33]. Spacers are often derived from phage and plasmid DNA [34]. During an infection, short CRISPR RNAs (crRNAs) are transcribed from the CRISPR loci and base pair with foreign (plasmid and bacteriophage) DNA via spacer encoded sequences [134]. Both the invading DNA and crRNA are cleaved by proteins encoded by CRISPR-associated *cas* genes, thereby eliminating the phage and plasmid DNA [32, 135, 138, 139]. Recent work has shown that bacteria readily lose CRISPRs under certain selective conditions. Specifically, enterococci lost CRISPRs allowing the acquisition of parasite-encoded antibiotic resistance genes and *Mycoplasma gallisepticum* lost CRISPRs following the shift to a new avian host [140, 141]. Because CRISPR loss is common in other organisms, we predicted that non-competent *A. actinomycetemcomitans* were more susceptible to phage and plasmids due to compromised CRISPR-mediated immunity. To test this prediction, we examined the CRISPR-*cas* content of competent and non-competent *A. actinomycetemcomitans*. Based on phylogeny, the common ancestor for *A. actinomycetemcomitans* possessed two CRISPR-*cas* systems, CRISPR1-*cas* and CRISPR3-*cas* duplicated from a single system in a more ancient ancestor (Fig. 3.4). Indeed, *Aggregatibacter aphrophilus* has only one CRISPR-*cas* system related to *A. actinomycetemcomitans* CRISPR1-*cas* and CRISPR3-*cas*, supporting the hypothesis that they were duplicated from an ancient system. The competent *A. actinomycetemcomitans* strains in our study contain both of these ancient CRISPR-*cas* systems and have acquired a third intact CRISPR-*cas* system (Fig. 3.4). In contrast, the CRISPR-*cas* systems in ancient non-competent lineage 1 and 2 strains have lost *cas* genes to deletion and/or acquired nonsense mutations in *cas* genes (Fig. 3.5A). Based on these mutations CRISPR elements in non-competent lineage 1 and 2 strains are predicted

to be non-functional defensively. CRISPR-*cas* loss was evident even in the most recently evolved (lineage 3) non-competent strains, indicating that loss of CRISPR-*cas* occurs quickly within evolutionary time. The most parsimonious evolutionary history is that competence loss preceded CRISPR-*cas* loss among *A. actinomycetemcomitans* strains. This inference can be drawn because lineage 2 strains share the same disruption of the *comM* competence gene; however, they have many different CRISPR-*cas* mutations and deletions. These data indicate that the emergence of non-competence in *A. actinomycetemcomitans* is correlated with the loss of CRISPR based adaptive immunity.

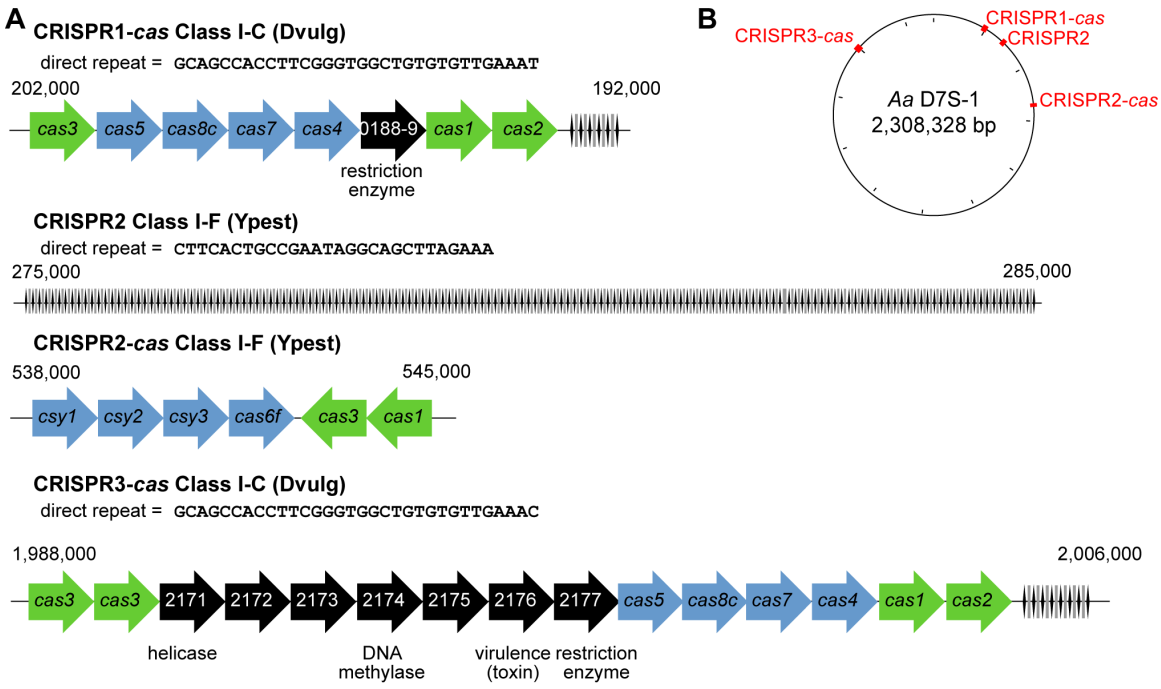


Figure 3.4 The representative *A. actinomycetemcomitans* CRISPR-*cas* systems in D7S-1.

(A) Representations include the *cas* genes (arrows) and CRISPRs (black diamonds = direct repeats, grey rectangles = spacers). CRISPR-*cas* locus names and types are indicated above each map. Approximate genomic coordinates of CRISPR-*cas* loci are indicated at right and left of each map. Direct repeat sequences are indicated above respective maps. Numbers within arrows represent the *A. actinomycetemcomitans* D7S-1 ORF number (e.g. "0188" represents D7S_0188). Green arrows represent core *cas* genes found in all loci, blue arrows represent subtype-specific *cas* genes, and variable genes not previously associated with CRISPRs are shown as black arrows. (**B**) Genome map with annotated CRISPR-*cas* loci. Boxes indicate locations of individual CRISPR-*cas*, names are shown, as well as CRISPR-*cas* class type.

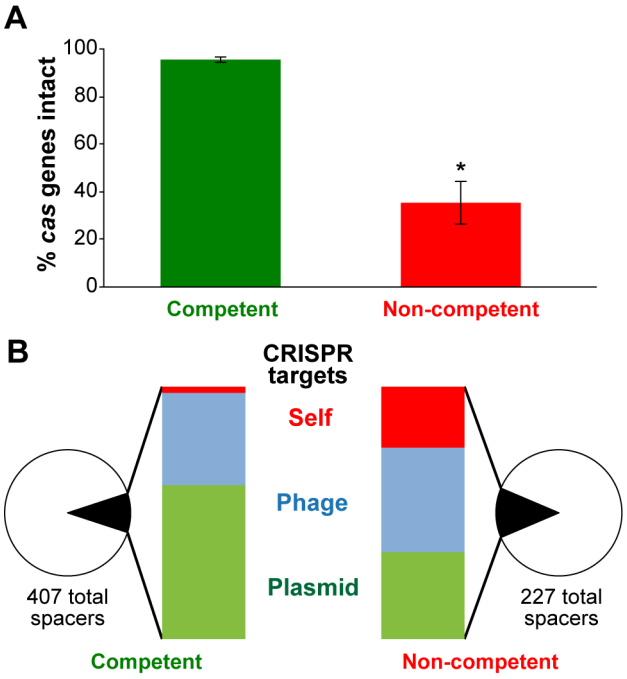


Figure 3.5 CRISPR-*cas* loss and the evolution of self-targeting CRISPR spacers in non-competent *A. actinomycetemcomitans*.

(A) Non-competent strains (red, $n = 14$) have a lower percentage of intact *cas* genes than competent strains (green, $n = 3$) (*two-tailed unpaired t-test, $P = 0.009$). Error bars represent standard error of the mean. **(B)** Non-competent strains' CRISPRs are enriched for self-targeting spacers. Spacer targets were determined by BLASTn analysis [142]. The white wedge represents spacers with unknown targets, the black wedge represents spacers with known targets. The bar graphs represent the proportion of spacers that are known to target self genes (red), phage (blue), plasmids (green).

3.2.5 Evolution of CRISPR self gene regulation in non-competent strains

Despite the fact that non-competent strains have lost many *cas* genes required for immunity, most of these strains maintain one CRISPR with a small number of spacers. A current hypothesis is that "broken" CRISPR-*cas* loci tend toward autoimmunity and acquire self-targeting spacers [35] that are used to control expression of chromosomal genes. To test this hypothesis, we identified spacers in all strains and predicted their targets [90, 143]. Most spacers target unknown sequences (Fig. 3.5B). However, identifiable targets in competent strains were predominantly phage and plasmids, while strains in non-competent lineages were enriched for spacers with specificity for *A. actinomycetemcomitans* genes (Fig. 3.5B). Some of these self-targeting spacers within the non-competent strains are antisense to coding genes, indicating that they could regulate self gene expression similar to that observed for an orphan crRNA in *Listeria monocytogenes* [144]. These data suggest that non-competent *A. actinomycetemcomitans* strains are co-opting CRISPRs for *A. actinomycetemcomitans* chromosomal gene regulation, reminiscent of eukaryotic microRNAs that have been co-opted from the RNA interference immune surveillance system in higher eukaryotes [145]. To test whether the self-targeting CRISPRs could be used for self-gene regulation, we used Northern blot analysis to probe for the self-targeting crRNA in *A. actinomycetemcomitans* VT1169. Similar to the orphan CRISPR in *L. monocytogenes*, the *A. actinomycetemcomitans* self-targeting spacer is primarily being expressed during exponential phase growth as part of an unprocessed ~650 nt crRNA (Fig. 3.6A). The VT1169 crRNA is antisense to *glgP*, glycogen phosphorylase, and is predicted to bind with perfect complementarity across 35 nt near the 3' end of the *glgP* mRNA (Fig. 3.6B). Further experiments will be required to

determine whether crRNA-based gene regulation impacts mRNA stability or translation, and if there is a fitness benefit to this recently evolved regulation.

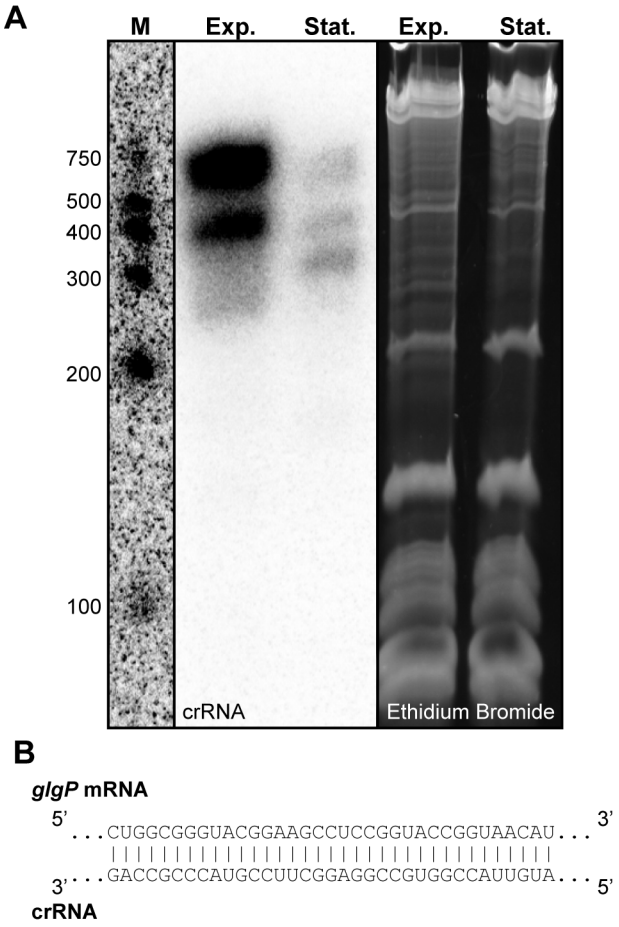


Figure 3.6 *A. actinomycetemcomitans* VT1169 self-targeting crRNA is expressed and predicted to target glcyogen phosphorylase mRNA.

(A) The self-targeting *A. actinomycetemcomitans* VT1169 crRNA is expressed primarily as an unprocessed transcript. Cells were grown to exponential phase (Exp.) and stationary phase (Stat.), total RNA was harvested and subjected to Northern blot analysis with a probe for the self-targeting crRNA. The ethidium bromide stained rRNA (right panel, top 2 bands) demonstrates equal RNA loading and RNA integrity. The radiolabeled RNA Century Marker-Plus (M) is overexposed at left to serve as the size standard. Northern blot is representative of three biological replicates. **(B)** Predicted binding interaction between *A. actinomycetemcomitans* VT1169 crRNA and the *glgP* mRNA.

3.2.6 An evolutionary model for the effects of competence and CRISPRs on fitness

We have discovered an instance where bacterial speciation is correlated with competence loss and the subsequent loss of bacterial adaptive immunity. While several models take into account our results we propose an evolutionary model where competence-mediated HGT promotes *A. actinomycetemcomitans* genetic diversity, while non-competence promotes genome stability. Based on this model, dynamic competent bacterial genomes quickly drive evolution, benefiting the organism in changing environmental conditions. However, non-competent siblings constantly arising in the population are selected for when the organisms have beneficial genome content in a relatively constant environment. One potential consequence of the concomitant loss of CRISPR-*cas* immunity by these non-competent strains is that they can sample and adjust to small changes in the environment through phage and plasmid-mediated HGT. An important component of this model is that competent bacteria have more dynamic genomes than non-competent strains, consistent with our observations that competent bacteria have larger genomes with increased genome rearrangements (Fig. 3.2). While this is an appealing model, the selective force(s) that drive loss of competence and CRISPR are not known and may not be linked. An alternate model that explains the link between CRISPRs and competence is the possible increased susceptibility of competent bacteria to genetic parasites relative to non-competent siblings. This is supported by the presence of large amounts of extracellular DNA in bacterial biofilms [61, 146], which potentially contains parasitic DNA from lysed neighboring cells. This hypothesis is bolstered by recent elegant experiments in *Streptococcus pneumoniae*, whereby an artificial CRISPR-*cas* system prevented natural transformation of a CRISPR-targeted capsule gene into a non-virulent rough strain, indicating the CRISPR-*cas* is capable of

preventing natural transformation [147]. Regardless, the evolutionary history of *A. actinomycetemcomitans* reveals a strong correlation between the evolution of competence loss and CRISPR adaptive immunity and provides new insights into mechanisms of bacterial diversity and speciation.

3.3 MATERIALS AND METHODS

3.3.1 DNA Isolation, genome sequencing, and genome assembly

A. actinomycetemcomitans 624, Y4, and VT1169 were routinely grown in tryptic soy broth supplemented with 0.5% yeast extract, and DNA was purified using standard methods for genome sequencing [148]. *A. actinomycetemcomitans* 624, VT1169, and Y4 genomes were sequenced and assembled *de novo* from Roche/454 FLX pyrosequencing reads using the Newbler assembler [149] (Table 3.1). Life Technologies SOLiD V4 and 454 sequencing reads from *A. actinomycetemcomitans* 624, VT1169, and Y4 were aligned to the complete *A. actinomycetemcomitans* reference genomes (*A. actinomycetemcomitans* HK1651, D7S-1, and D11S-1) using CLC Genomics Workbench software (Table 3.1). Supplemental sequencing of VT1169 CRISPR1 was carried out by capillary sequencing of a PCR product (C1-For 5'-ACGCAAATTCACACCCAC-3' and C1-Rev 5'-TGGATGGTTTTGAGTGAC-3').

3.3.2 Multiple sequence alignments, ANI calculations, and phylogenetic tree construction

Complete and draft genome sequences for all *A. actinomycetemcomitans* strains were generated in this study or downloaded from Genbank and Oralgen (http://www.oralgen.lanl.gov/_index.html) [44, 124-126]. When complete gene sequences were not found in the *de novo* genome assembly of VT1169, consensus gene sequences were inferred from alignment to HK1651. The core gene phylogenetic tree was constructed from the sequence alignment of 30 concatenated core genes (*pyrG*, *rplA*, *rplB*, *rplC*, *rplD*, *rplF*, *rplK*, *rplM*, *rplP*, *rplS*, *rpsC*, *rpsE*, *rpsI, rpsJ*, *rpsM*, *rpsS, smpB*, *tsf*, *frr, rplE*, *rpoB*, *rpsB*, *dam, dnaN*, *dnaQ*, *holB*, *holD, recR*, *rpoA*, and *rpoH*) by the Maximum Likelihood method using the Tamura-Nei model in MEGA5 using MUSCLE with 100 bootstrap replicates [128-131]. Whole genome alignments were performed with progressive MAUVE to identify genome rearrangements, local collinear genomic blocks, and variable mobile elements [150]. ANI values for pairwise genome comparison were calculated using JSpecies with BLAST (ANIb) and MUMMER (ANIm) [133]. Individual competence genes were aligned with MUSCLE in MEGA5 [128, 129, 131]. Whole *cas* operons were aligned with CLC Sequence Viewer. The consensus gene sequences from VT1169 sequencing read alignment to HK1651 were used when genes were unassembled *de novo*.

3.3.3 CRISPR detection and characterization

CRISPRs were identified with CRISPRfinder [151]. CRISPR spacers were subjected to BLASTn analyis against the nr/nt database and hits with > 94% identity

were recorded [152]. Hits were classified as directed toward plasmid, prophage, phage, or self, while negative results were classified as unknowns.

3.3.4 Northern blot analysis

A. actinomycetemcomitans VT1169 was grown to mid-exponential (OD_{600} 0.80) and stationary phase (OD_{600} 1.60) in TSBYE and total RNA was harvested using RNA bee (Tel-test), as described previously [153]. Northern blotting was carried out using previously described methods [153]. Briefly, 10 μg RNA was separated on an 8% polyacrylamide 8 M urea denaturing gel, stained with ethdium bromide, transferred to nitrocellulose, UV crosslinked, blocked for 1 h at 55°C in UltraHyb Buffer (Ambion), and probed for 16 h at 38.5°C with a radiolabeled DNA oligonucleotide anti-crRNA probe (5'GGCGGGTACGGAAGCCTCCGGCACC-3'). The radiolabeled RNA Century Marker-Plus (Ambion) was synthesized according to the manufacturer's protocol and served as a size standard. Prior to transfer, gels were stained with ethidium bromide and imaged with a G:BOX apparatus (Syngene) to visualize rRNA, which served as the loading control.

Chapter 4. Probing Bacterial Metabolism during Infection using High-Resolution Transcriptomics[3]

4.1 INTRODUCTION

Over 100 years ago, Louis Pasteur emphasized the importance of understanding the metabolic processes important for bacterial colonization and persistence during infection [154]. However, progress in this area has proven challenging in part due to the difficulty in characterizing the nutritional content of infection sites. One approach that has been successful in probing the growth environment of the infection site is transcriptomics using microarrays and RNA sequencing (RNA-seq). Microarrays have yielded insights into the carbon sources and metabolic pathways used by pathogens during infection [155-158], although these studies are technically challenging due to the large amounts of host RNA often present in disease samples. In contrast, high-throughput RNA-seq provides a robust tool for global gene expression analyses *in vivo* since it can be performed with low amounts of RNA, and bacterial and host gene expression can be separated computationally [159]. However to date, very few *in vivo* RNA-seq studies have been performed for mammalian pathogens. One study revealed significant insight into the *in vivo* physiology of *Vibrio cholera* during infection [160]; however, the undefined media used as an *in vitro* control limited the conclusions that could be drawn about the metabolic pathways active *in vivo*.

Our laboratory has studied nutrition in numerous pathogenic bacteria including the opportunistic human pathogen *Aggregatibacter actinomycetemcomitans* [14, 22, 161-

[3]This Chapter was adapted from the following reference (221, © American Society for Microbiology, Journal of Bacteriology, doi:10.1128/JB.00875-13).

165]. *A. actinomycetemcomitans* is a Gram negative, facultative anaerobe that resides in the oral cavity of human and old-world primates. Specifically, *A. actinomycetemcomitans* colonizes the subgingival crevice, defined as the area around the tooth bounded by the gingival epithelium on one side and the tooth surface on the other. *A. actinomycetemcomitans* is the proposed causative agent of localized aggressive periodontitis [166, 167], an acute disease characterized by massive tissue destruction and tooth loss. *A. actinomycetemcomitans* also causes extra-oral infections including abscess infections [63, 64, 168], and our laboratory has used a murine abscess infection model to characterize *A. actinomycetemcomitans* genes required for growth *in vivo* [22]. In this study, we utilized high-resolution transcriptomics to examine the physiology of *A. actinomycetemcomitans* during growth in the murine abscess. These analyses revealed significant insights into the transcriptional organization of the *A. actinomycetemcomitans* genome including transcription start sites (TSS) and non-coding RNAs (ncRNAs). In addition, numerous metabolic genes displaying enhanced transcription during *in vivo* growth were identified. Targeted mutagenesis of these differentially regulated genes revealed roles for fermentative metabolism and anaerobic respiration for *in vivo* persistence.

4.2 MATERIALS AND METHODS

4.2.1 Bacterial growth conditions

A. actinomycetemcomitans strain 624, a rough serotype A clinical isolate was used in this study. *A. actinomycetemcomitans* 624 was routinely cultured in brain heart

infusion (BHI) broth or tryptic soy broth supplemented with 0.5% yeast extract (TSBYE) shaking at 150 rpm in 5% CO_2 atmosphere at 37°C. For mutant *A. actinomycetemcomitans*, BHI and TSBYE were supplemented with 50 μg/ml spectinomycin. For RNA-seq, *A. actinomycetemcomitans* was grown in a chemically defined modified Socransky's medium supplemented with 20 mM glucose, 50 mM MOPS, pH 7.2 (CDM) [14, 73], shaking at 150 rpm in 5% CO_2 atmosphere at 37°C. Prior to biofilm growth, overnight liquid cultures of *A. actinomycetemcomitans* were diluted 1:2 with fresh CDM, grown 2 h, centrifuged at >16,000 *g*, resuspended in CDM, and inoculated onto CDM agar plates as colony biofilms [153]. Colony biofilms were grown 4 hours on CDM agar, transferred to fresh CDM agar for 4 more hours of growth, and collected in RNA later solution.

4.2.2 Determining generation times

Wild-type *A. actinomycetemcomitans* 624 and the Δ*fdhF1F2*, Δ*frdABCD*, and Δ*pfl* mutants were grown shaking at 150 rpm in CDM liquid medium aerobically at 37°C in 5% CO_2 atmosphere and statically in an anaerobic chamber (Coy). Because *A. actinomycetemcomitans* 624 grows in large aggregates *in vitro* that cannot be adequately dispersed, cellular protein concentrations were determined using a Bradford assay (BioRad) to serve as a proxy for cellular growth and calculate generation times. To prepare lysates for the Bradford assay, cells were removed throughout exponential growth phase, pelleted in a microcentrifuge, resuspended in 6 M urea, and boiled for 30 m at 100°C to lyse the cells. Generation times were calculated by plotting cellular protein concentrations over time.

4.2.3 *A. actinomycetemcomitans* murine abscess infections

Three-day murine abscess infections were established with wild-type and mutant *A. actinomycetemcomitans* strains, as described previously [22]. Severity of infections was determined by plate counts on BHI agar and BHI agar supplemented with 50 μg/ml spectinomycin, for wild-type and mutant *A. actinomycetemcomitans* strains, respectively. The experiments described here were conducted according to the Guide for Care and Use of Laboratory Animals of the National Institutes of Health. The Texas Tech University Health Sciences Center Institutional Animal Care and Use Committee approved the protocol (Protocol # 09039).

4.2.4 RNA isolation and high-throughput sequencing library preparation

A. actinomycetemcomitans biofilm cells stored in RNA later solution were centrifuged at 5,000 *g*, resuspended in 100 μl 10 mg/ml lysozyme in TE buffer and incubated for 10 m at 25°C. After lysozyme treatment, 200 μl PBS was added to each lysed cell solution and further lysis was carried out by mechanical disruption of each sample four times for 30 s at maximum speed in a Mini-Beadbeater (Biospec Products) and lysed cell solutions were stored in an ice bath for 2 m in between each bead beating. Total RNA was isolated from resulting lysed cells using 1 ml RNA Bee solution (Tel-Test) according to the manufacturer's protocol. Murine abscesses were resuspended immediately in 1 mL RNA-Bee, bead-beaten 4 times as described above, and RNA was purified from pooled abscesses (Table 4.1). DNA contamination in RNA samples was removed by treating 5 μg total RNA with 2.5 U RQ1 DNase (Promega) for 30 m at 37°C and RNA was purified with 500 μl RNA Bee solution. DNA removal was verified by PCR amplification of the *A. actinomycetemcomitans clpX* protease gene from DNase-

treated RNA (Table 4.2). Bacterial and host rRNAs were depleted using commercially available capture methods (Ambion MICROBExpress and MICROBEnrich kits) or enzymatic degradation (Epicentre Terminator 5'-monophosphate-dependent ribonuclease) according to manufacturers' protocols. To produce RNA between 20-500 nt that is optimal for sequencing, 1 μg rRNA-depleted RNA samples was treated with NEBNext RNA fragmentation buffer (NEB). Strand-specific cDNA libraries were prepared using commercially available T4 RNA ligase-based kits and sequenced on the SOLiD V4 and Illumina HiSeq2000 platforms. Both SOLiD and Illumina cDNA libraries were subjected to polyacrylamide gel extraction to purify cDNA between 130-500 nt, removing library adapters and primers from the samples. Sequencing library preparation kits and sequencing outputs are summarized in Table 4.1. Sequencing reads have been deposited in the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra), accession number SRP022893.

Table 4.1 RNA-seq sample and sequencing information.

Sample[1]	[2]rRNA deplete	[3]Prep	Platform	[4]Total reads	[5]map	[6]m/ncRNA
B1.1	Express	NEB (S)	SOLiD V4	6,323,061	886,203	85,547
B1.2	Express	NEB (S)	SOLiD V4	9,817,673	2,354,500	200,164
B2.1	Tex	NEB (S)	SOLiD V4	14,060,736	3,165,547	50,282
B2.2	Tex	Epicentre	HiSeq2000	12,632,073	132,635	16,235
B2.3	NA	Epicentre	HiSeq2000	8,208,689	557,369	120,289
A1(6)	Enrich	NEB (I)	HiSeq2000	25,297,104	5,798,615	434,191
A2(2)	Enrich	NEB (I)	HiSeq2000	42,804,092	25,539,796	839,805
A3(2)	Enrich	NEB (I)	HiSeq2000	52,165,722	31,983,011	915,676

[1]Bx.y: B = biofilm, x = biological replicate, y = technical replicate; A(#): A = abscess, # = number of pooled abscesses

[2]rRNA depletion method used: Express=Ambion MICROBExpress kit, Enrich=Ambion MICROBEnrich kit, Tex=Epicentre Terminator 5' monophosphate dependent exonuclease, and NA=no depletion method was utilized

[3]Prep: Library preparation kit used. NEB (S)=NEB Next Small RNA Sample Prep Set 3 for SOLiD, Epicentre=Epicentre ScriptMiner Small RNA Library Preparation kit, NEB (I)=NEB Next Multiplex Small RNA Library Prep Set for Illumina.

[4]Total reads: the number of reads sequenced

[5]map: the number of reads mapping to the *A. actinomycetemcomitans* genome

[6]m/ncRNA: the number of reads mapping to an mRNA or ncRNA

Table 4.2 Primer sequences.

Primer	Sequence (5'-3')
fdhKO-P1F-USS	AAGTGCGGTCCGGTCAGATTCGTACATACAAC
fdhKO-P1R-Sp	CATGTATTCACGAACGAAAATCGAGTAATGTCCTTATTTGAGCTACAAC
fdhKO-P2F-Sp	GAAAACAATAAACCCTTGCATATGATTTAACGAAAATTACAGCCGC
fdhKO-P2R-USS	AAGTGCGGTTGGCTATCTCTAATAAAGTTGGGAG
fdhKO-verify	CTTCAACGCCACTCAATATTTATCC
frdKO-P1F-USS	AAGTGCGGTGCCACCCGTAAATCTTCCG
frdKO-P1R-Sp	CATGTATTCACGAACGAAAATCGGATATTCCTCCAAGTAAAGGGATAG
frdKO-P2F-Sp	GAAAACAATAAACCCTTGCATATGCAACAAAAAACGCCCTCAATG
frdKO-P2R-USS	AAGTGCGGTGACCGCACTTTTGCTAACG
frdKO-verify	CGCTGGTGATTATCCTGTTC
pflKO-P1F-USS	AAGTGCGGTTGGTATCTCCGGCAGAAATG
pflKO-P1R-Sp	CATGTATTCACGAACGAAAATCGAACAAATTATCTACGACCGGGAAAG
pflKO-P2F-Sp	GAAAACAATAAACCCTTGCATATGCAGCCGTTTACCATTTAGGTAAG
pflKO-P2R-USS	AAGTGCGGTGCGACTATTGTAGTGCGCTTTG
pflKO-verify	GAATTCACCCGAACGGGTTATG
Spec-F	CGTTTTTCGTTCGTGAATACATG
Spec-R	CATATGCAAGGGTTTATTGTTTTC
clpX-F	GCAGAAACCATGGCGCGTATG
clpX-R	CAATTCGCTTAATGGCGCCAC

4.2.5 Computational methods

The *A. actinomycetemcomitans* D7S-1 reference genome (Genbank accession ADCF01000001.1) was used for RNA-seq read alignment [125]. The 50 bp single-end SOLiD sequencing reads were aligned in colorspace to the reference genome using SHRiMP version 2.2.1 to produce sam format read alignment files [169]. The longer ~100 bp Illumina sequencing reads were pre-filtered using the Flexbar version 2.0 [170] to remove library adapter sequences from the sequencing reads (Illumina index sequencing primer sequence: 5'-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3', Illumina 3' adapter sequence: 5'-TCGTATGCCGTCTTCTGCTTG-3') and improve alignment to the reference genome. Filtered Illumina reads were aligned to the reference using Bowtie version 2.0.0 to produce read alignment files in sam format [171]. The read alignment files were converted from sam to bam format, sorted and indexed with Samtools [172]. Individual read alignments in bam file format from each condition were visualized with the Integrative Genomics Viewer (IGV) [173, 174]. Using the read alignments visualized with IGV, TSS were manually recorded in GFF format by identifying the position where reads aligned upstream of genes. Using an analogous approach, ncRNAs visualized in IGV were manually annotated by the identification of reads mapping to intergenic regions and antisense to coding sequences. Whole genome annotations for TSS and ncRNAs are provided in GFF and Genbank formats with instructions for opening the files on our laboratory's website at http://web.biosci.utexas.edu/whiteley_lab/pages/resources.html (DataSetS1.gff and DataSetS2.gbk). Prior to determining differential gene expression, read alignment files produced by SHRiMP were reformatted for HTSeq read counting with a custom Perl

script. The number of reads aligned to each gene and ncRNA in the newly annotated reference genome were calculated using HTSeq (http://www.huber.embl.de/users/anders/HTSeq), not including tRNAs and rRNAs. Prior to calculating differential gene expression, read counts per gene were summed for technical replicates, per DESeq recommendation. Differential gene and ncRNA expression between *A. actinomycetemcomitans* biofilms and abscesses were determined based on a negative binomial distribution using the R package DESeq version 1.6.1 [175]. Clusters of orthologous groups (COG) gene enrichment analysis was conducted using COGs determined previously [44]. Enrichment of differentially regulated genes in a given COG category was determined by comparing the prevalence of up- or down-regulated genes assigned to a specific COG category and comparing to the prevalence of genes in the entire genome assigned to that COG category. Enrichment of a COG category in either the up- or down-regulated gene set relative to the genome was calculated using a Fisher's exact test custom macro in Microsoft Excel. Whole genome diagrams were produced with Circos version 0.55 [176].

4.2.6 Construction of *A. actinomycetemcomitans* deletion mutants

A. actinomycetemcomitans mutants were constructed for the FDH-H operon (*fdhF1F2;* D7S_2219-D7S_2220), fumarate reductase operon (*frdABCD*; D7S_1533-D7S_1536), and pyruvate formate lyase (*pfl*; D7S_2028) by double homologous recombination. Wild-type genes of interest were substituted with *aad9* (encoding spectinomycin resistance) from pVT1461 by overlap extension PCR and natural transformation of *A. actinomycetemcomitans* 624 [127, 177]. The *aad9* spectinomycin resistance gene was amplified from pVT1461 using Spec-F and Spec-R to generate Spec^r.

The PCR overlap extension constructs were amplified for the *fdhF1F2* operon using primers fdhKO-P1F-USS and fdhKO-P1R-Sp to generate the fdhKO-P1 PCR product and primers fdhKO-P2F-Sp and fdhKO-P2R-USS to generate the fdhKO-P2 PCR product. Overlap extension PCR mixing 500 ng each of fdhKO-P1, Spec^r, and fdhKO-P2 with primers fdhKO-P1F-USS and fdhKO-P2R-USS, making the fdhKO PCR product containing the ~1 kb region upstream of the *fdhF1F2* operon, *aad9*, and the ~1 kb region downstream of the *fdhF1F2* operon and *A. actinomycetemcomitans* specific DNA uptake signal sequences on each end to enhance natural transformation. Prior to natural transformation, the ~3kb fdhKO overlap extension PCR product was gel extracted using a Fermentas gel extraction kit. The *fdhF1F2* operon was substituted by natural transformation on tryptic soy agar supplemented with 0.5% yeast extract (TSAYE) with 5% heat inactivated horse serum 1 mM cAMP in an anaerobic chamber via double homologous recombination with 1 μg of the fdhKO overlap extension PCR product. The resulting Δ*fdhF1F2* mutants were selected on TSAYE with 50 μg/ml spectinomycin. The Δ*fdhF1F2* mutant was verified by growing the mutant in TSBYE with 50 μg/ml spectinomycin, purifying genomic DNA using a Qiagen DNeasy kit, and PCR amplifying the *aad9* gene with the SpecR primer specific to the *aad9* gene and the fdhKO-verify primer, which anneals upstream of the fdhKO-P1F-USS primer. The verification product was observed for Δ*fdhF1F2* mutant genomic DNA but not wild-type genomic DNA. Analogous methods were used to generate the Δ*frdABCD* mutant using primers frdKO-P1F-USS, frdKO-P1R-Sp, frdKO-P2F-Sp, frdKO-P2R-USS, and frdKO-verify, as well as the Δ*pfl* mutant using primers pflKO-P1F-USS, pflKO-P1R-Sp, pflKO-P2F-Sp, pflKO-P2R-USS, and pflKO-verify. Primer sequences are available in Table 4.2.

4.3 RESULTS

4.3.1 *In vitro* and *in vivo* RNA-seq

The goal of this study was to utilize RNA-seq to provide a high-resolution analysis of the *A. actinomycetemcomitans* genetic elements expressed during *in vivo* growth, with a specific interest in metabolic genes. While *A. actinomycetemcomitans* is most noted for its ability to cause oral infections, our laboratory has utilized an extra-oral abscess model to study *A. actinomycetemcomitans* virulence [22]. We prefer this model for the following reasons: *A. actinomycetemcomitans* causes extra-oral infections, including abscess infections, thus the model has clinical relevance [63, 64, 168]; the infectious dose is easily controlled and results in a contained infection that has been used to study pathogenesis of oral bacteria [22-24]; and unlike periodontal models of infection, infected tissue can be easily removed to assess disease severity [22] or for RNA isolation. In addition to performing transcriptome analysis in a relevant animal model of infection, it is also critical to examine *A. actinomycetemcomitans* gene expression during *in vitro* growth under defined conditions. Growth under known nutritional conditions is essential to provide a well-defined control transcriptome for comparing to the undefined *in vivo* transcriptome. Therefore, we grew *A. actinomycetemcomitans* as a colony biofilm on a solid defined medium with glucose as the sole energy source. Since glucose metabolism is well characterized and *A. actinomycetemcomitans* forms robust biofilms, this growth condition provided a well-defined, relevant control for the *in vivo* experiments.

Total RNA was harvested from biofilms and abscesses and strand-specific cDNA libraries were subjected to RNA-seq analysis. Because murine abscesses contain small

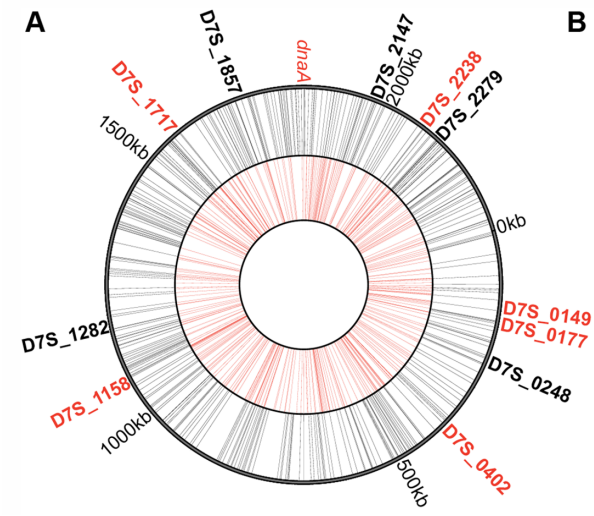
populations of bacteria (approximately 10^6 bacteria per abscess) it was essential to develop methodology to efficiently purify RNA from the infections, deplete the abundant murine ribosomal RNA, and enrich for bacterial transcripts in order to obtain sufficient bacterial sequencing reads for analysis. Since sequencing technologies are constantly evolving, we used two sequencing platforms to quantify transcript levels. However, it is important to note that all sequencing libraries were prepared using similar RNA ligation-based protocols and each technology produces reads of sufficient length to map specifically to reference genomes. For both technologies, gene expression data was normalized among all replicates for differential expression analyses. The resulting sequencing reads were processed and aligned to the *A. actinomycetemcomitans* genome to determine the *A. actinomycetemcomitans* transcriptome during *in vitro* and *in vivo* growth. In total, 7.1 million biofilm and 63.3 million *in vivo* sequencing reads were aligned to the *A. actinomycetemcomitans* genome (Table 4.1) resulting in an average of ~190 and 870 reads/gene for *in vitro* and *in vivo* grown *A. actinomycetemcomitans*, respectively. Using these data we manually annotated *A. actinomycetemcomitans* TSS and ncRNAs. Differential expression of protein-coding genes and ncRNAs was determined after normalizing for read number to identify *in vivo* regulated genes.

4.3.2 Transcription start site mapping

To begin characterizing the primary transcriptome of *A. actinomycetemcomitans*, TSS were manually annotated using the aligned RNA-seq data (as detailed in the 'computational methods' section of the Materials and Methods). TSS were identified by manually recording the position where reads aligned upstream of annotated genes using IGV [173, 174]. In total, 691 mRNA TSS were identified in *A. actinomycetemcomitans*

(Fig. 4.1). Among the TSS, several previously characterized TSS including *katA, apiA,* and *lysT* were verified by our RNA-seq analyses, thus validating our approach [16, 153]. Due to the methods used to generate the RNA-seq libraries, it is important to note that the TSS identified represent authentic TSS as well as the 5' ends of processed RNAs. However, most bacterial transcripts are not specifically processed at a defined nucleotide but are processively degraded [178]. Moreover, since several of the TSS identified here correspond with TSS identified by other approaches [16, 153], many of these TSS likely represent true start sites.

Secondary TSS have been observed among genes in *Helicobacter pylori* and in a cyanobacterium, but the roles of alternate TSS are relatively obscure [179, 180]. When TSS for genes differ from one growth condition to another, it suggests that the gene is transcribed from multiple promoters or the mRNA is differentially processed. We observed 11 genes with different TSS *in vivo* and *in vitro* (Fig. 4.1). One potential hypothesis is that alternative promoters are used to differentially regulate genes during *in vivo* and *in vitro* growth; however, 9 of the 11 genes were not differentially expressed suggesting that the utilization of alternate promoters does not correlate with differential regulation.



B

Locus	Diff.	Product
D7S_0149	-9 bp	H-NS histone family
D7S_0177	-31 bp	ribosomal protein S1
D7S_0248	+31 bp	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase
D7S_0402	-42 bp	OmpA-like outer membrane protein
D7S_1158	-30 bp	nucleotidyltransferase
D7S_1282 -4.6-fold	+22 bp	putative FHA domain protein
D7S_1717	+23 bp	hypothetical protein
D7S_1857	+16 bp	RpsJ protein
D7S_2147	-39 bp	glycoside hydrolase, family 19, putative
D7S_2238 -9.4-fold	+5 bp	formate dehydrogenase N, alpha subunit
D7S_2279	+16 bp	transcription elongation factor GreA

Figure 4.1 ***A. actinomycetemcomitans*** **mRNA 5' ends and 5' end switching** ***in vivo.***

(A) *A. actinomycetemcomitans* 5' ends mapped along the genome to the positive strand (outer circle, black) and negative strand (inner circle, red). The putative origin of replication is marked by *dnaA*. Locus tags for genes with different start sites *in vivo* and *in vitro* are marked along the outermost circle and colored according to read strand. **(B)** Genes with different start sites *in vivo*. Locus tags for genes are indicated, significant differences in mRNA expression (fold-change) during *in vivo* growth are listed below the locus tags, the location of the *in vivo* TSS relative to the *in vitro* TSS, and the gene function follow. When no fold-change is indicated, genes were not differentially expressed *in vivo*.

4.3.3 *A. actinomycetemcomitans* ncRNA discovery

RNA-seq provides tremendous insight into the identification and potential functions of ncRNAs transcribed from intergenic regions and antisense to coding genes. Like TSS and operons, ncRNAs were identified based on contiguous reads aligning to intergenic regions and antisense to protein-coding genes observed in IGV (Fig. 4.2). The ncRNA sequences were collected and compared to the Rfam database to determine homology to ncRNAs characterized in other bacteria [181, 182]. In biofilms and *in vivo*, *A. actinomycetemcomitans* expressed a number of ncRNAs, including riboswitches, cis-anstisense RNAs (asRNAs), small ncRNAs (sRNAs), and clustered regularly interspaced short palindromic repeat RNAs (crRNAs) (Fig. 4.2). In total, *A. actinomycetemcomitans* expressed 210 ncRNAs, including 127 asRNAs, 3 pre-crRNAs, 3 riboswitches, 75 sRNAs, and 1 unidentified tRNA. Among these ncRNAs were housekeeping ncRNAs involved in transcription, translation, and protein secretion (Table 4.3), and several previously discovered and predicted ncRNAs [183-189]. In addition, many ncRNAs were differentially expressed *in vivo* compared to *in vitro* biofilm growth. Indeed, 80 out of 210 ncRNAs were differentially expressed *in vivo*, including 39 up-regulated and 41 down-regulated ncRNAs (Fig. 4.2). This result is significant, but perhaps expected, since the roles of ncRNAs for *in vivo* growth and persistence has been described for other bacteria [190-192].

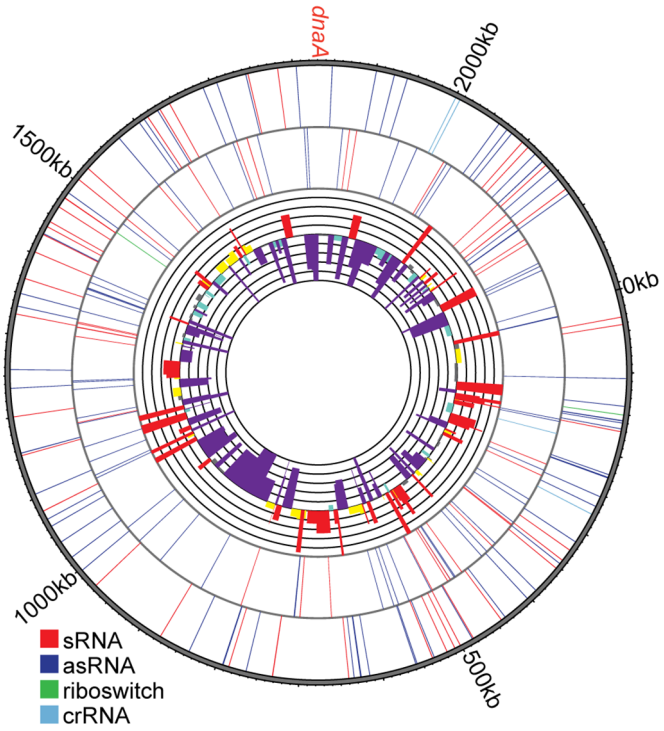


Figure 4.2 ***A. actinomycetemcomitans*** **ncRNAs are differentially regulated** ***in vivo.***

A. actinomycetemcomitans ncRNAs mapped along the genome to the positive strand (outer circle) and negative strand (inner circle). ncRNAs are colored according to type: sRNA (red), asRNA (blue), riboswitch (green), and pre-crRNA (light blue). The circular histogram indicates the $\log_2$ fold-change of ncRNA expression *in vivo* (-10 to 10, inner to outer line). Bars are colored relative to change in ncRNA expression, > 2-fold up-regulated (yellow), > 5-fold up-regulated (red), > 2-fold down-regulated (light blue), > 5-fold down-regulated (purple), and no change (grey).

Table 4.3 Rfam predictions for *A. actinomycetemcomitans* ncRNAs.

					Rfam predictions		
Locus tag	**Start**	**Stop**	**Strand**	**RNA family**	**Start**	**End**	**E-value**
D7S_0043.1	41437	41836	+	sRNA, RNaseP A	2	390	4.91E-40
D7S_0235.1	240380	240648	-	sRNA, GcvB	18	174	8.90E-23
D7S_0343.1	340697	340887	+	sRNA, tfoR	91	181	5.10E-03
D7S_0661.1	640001	640182	+	sRNA, his leader	42	171	4.28E-17
D7S_0743.1	704845	704936	-	tRNA	2	74	5.29E-14
D7S_0849.1	794824	795016	-	sRNA, 6S RNA	1	184	1.52E-23
D7S_1442.2	1323184	1323273	-	sRNA, C4	1	74	4.90E-12
D7S_1454.1	1330343	1330779	-	sRNA, tmRNA	1	366	7.92E-92
D7S_1607.1	1477807	1477998	-	riboswtich, FMN riboswitch	1	182	5.17E-29
D7S_1716.1	1572027	1572259	-	riboswitch, glycine riboswitch	87	231	1.00E-10
D7S_2270.1	2096164	2096331	+	sRNA bacterial SRP	47	145	2.08E-18

Nucleotide sequences were collected for ncRNAs and searched against Rfam database families (RNA family). Locus tags are in Table S3. Start and stop refer to the location of the ncRNA on the *A. actinomycetemcomitans* reference genome. The portions of the *A. actinomycetemcomitans* ncRNA sequences that possessed sequence homology to RNAs in the Rfam database are indicated (Start and End), as well as the Expectation value (E-values).

4.3.4 Fermentative metabolism and anaerobic respiration promote *in vivo* survival

Our main goal was to use RNA-seq to identify metabolic pathways that impact *A. actinomycetemcomitans* fitness *in vivo*. The hypothesis was that genes differentially regulated *in vivo* will be enriched for those important for growth in the murine abscess. To identify differentially regulated genes, *in vitro* and *in vivo* transcriptomes were compared. This analysis yielded 337 genes (~14% of all predicted genes) that were differentially regulated 2-fold or greater (Fig. 4.3A) in the abscess infection relative to the *in vitro* biofilm. Of these genes, 107 were differentially regulated 5-fold or greater. Notably, genes encoding the *A. actinomycetemcomitans* virulence factors leukotoxin (locus tag D7S_0615) and cytolethal distending toxin (locus tag D7S_2348) [37-40, 193] were up-regulated *in vivo*. While leukotoxin is upregulated in low oxygen conditions [193], prior to this study, neither toxin was known to be up-regulated during infection.

To gain a broader perspective of genes differentially regulated *in vivo*, we performed an enrichment analysis to define clusters of orthologous groups (COGs) [194, 195] for differentially regulated genes. This analysis first clusters genes based on their putative function then examines whether a particular category is enriched for differentially regulated genes compared to what would be expected by chance. Using the COGs defined by Kittichotirat et al. [44], only one COG category (COG C) was significantly enriched for both up- and down-regulated genes. COG C includes genes involved in energy production and conversion (Fig. 4.3B), suggesting that *A actinomycetemcomitans* undergoes substantial changes in metabolism during *in vivo* growth.

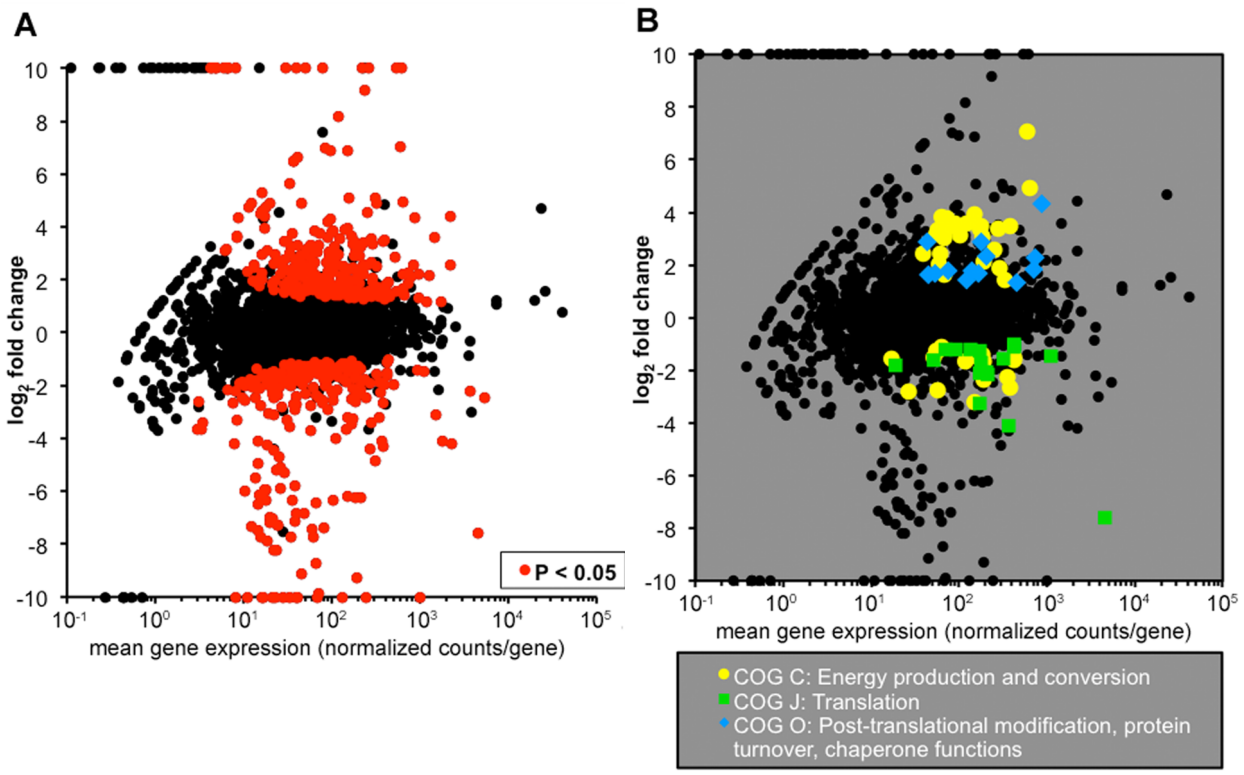


Figure 4.3 Differential RNA-seq and COG enrichment analyses reveal the importance of energy metabolism *in vivo*.

(A) *In vivo* differential gene expression plot. Fold-change for each gene *in vivo* is plotted against mean gene expression. Red points represent the 337 genes with significant differential gene expression *in vivo* ($P < 0.05$, DESeq). **(B)** *In vivo* COG enrichment analysis. Enrichment of COGs among differentially regulated genes compared to the abundance of the COGs in the genome was determined using Fisher's exact test ($P < 0.05$).

Included within the differentially regulated metabolic genes were operons encoding formate dehydrogenase H (*fdhF1F2*) and fumarate reductase (*frdABCD*), which are involved in fermentative metabolism and anaerobic respiration, respectively (Fig. 4.4). Thus we hypothesized that these operons would be important for *A. actinomycetemcomitans* growth in the murine abscess. To test this hypothesis, individual mutants containing deletions of *fdhF1F2* or *frdABCD* were constructed via allelic exchange. In addition, a strain containing a deletion of the gene encoding pyruvate formate lyase (*pfl*) was also constructed. Since *pfl* was not differentially regulated *in vivo*, we hypothesized that it would not be critical for *A. actinomycetemcomitans* growth in the abscess; thus this mutant served as a control for subsequent *in vivo* experiments.

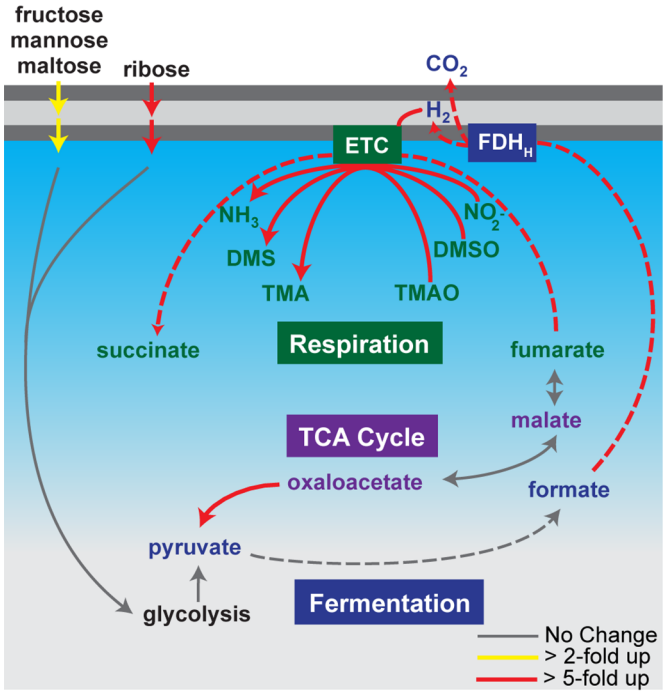


Figure 4.4 Metabolic genes and pathways up-regulated *in vivo*.

Genes putatively encoding fructose, glucose, mannose, and ribose transporters were up-regulated in the murine abscess along with multiple fermentative and anaerobic respiratory pathways. (red: > 5-fold up-regulated, yellow: > 2-fold up-regulated, grey: no change). Dashed lines indicate genes/pathways selected for mutagenesis. TMAO (trimethylamine-*N*-oxide), TMA (trimethylamine), DMSO (dimethylsulfoxide), DMS (dimethyl sulfide), FDH (formate dehydrogenase), ETC (electron transport chain).

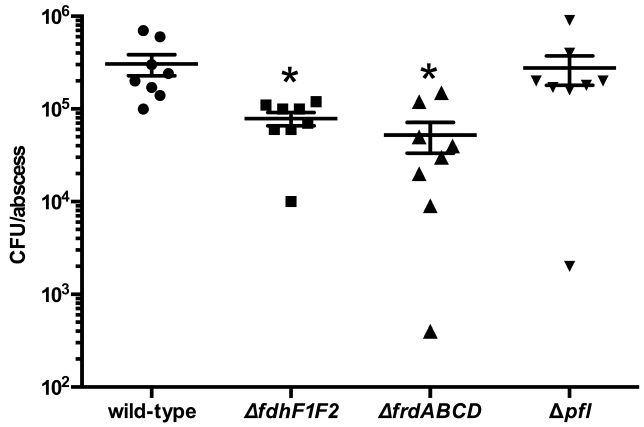


Figure 4.5 Metabolic mutants are attenuated *in vivo*.

Monoculture abscess infections with wild-type, Δ*fdhF1F2*, Δ*frdABCD*, and Δ*pfl* strains were processed 3 days post-infection and CFU/abscess were determined by plate counts (*$P < 0.05$, Mann-Whitney U Test, compared to wild-type). Each symbol represents infection of an individual mouse.

Abscesses produced by both the Δ*fdhF1F2* and Δ*frdABCD* mutants contained ~10-fold less bacteria than wild-type infections (Fig 4.5), indicating that these pathways are critical for *A. actinomycetemcomitans* fitness in the murine abscess. In contrast, the Δ*pfl* mutant showed similar bacterial numbers to wild-type *A. actinomycetemcomitans* (Fig. 4.5). Importantly, deletion of *fdhF1F2* and *frdABCD* operons did not impact aerobic or anaerobic *in vitro* growth in a glucose defined medium (data not shown). Collectively, these data suggest that the *fdhF1F2* and the *frdABCD* operons are not only highly up-regulated during *in vivo* growth but also impact *A. actinomycetemcomitans* fitness in the murine abscess.

4.4 DISCUSSION

In this study, a high-resolution transcriptome analysis was performed for *in vitro* and *in vivo* grown *A. actinomycetemcomitans* using RNA-seq. Protocols were developed to isolate total RNA from abscess infections and enrich for bacterial RNA before RNA-seq analysis. From these data, the terminal 5' sequences of ~700 RNAs were mapped, most of which likely represent authentic transcriptional start sites. In addition, over 300 *A. actinomycetemcomitans* genes were shown to be differentially regulated during abscess infection as compared to *in vitro* grown bacteria. This study provides the first high-resolution transcriptome analysis of a bacterium during growth in an abscess and establishes methodology for performing RNA-seq on samples containing predominantly host RNA.

The primary goal of this work was to identify metabolic pathways utilized by *A. actinomycetemcomitans in vivo* by focusing on metabolic genes differentially

regulated during growth in the abscess. The use of defined *in vitro* growth conditions provided a robust control to probe *A. actinomycetemcomitans* metabolism during *in vivo* growth, allowing identification of both fermentative and respiratory pathways important for *A. actinomycetemcomitans* growth in the murine abscess (Fig. 4.4). The enrichment of pathways, including *fdhF1F2* and *frdABCD*, primarily involved in regeneration of NAD^+ during anaerobic growth, suggests that *A. actinomycetemcomitans* experiences low oxygen in the abscess. This observation is curious in regard to previous studies from our group demonstrating that during co-culture infection with the peroxigenic oral bacterium *Streptococcus gordonii*, *A. actinomycetemcomitans* grows aerobically in the abscess [22]. Why the *A. actinomycetemcomitans-S. gordonii* co-culture abscess infection is aerobic is currently unknown, but these results emphasize that a pathogen may require distinct metabolic pathways to persist in an infection site dependent on whether it is alone or in the presence of other microbes. Of course the infection site may also impact gene expression, and it will be interesting to examine *A. actinomycetemcomitans* gene expression in the oral cavity during mono- and co-culture.

One of the primary advantages of RNA-seq is that it provides unparalleled insight into ncRNAs expressed in a given growth condition. Here we reported the differential regulation of 80 *A. actinomycetemcomitans* ncRNAs and increased expression of *hfq in vivo*. The up-regulation of *hfq* is particularly striking since it encodes an sRNA chaperone that promotes ncRNA interactions with mRNA targets [196, 197]. Similar to other studies, this suggests that *A. actinomycetemcomitans* ncRNAs may play important regulatory roles *in vivo* [190, 191, 198]. Differential regulation of ncRNAs also provided insight into the nutritional content of the infection site. For example, the

A. actinomycetemcomitans lysine riboswitch and the sRNA GcvB [114, 117, 153] showed reduced abundance during infection. These ncRNAs regulate expression of lysine and glycine transport proteins respectively, and their decreased levels suggest that these amino acids are found at reduced levels in the abscess.

Transcriptome analysis using RNA-seq is a powerful and sensitive tool for studying infectious disease processes *in vivo*. Development of methodologies for performing RNA-seq on *in vivo* samples dominated by host RNA allowed tremendous insight into the physiology and metabolism of *A. actinomycetemcomitans* during *in vivo* growth. As evidenced by the abscess CFU counts, the methodology employed here can be used to perform gene expression analyses on as few as 10^5 *in vivo* bacteria. Ultimately, applying this technology to defined polymicrobial infections, as well as undefined complex human infections, will provide a window into the physiology of bacterial pathogens in human infections.

Chapter 5. Gene Expression of the Human Microbiome during Health and Disease

5.1 Introduction

Periodontal disease is a "microbial shift" disease associated with massive reorganization of the microbiota residing in the subgingival crevice, the region between the tooth surface and the gingival epithelium [6]. Studies using culture-independent methods have shown that the presence and prevalence of certain species or genes in the microbiome are correlated with disease [6, 199-202]. In light of these results, microbiome-associated diseases such as periodontitis are increasingly examined through an ecological lens. Yet as communities occupying various niches throughout the human body shift into a diseased state, it is unclear how the behavior of the microbiota changes during disease. Thus the question arises, to what extent are changes in the ecosystem attributable to alterations in the abundance of certain community members or changes in the behavior of existing organisms? Recent advances in high-throughput sequencing have greatly accelerated the cataloging of changes to the organismal and metagenomic content of microbial communities in health and disease [203, 204], and they also provide an avenue to analyze the behavior of the community and its constituent members. In this study, we used massively parallel RNA sequencing to profile changes in both the composition and gene expression of the human microbiome in periodontitis.

5.2 RESULTS AND DISCUSSION

5.2.1 Quantifying microbiota with qrRNA-seq

Patient-matched healthy and diseased periodontal samples were collected from 10 patients with aggressive periodontitis (Table 5.1). Quantification of rDNA has been used to identify different bacterial species correlated with periodontal disease or health [6]; however, these studies also likely captured extracellular DNA, an important structural component of surface attached biofilms [205, 206]. Moreover, these studies do not differentiate between metabolically active and dead cells. Since rRNA content is proportional to growth rate [207], direct rRNA quantification provides a measurement of metabolic activity of individual members of the community. To this end, we developed quantitative rRNA sequencing (qrRNA-seq). Our qrRNA-seq method sequences two adjacent variable regions of the 16S rRNA of bacterial ribosomes, aligns these to the Greengenes rRNA database [208], and quantitates ribosomal RNA using Qiime [209](Table 5.2). Similar to studies that used rDNA-based procedures [5, 6], we found that many bacteria are present in both healthy and diseased communities; however, our data indicated that many of the most metabolically active microbes in disease samples were those previously associated with infection, including *Tannerella sp., Prevotella sp., Treponema sp.,* and *Porphyromonas sp*. (Fig. 5.1).

Table 5.1 Aggressive periodontitis patient data.

								Mean PD[1] per sampling site	
Sample	**Age**	**Gender**	**PD(full)[1]**	**CAL(full)[2]**	**Plaque**	**Bleeding (%)**	**Smoking**	**Disease**	**Healthy**
1	34	F	6.5	6.5	1	60	no	5	3
2	36	M	6	7	2	80	yes	5	3
3	30	F	5	5	2	70	yes	6	2
4	34	M	5	6.5	1	70	no	5	2
5	36	M	5	6	2	90	yes	5	2
6	40	F	5	5	1	60	no	5	3
7	34	M	5.5	5.5	1	70	no	7	2
8	33	F	6	6	2	70	yes	6	2
9	36	F	5.5	5.5	2	80	no	5	2
10	37	F	5	5	1	60	no	5	3

[1]PD: Probing depth of the subgingival crevice (mm)
[2]CAL: Clinical attachment loss of gingival epithelium (mm)

Table 5.2 qrRNA-seq sequencing and analysis information.

Sample	Total reads[1]	Assembled reads[2]	Assigned OTUs[3]
D1	168,406	31,314	28,421
D2	171,933	40,849	36,916
D3	165,777	26,637	23,416
D4	179,507	35,624	32,597
D5	167,569	29,838	27,126
D6	157,970	46,462	42,953
D7	154,688	35,032	32,334
D8	146,168	38,447	35,263
D9	151,182	20,432	18,009
D10	180,610	42,818	39,229
H1	148,195	17,022	16,629
H2	106,273	17,308	15,087
H3	175,362	26,423	23,775
H4	128,998	5,520	5,008
H5	140,917	8,690	7,836
H6	144,023	7,928	7,314
H7	167,730	19,677	18,032
H8	180,986	7,308	6,565
H9	189,386	17,090	15,302
H10	167,501	12,443	11,300

[1]Total paired MiSeq sequencing reads.

[2]Number of reads assembled into contiguous 16S rRNA V4/V5 regions.

[3]Number of assembled reads that were assigned to operational taxonomic units (OTUs).

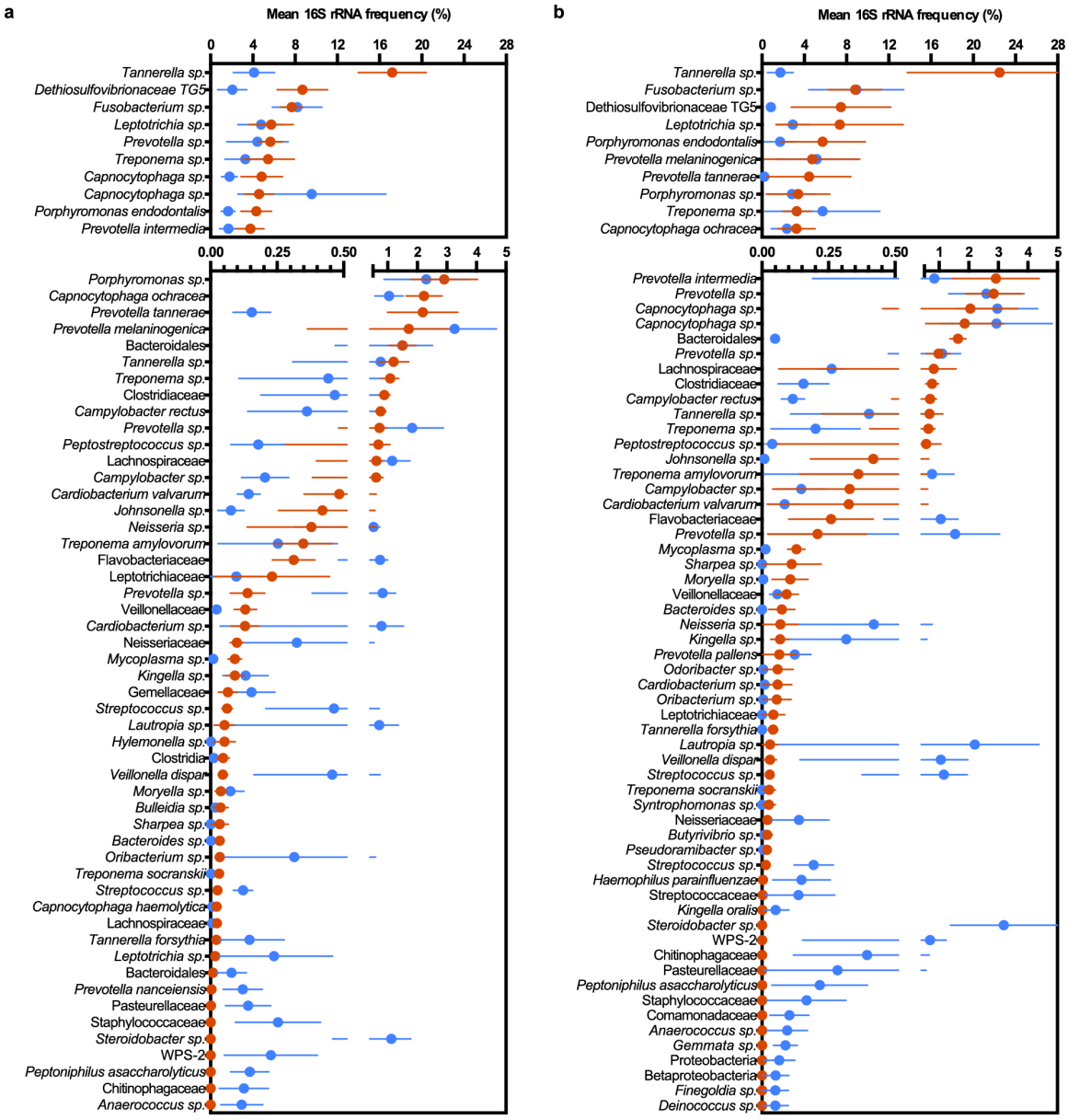


Figure 5.1 Average bacterial species abundances in healthy and diseased periodontal samples determined by qrRNA-seq.

(a) qrRNA-seq was used to determined OTUs present in health and disease. Shown are average abundances for healthy (blue) and diseased (red) samples. Error bars indicate s.e.m. (n = 10). **(b)** OTU abundances for healthy and diseased samples from patients 5, 6, and 8, which were selected for metaRNA-seq.

5.2.2 Diseased communities are less diverse than healthy communities

Since qrRNA-seq provides a quantitative assessment of ribosomes from individual species within the community, it allows assessment of the overall metabolic activity of each member of the population. Examination of ribosome content in healthy and diseased populations showed that diseased samples had significantly lower alpha diversity, or within-population diversity, than healthy populations (Fig. 5.2a). In addition to having fewer metabolically active species, the most prevalent members of diseased populations contributed more highly to overall rRNA content than prevalent species in healthy populations (Fig. 5.2b). These data together with a previous study [6] support a model wherein aggressive periodontitis is marked by the bloom of pre-existing pathogenic species that out-compete rare microbiota. Beta diversity analysis comparing the relatedness of diseased and healthy populations from multiple individuals showed that diseased populations clustered together tightly around the average disease state (Fig. 5.3). Moreover, diseased populations were more related to this average state than to paired healthy populations from the same individual. This did not hold true for healthy populations, which appeared to have a more random rRNA composition, as demonstrated by their scattered distribution in the multidimensional scaling plot (Fig. 5.3). These data show that healthy periodontal populations are diverse, patient-specific, and contain a large population of rare active microbiota, while a few commonly found, highly prevalent active organisms overwhelm rare healthy microbiota during aggressive periodontitis.

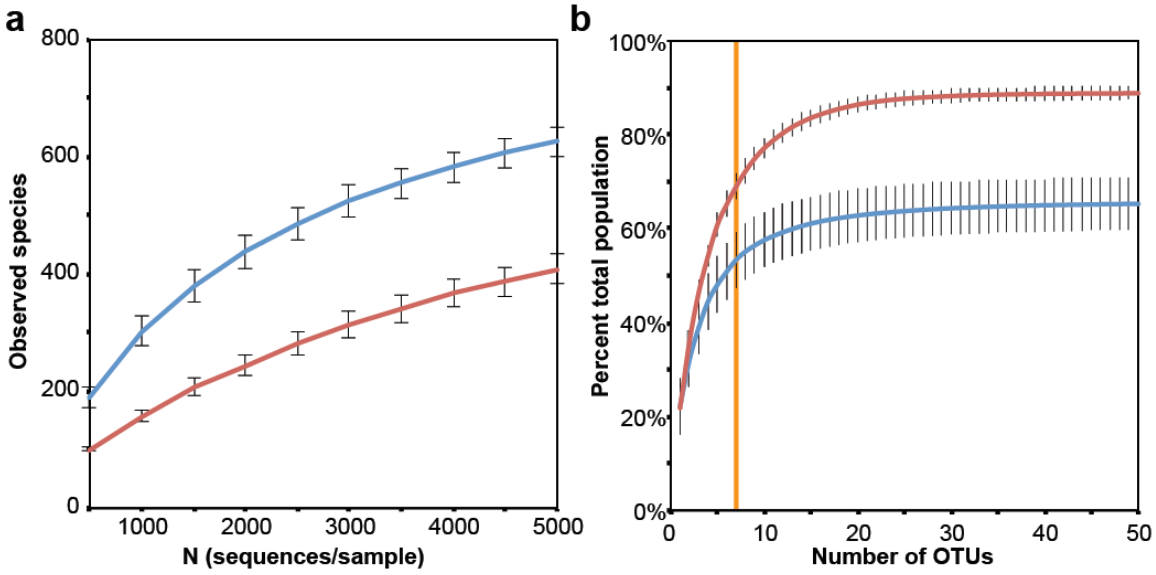
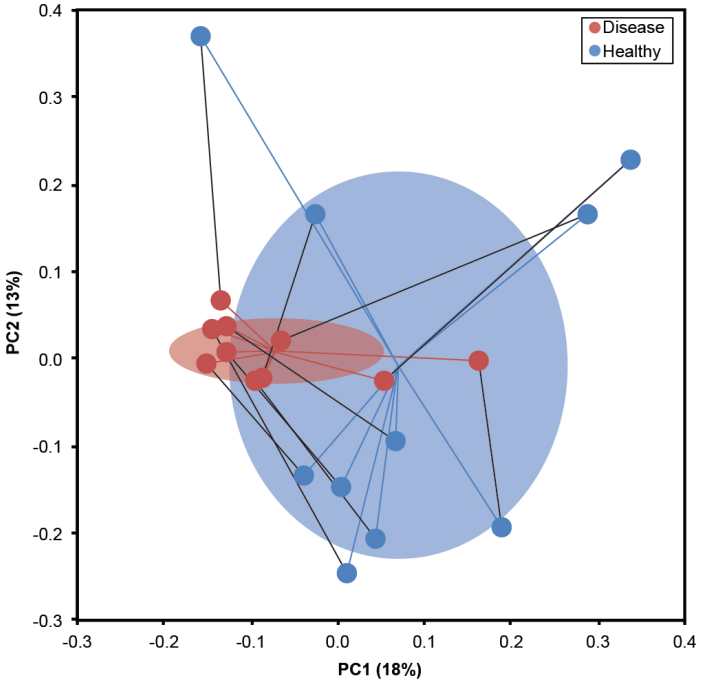


Figure 5.2 Ribosome quantification reveals that diseased periodontal microbiota are less diverse and contain fewer low abundance species than healthy populations.

(a) Number of distinct 16S rRNA sequences (operational taxonomic units, OTUs) observed in healthy (blue) and diseased (red) samples with increasing numbers of sequences sampled from each population. Error bars indicate S.E.M. (n = 10). **(b)** Number of OTUs contributing to the total population in healthy (blue) and diseased (red) samples. OTUs for each population were sorted from highest abundance to lowest and the running percentage was calculated for each additional OTU. OTUs at the Domain level were ignored. Grey bars indicate S.E.M. (n = 10). Orange line indicates point at which diseased OTUs contribute to a significantly greater percentage of the population than the healthy OTUs ($P < 0.05$, unpaired two-tailed Student's t-test).



Avg. distance to:	Centroid	Pair	*P*
Healthy	0.272 ± 0.112	0.261 ± 0.079	0.79634
Disease	0.081 ± 0.062	0.261 ± 0.079	0.00003*

Figure 5.3 Diseased oral microbial populations have similar population composition.

Beta diversity was measured using the unweighted Unifrac method to calculate relatedness of paired healthy (blue) and diseased (red) microbial populations by assessment of shared and unique species in each community. Principle coordinates 1 and 2 are plotted. Mean diseased and healthy centroids (mean + S.D.) are indicated by ellipses. Distances between samples and corresponding centroids are shown as blue and red lines, respectively. Black lines show distances between paired populations from the same patient. The table summarizes the mean distance (mean ± S.D.) from each sample to corresponding centroids and corresponding paired sample from the same patient (*P*, unpaired two-tailed Student's t-test).

5.2.3 Quantifying gene expression with metaRNA-seq

Changes in the composition of the microbiome have previously been associated with numerous diseases, including periodontitis, and the results of our qrRNA-seq analysis show that specific members of the community metabolically dominate during periodontal disease. Yet it remains unclear whether the behavior of the microbiota also changes during disease. To address this question, we used high-resolution community transcriptional profiling, which we have termed metaRNA-seq. Total RNA was extracted from patient-matched healthy and diseased samples, depleted of highly abundant human and bacterial rRNA, and sequenced on an Illumina HiSeq. In total, 1.5 billion metaRNA-seq reads were sequenced (Table 5.3). The majority (> 99%) of reads were of bacterial origin, as determined by alignment to the oral microbiome [210], human cDNA, and the viral genome databases (Table 5.4). To quantify gene expression, reads were aligned to a 60-organism "metagenome", representing microbes comprising 60-90% of total healthy or diseased rRNA. Since we characterized both healthy and diseased communities from the same patient with metaRNA-seq, we could analyze differential expression of >160,000 bacterial genes simultaneously between health and disease. For each sample, 28-85 million metaRNA-seq reads mapped to the 60-organism metagenome, representing an average coverage of ~170 reads per gene.

Table 5.3 metaRNA-seq sequencing information.

Sample	Total reads[1]	Trimmed reads[2]	Mapped[3]
D5	140,019,932	113,647,258	55,160,160
D6	190,430,010	160,583,365	70,164,137
D8	246,462,918	179,248,020	71,603,870
H5	261,039,989	90,831,243	34,054,085
H6	265,255,404	72,454,286	28,207,802
H8	199,443,355	156,541,929	85,709,824

[1]Total reads sequenced from each sample.
[2]Reads remaining after adapter removal.
[3]Total reads aligned to the 62-organism metagenome.

Table 5.4 Percent specific metaRNA-seq reads aligned to the HOMD database, human mRNA database, and viral genome database.

	Sample					
Database	**D5**	**D6**	**D8**	**H5**	**H6**	**H8**
HOMD	99.9999%	99.9999%	100.0000%	99.9632%	100.0000%	100.0000%
Human	0.0001%	0.0000%	0.0000%	0.0002%	0.0000%	0.0000%
Virus	0.0000%	0.0001%	0.0000%	0.0366%	0.0000%	0.0000%

5.2.4 Communal shift toward amino acid consumption

Previous studies have used genomic information to predict disease-associated shifts in metabolism; however these studies are based on the genetic capacity of the population rather than expression of metabolic genes [211, 212]. Our metaRNA-seq approach allows, for the first time, modeling of the metabolism of the microbiome during health and disease based solely on differential gene expression. In this approach, we used Enzyme Commission (EC) numbers to assign biochemical function to the >160,000 genes present in the 60-organism metagenome. EC numbers classify enzymes based on the reaction that they catalyze (i.e. enzymes catalyzing the same reaction will have the same EC number). This allowed us to calculate changes in expression of metabolic enzymes for the entire community during health and disease, resulting in a quantitative, high-resolution view of metabolism. Among ~1100 unique enzyme-coding gene families in the oral metagenome, ~18% were differentially expressed ($P < 0.05$) at the microbiome level during disease (Fig. 5.4). Differentially expressed gene families were mapped to the Kyoto Encyclopedia of Genes and Genomes [213] metabolic pathway database to reveal enzymatic steps whose genes were up-regulated, down-regulated, or unchanged in the microbiome during disease (Fig. 5.5). Many enzyme-coding genes related to amino acid metabolism were up-regulated, including pathways involved in lysine and histidine degradation as well as genes involved in butyrate production. These results suggest that histidine and lysine consumption by the disease-associated microbiome is important during aggressive periodontitis. This is further supported by up-regulation of numerous protease genes during disease, which degrade extracellular proteins and liberate amino acids [214-216] (Fig. 5.6a, Table 5.5).

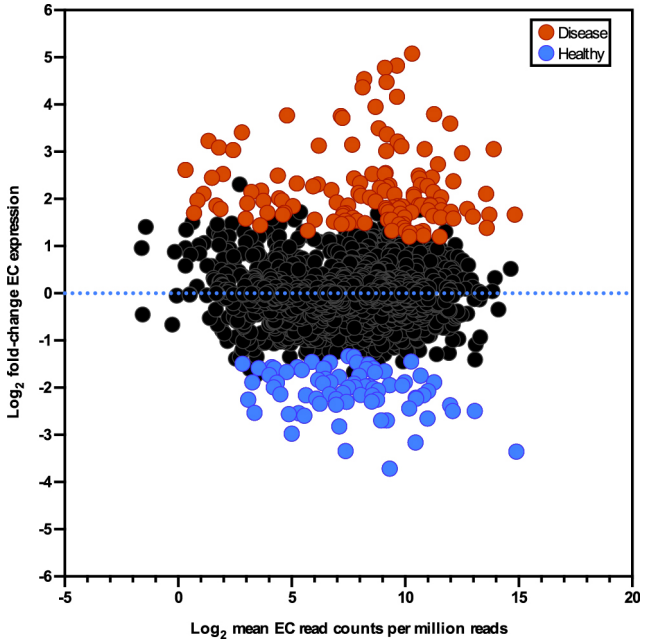


Figure 5.4 Differential expression of enzyme gene families in health and disease.

Log_2 fold-change during disease is plotted against the log_2 mean read counts per million total reads for each EC enzyme-coding gene family. Gene families up-regulated in health are shown in blue, while gene families up-regulated in disease are shown in red.

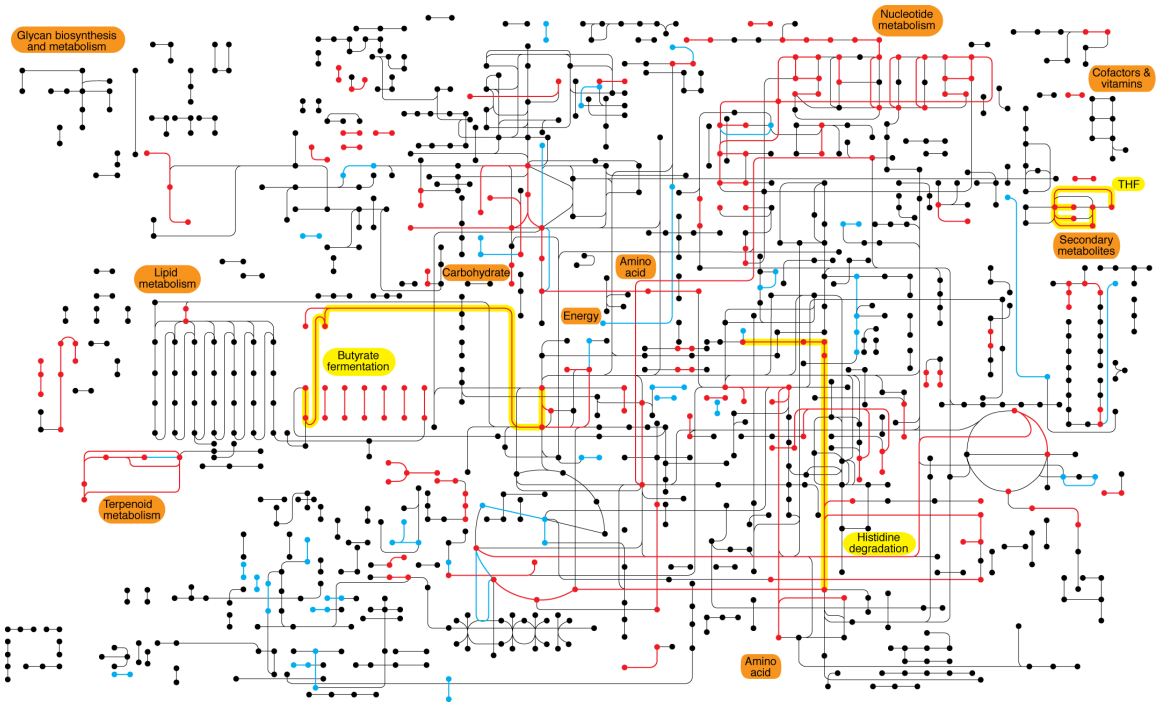


Figure 5.5 Differential metabolic gene expression in the diseased periodontal microbiome.

Metabolic network reconstruction. Black lines indicate enzyme-coding genes that were expressed and unchanged in health and disease, red lines indicate up-regulated genes during disease, and blue lines indicate up-regulated genes during health. Orange labels identify different regions of the metabolic pathway map. Those highlighted in yellow represent whole pathways that were up-regulated in disease. THF: tetrahydrofolate metabolism.

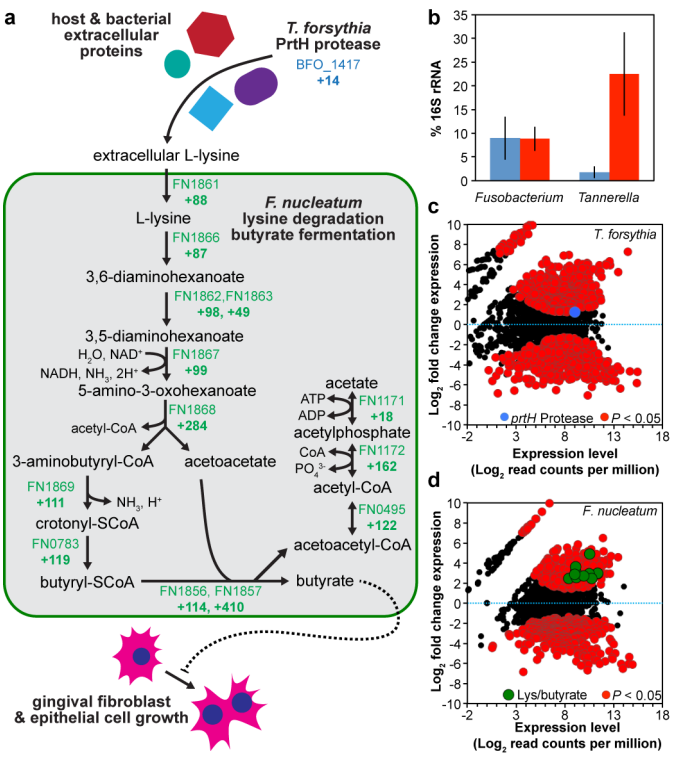


Figure 5.6 The microbiome alters behavior during periodontal infection.

(a) Although less than 10% of the microbiota in diseased and healthy communities, the keystone species *F. nucleatum* mediates enhanced lysine fermentation in diseased communities to produce butyrate. Butyrate is present at high levels in periodontal disease and prevents gingival fibroblast growth. *T. forsythia* enhances expression of the protease-encoding gene *prtH*, which likely liberates amino acids in diseased sites. Locus tags are indicated for each enzyme-coding gene, and the fold-change in gene expression at the community level during disease is indicated below each locus tag. **(b)** Relative abundance of *Fusobacterium sp.* and *Tannerella sp.* rRNA during health and disease. Relative abundances were determined using qrRNA-seq (Extended Data Fig. 1b). **(c)** Differential gene regulation of *T. forsythia* in diseased sites. Log_2 fold-change in gene expression during disease is plotted against log_2 mean read count per million reads (n=3). **(d)** Differential gene regulation of *F. nucleatum* in diseased sites. Graph is depicted as in (c).

Table 5.5 Virulence factors are up-regulated during disease due to differential gene regulation and change in population composition.

Species	Locus ID	Gene FC*	rRNA FC†	Gene *P*-value‡	Gene/product§
Tannerella forsythia	BFO_1352	53	13	3.37E-04	Collagenase
Tannerella forsythia	BFO_1417	15	13	1.49E-02	*prtH*, protease
Tannerella forsythia	BFO_2000	34	13	1.10E-02	*bspA*, bacterial group 2 Ig-like protein
Porphyromonas gingivalis	PG_0026	3	1	2.68E-01	Gingipain
Porphyromonas gingivalis	PG_0506	3	1	4.68E-01	Gingipain, arginine-specific cysteine proteinase
Porphyromonas gingivalis	PG_1542	2	1	5.13E-01	Collagenase
Porphyromonas gingivalis	PG_2024	16	1	9.94E-02	*hagE*, hemagglutinin protein HagE
Fusobacterium nucleatum	FN1668	37	1	2.92E-06	Cholinephosphate cytidylyltransferase, LPS modification
Fusobacterium nucleatum	FN1669	15	1	1.65E-04	Choline transport protein, LPS modification
Fusobacterium nucleatum	FN1670	4	1	2.47E-02	Choline kinase, LPS modification
Prevotella tannerae	ptan_c_1_1221	5	19	9.11E-02	*prtC*, Collagenase
Prevotella intermedia	PIN17_0033	11	4	1.32E-02	*inpA*, interpain A peptidase C10 family protein

*Fold-change in gene expression at the community level during disease.

†Fold-change in rRNA levels during disease.

‡*P*-value for change in gene expression at the community level during disease.

§Gene product description.

5.2.5 The keystone species *F. nucleatum* delays wound healing through differential gene expression of lysine fermentation genes

The observed metabolic changes in the diseased community could have resulted from either increased abundance of organisms carrying out those processes or differential gene regulation by individual organisms during infection. Our high-resolution metaRNA-seq analyses allow differentiation of these two possibilities through analysis of gene expression of individual species during health and disease. These analyses showed that the Gram-negative organisms *Fusobacterium nucleatum* and *Tannerella forsythia* were primary sources of metabolic gene expression changes in the diseased community (Fig. 5.5 and 5.6a). Increased expression of *T. forsythia* genes in the diseased community could be attributed to both changes in abundance of this bacterium (i.e. changes in gene expression were equivalent to abundance changes, *prtH* in Fig. 5.6bc and Table 5.5) and differential gene expression. In contrast, *F. nucleatum* maintained constant relative abundance in both healthy and diseased communities (Fig. 5.6b), yet highly up-regulated many genes, including those involved in lysine fermentation to produce butyrate (Fig. 5.6d). Butyrate production has been observed at high levels in infected periodontal sites [217] and inhibits growth of gingival fibroblasts [218], thus potentially delaying wound healing. Therefore, *F. nucleatum* likely plays an important role in disease exacerbation and delayed healing via butyrate production. We propose *F. nucleatum* is a keystone species during periodontitis, functioning not by changing its prevalence during disease but by shifting metabolism of the community to produce a metabolite known to be important for maintaining the diseased state.

This study provides the first differential gene expression analysis of the human microbiome during health and disease. Using high-resolution transcriptomics combined

with microbiome community analyses, we have shown changes in both composition and gene expression of the microbiome during disease. Furthermore, our results emphasize the importance of studying behavior of the microbiome during disease, since constant-abundance species, like *F. nucleatum*, can differentially regulate pathways that have a profound impact on infection.

5.3 METHODS

5.3.1 Study population

A total of 10 individuals seeking dental treatment in the School of Dentistry, Ege University, İzmir, Turkey, were involved in the present study. Systemically healthy, untreated 10 patients with generalized aggressive periodontitis (AgP) were recruited from September 2011 to August 2012 (Table 5.1). The study was conducted in full accordance with ethical principles, including the World Medical Association's Declaration of Helsinki, as revised in 2000. The study protocol was explained, and written informed consent was received from each individual before clinical periodontal examinations and subgingival plaque sampling. Medical and dental histories were obtained and smoking habits were recorded. Individuals with medical disorders, such as diabetes mellitus, immunological disorders and those who had antibiotic or periodontal treatment in the last 6 months were excluded from the study.

Individuals with AgP were diagnosed in accordance with the clinical criteria stated in the consensus report of the World Workshop in Periodontitis. Individuals had at

least 6 permanent teeth, including incisors and/or first molars, with at least one site with probing depth (PD) and clinical attachment loss (CAL) ≥ 5 mm and 6 teeth other than first molars and incisors with similar PD and CAL measurements, familial aggregation (all individuals were asked if they had any family member with current or history of severe periodontal disease).

5.3.2 Subgingival plaque sampling

For the diseased samples, the deepest 3 pockets were selected and pooled in a single eppendorf tube. Supragingival plaque was first removed from the sample teeth with sterilized Gracey curettes and sterilized gauze. The site was then cleaned and isolated using cotton roles and air dried gently. Another sterilized Gracey curette was inserted to the deepest part of the pocket and removed applying a slight force towards the root surface. The tip of the curette was then inserted in the eppendorf tube containing RNA Later and shaken until the plaque was removed from the curette. For the healthy subgingival plaque samples, in the same patient 3 healthy sites that did not show any sign of inflammation and bleeding on probing were chosen and pooled in an eppendorf tube. The same procedures were followed for the subgingival sampling. After 24 h, the samples were frozen and stored at -40ºC until the sample collection period was completed.

5.3.3 Clinical periodontal measurements

Subsequent to saliva and serum sampling, clinical periodontal recordings, including plaque index, PD, CAL, and bleeding on probing (BOP) (+/-) were performed

at 6 sites (mesio-buccal, mid-buccal, disto-buccal, mesio-lingual, mid-lingual and disto-lingual locations) on each tooth present, except the third molars, using a Williams periodontal probe. CAL was assessed from the cement enamel junction to the base of the probable pocket. BOP (deemed positive if it occurred within 15 seconds after periodontal probing) was recorded dichotomously by visual examination. All measurements were performed by two precalibrated examiners. Inter-examiner and intra-examiner calibration was analyzed using the Kappa-Cohen test. The initial intraexaminer kappa values were 0.96 (PD) and 0.86 (CAL) for PG and 0.93 (PPD) and 0.79 (CAL) for NN. The interexaminer values were 0.92 (PD) and 0.75 (CAL).

5.3.4 Total RNA isolation

Subgingival plaque samples stored in RNA Later were centrifuged at 16,100 *g* to collect whole cells. Cell pellets were resuspended in 1 ml RNA Bee and transferred to a bead-beating tube. Cells were lysed by bead-beating 3 times for 60 s and incubating on ice 1 m in between each bead beating. Lysed cell solutions were transferred to new microcentrifuge tubes, 200 µl chloroform was added. Tubes were shaken vigorously for 1 m to mix, and incubated 5 m in an ice bath. Samples were centrifuged at 13,100 *g* for 30 m at 4 °C to separate aqueous and organic phases. The aqueous phase from each sample was transferred to a new microcentrifuge tube and RNA was precipitated with an equal volume of isopropanol and 2 µg linear acrylamide for 16 h at -80°C. Samples were thawed in an ice bath and centrifuged for 30 m at 13,100 *g* at 4°C. Supernatants were removed, and RNA pellets were washed with twice with ice-cold 75 % ethanol by resuspension and centrifugation for 10 m at 16,100 *g* at 25°C. Following the second ethanol wash, RNA pellets were air-dried 5 m at 25°C and resuspended in 22 µl RNase-

free water. RNA concentrations for each sample were determined with a Nanodrop spectrophotometer (Thermo Scientific).

5.3.5 qrRNA-seq

qrRNA-seq was developed by modifying 2 protocols from previous studies which sequenced bacterial 16S rDNA [203, 204]. Total subgingival plaque RNA for all 10 healthy and diseased samples was used to reverse transcribe 16S cDNA with SSII reverse transcriptase (Invitrogen) and the universal bacterial 16S 926 RT gene-specific primer (Table 5.6) which anneals immediately downstream of the 16S rRNA V5 variable region. Negative control reactions lacking SSII were conducted on all RNA samples in parallel to ensure that DNA was not co-purified with the total RNA. From each RT reaction, including negative control reactions, 2 μl were removed and cDNA was used as template to minimally PCR amplify the 16S rRNA V4/V5 variable region using indexed sample-specific primers 16SV5926R-BC0 through 16SV5926R-BC19 and the common primer 16SV4515F (Table 5.6). All RT-PCR reaction products were separated by agarose gel electrophoresis, stained with ethidium bromide, and viewed with a GBox imaging system. Distinct cDNA bands were visible for all positive control reactions, while negative control reactions lacking RT showed no product, verifying the absence of DNA contamination in the original RNA preparations. Paired end 250 bp sequencing was performed on the 16S cDNA libraries using an Illumina MiSeq at the University of Texas Genomic Sequencing and Analysis Facility (UTGSAF) with custom MiSeq16SV4515F forward, MiSeq16SV5926R reverse, and MiSeq16SV4V5Index index sequencing primers (Table 5.6).

Table 5.6 Primer information.

Primer name	Sequence	Sample
16S926RT	CCGTCAATTYYTTTRAGTTT	
16SV4515F	AATGATACGGCGACCACCGAGATCTACACTATGGTAATTGTGTGC CAGCMGCCGCGGTAA	
16SV5926R-BC0	CAAGCAGAAGACGGCATACGAGATTCCCTTGTCTCCAGTCAGTCA G CC CCGTCAATTYYTTTRAGTTT	D1
16SV5926R-BC1	CAAGCAGAAGACGGCATACGAGATACGAGACTGATTAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	D2
16SV5926R-BC2	CAAGCAGAAGACGGCATACGAGATGCTGTACGGATTAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	D3
16SV5926R-BC3	CAAGCAGAAGACGGCATACGAGATATCACCAGGTGTAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	D4
16SV5926R-BC4	CAAGCAGAAGACGGCATACGAGATTGGTCAACGATAAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	D5
16SV5926R-BC5	CAAGCAGAAGACGGCATACGAGATATCGCACAGTAAAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	D6
16SV5926R-BC6	CAAGCAGAAGACGGCATACGAGATGTCGTGTAGCCTAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	D7
16SV5926R-BC7	CAAGCAGAAGACGGCATACGAGATAGCGGAGGTTAGAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	D8
16SV5926R-BC8	CAAGCAGAAGACGGCATACGAGATATCCTTTGGTTCAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	D9
16SV5926R-BC9	CAAGCAGAAGACGGCATACGAGATTACAGCGCATACAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	D10
16SV5926R-BC10	CAAGCAGAAGACGGCATACGAGATACCGGTATGTACAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	H1
16SV5926R-BC11	CAAGCAGAAGACGGCATACGAGATAATTGTGTCGGAAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	H2
16SV5926R-BC12	CAAGCAGAAGACGGCATACGAGATTGCATACACTGGAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	H3
16SV5926R-BC13	CAAGCAGAAGACGGCATACGAGATAGTCGAACGAGGAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	H4
16SV5926R-BC14	CAAGCAGAAGACGGCATACGAGATACCAGTGACTCAAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	H5
16SV5926R-BC15	CAAGCAGAAGACGGCATACGAGATGAATACCAAGTCAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	H6
16SV5926R-BC16	CAAGCAGAAGACGGCATACGAGATGTAGATCGTGTAAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	H7
16SV5926R-BC17	CAAGCAGAAGACGGCATACGAGATTAACGTGTGTGCAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	H8
16SV5926R-BC18	CAAGCAGAAGACGGCATACGAGATCATTATGGCGTGAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	H9
16SV5926R-BC19	CAAGCAGAAGACGGCATACGAGATCCAATACGCCTGAGTCAGTC AG CC CCGTCAATTYYTTTRAGTTT	H10
MiSeq16SV4515F	TATGGTAATTGTGTGCCAGCMGCCGCGGTAA	
MiSeq16SV5926R	AGTCAGTCAGCCCCGTCAATTYYTTTRAGTTT	
MiSeq16SV4V5Index	AAACTYAAARRAATTGACGGGGCTGACTGACT	

5.3.6 Bacterial population analyses

The paired 250 bp forward and reverse MiSeq sequencing reads were assembled using fastq-join [219]. Qiime [209] was used to search the assembled 16S cDNA sequences from all 10 healthy and 10 diseased samples with Uclust against the 97% Greengenes reference database [208] for species-level identification of Operational Taxonomic Units (OTUs) using the Qiime python script pick_otus_through_otu_table.py. Prior to alpha diversity analyses, samples were rarefied, or sub-sampled, 10 times at each step from 500 to 5000 sequences with a 500-sequence step-size. Mean alpha diversity, or within sample diversity, was calculated using the Qiime python scripts alpha_diversity.py and collate_alpha.py to determine the number of observed species at each sub-sampling depth in the rarefication analysis. To determine the relative contributions of high abundance species to total diseased and healthy populations, unclassified and non-specific OTUs assigned to bacteria were removed from each sample. Within each sample, abundances of OTUs were sorted from most to least abundant, and a running sum percentage of the total population was calculated for the addition of each additional species. Jackknifed beta-diversity, or between sample diversity, was determined for 5000 sequences per sample using the Qiime python script jackknifed_beta_diversity.py. A multidimensional scaling (MDS) analysis plot was generated from the average of 10 distance matrices determined by unweighted Unifrac analysis [220] calculated by the jackknifed_beta_diversity.py script and was used to determine the similarity between sample populations. Briefly, Unifrac analysis takes into account the number of shared and unique species between two populations, and provides a distance metric that represents the overall similarity of the two populations [220]. Healthy and disease centroids on the MDS plot were determined from the mean positions of the respective samples on the plot.

Euclidean distances were calculated to determine relatedness between paired healthy and diseased samples and from samples to their respective healthy or disease centroids.

5.3.7 metaRNA-seq

Patients 5, 6, and 8 were selected for total RNA sequencing to analyze microbial population gene expression in periodontal health and disease. Total RNA samples were treated with the RiboZero Epidemiology kit (Epicentre) to deplete bacterial and eukaryotic rRNA and purified by ethanol precipitation using 20 μg linear acrylamide to precipitate the RNA. Depleted RNA was fragmented with NEB RNA Fragmentation Buffer, according to the manufacturer's protocol. Fragmented RNA was ethanol precipitated with linear acrylamide and eluted in RNase-free water. RNA-seq libraries were prepared using the NEB Next Multiplex Small RNA Library Prep Set for Illumina, according to the manufacturer's protocol. The resulting strand-specific cDNA libraries were stained with SYBR gold nucleic acid stain (Invitrogen), visualized on a GBox imaging system, and cDNA between ~150-300 bp was extracted, corresponding to fragmented RNA between 31-181 nt. Gel extracted cDNA was eluted in NEB polyacrylamide gel elution buffer, ethanol precipitated, and resuspended in TE buffer (NEB). Libraries were quantified and analyzed using a Nanodrop spectrophotometer (Thermo Scientific) and a Bioanalyzer (Agilent). Single end 50 bp sequencing was conducted at the UTGSAF on an Illumina HiSeq2000 producing ~1.5 billion sequencing reads (Table 5.3).

5.3.8 metaRNA-seq fastq read processing

HiSeq reads were trimmed with Flexbar [170], as described previously [221], to remove contaminating adapter sequences from the cDNA library preparation. Flexbar was run with settings to collect reads 15-50 bp following adapter trimming for further analysis, because these reads are specific: 15 bp sequences are predicted to occur randomly only once per ~1 billion bp.

5.3.9 Determining origin of metatranscriptome sequencing reads

All metatranscriptome data analysis was conducted on the Texas Advanced Computing Center Stampede supercomputer. Human oral bacteria genome sequences (oral_microbiome.na.zip, >4 billion bp) were downloaded from the Human Oral Microbiome Database [210] (HOMD) available on the world wide web via ftp://ftp.homd.org/human_oral_microbial_genomic_sequences/20130520/, the human RNA database (human.rna.fna.gz) was downloaded on the world wide web through NCBI RefSeq via ftp://ftp.ncbi.nlm.nih.gov/refseq/H_sapiens/mRNA_Prot/, and the viral genome database (viral.1.1.genomic.fna.gz, ~121 million bp) consisting of sequenced viruses and bacteriophage available was downloaded through NCBI RefSeq via ftp://ftp.ncbi.nih.gov/refseq/release/viral/. Reference sequences were indexed with Bowtie 2.0 [171]. Since the HOMD sequences exceeded the size limit for Bowtie 2.0 [171], the sequences were split into two files with the custom Perl script FastaSplit.pl (http://github.com/khturner/metaRNA-seq) and then each file was indexed. Trimmed fastq sequencing reads were split into chunks of 10 million reads using the UNIX split command. Each read chunk was mapped separately to the four indexed references sequences using Bowtie 2.0 [171], keeping only 1 match for each fastq read for each

reference. Unmapped reads were discarded and mapped reads were labeled to indicate whether they mapped to either of the 2 human oral microbiome indexed databases (HOMD1 and HOMD2), the indexed human RNA database, or the indexed viral RNA database. The resulting labeled mapped reads in sam file format for each read chunk in each sample were concatenated and sorted by read name using the UNIX cat and sort commands. Since initially we were interested in whether a read was of bacterial origin, human origin, or viral origin, if a single fastq read mapped to both HOMD1 and HOMD2, the read mapped to HOMD2 was discarded; however, if a read mapped to multiple references (e.g. HOMD and human) it was labeled in the file using the custom Perl script MatchMarker.pl (http://github.com/khturner/metaRNA-seq). The number of uniquely mapping reads and reads mapping to multiple reference databases for each sample was determined using the UNIX uniq and pattern-matching grep commands.

5.3.10 Generating a reference metagenome for differential gene expression analysis

Genomes for differential gene expression analyses were selected using 16S rRNA sequencing for patients 5, 6, and 8. Reference genome sequences and annotations were downloaded in Fasta and GFF formats, respectively. Genomes were downloaded, concatenated, and processed to only include protein-coding genes using the custom Perl scripts GenomeMerge.pl and HOMDpull.sh (http://github.com/khturner/metaRNA-seq) to generate a metagenome to serve as a reference. Individual genome sequences and annotations were obtained from NCBI Genbank (ftp://ftp.ncbi.nih.gov/genbank) and HOMD (http://www.homd.org/index.php?&name=seqDownload&type=G). When available, EC numbers for genes were downloaded from the Kyoto Encyclopedia of

Genes and Genomes (KEGG) [213] using the custom Perl scripts PullEC.pl and HOMD_GenomeMerge.pl (http://github.com/khturner/metaRNA-seq).

5.3.11 Differential gene expression analyses

Trimmed RNA-seq reads produced by Flexbar[27] were mapped against the indexed reference metagenome and reads mapping to each gene were counted using the custom UNIX shell script MapCount_RNASeq.sh (http://github.com/khturner/metaRNA-seq), which depends on Bowtie 2.0 [171] and the Python package HTSeq (https://pypi.python.org/pypi/HTSeq). The trimmed sequencing reads are read into the script, mapped to the metagenome, and the number of reads in each sample mapping to each annotated gene in the metagenome are counted. Paired differential gene expression was determined using the custom UNIX shell script calcRNASeqPaired.sh (http://github.com/khturner/metaRNA-seq), which depends on the R package edgeR[29] and the supporting R script Pairwise_edgeR.r (http://github.com/khturner/metaRNA-seq). This analysis normalizes read counts between samples, fits the data to a negative binomial distribution, and determines pairwise differential expression using the patient-matched samples.

5.3.12 Differential expression analysis of KEGG enzymes

EC numbers obtained from HOMD and the KEGG databases were added to the table containing raw read counts per gene produced by MapCount_RNASeq.sh (above). Genes lacking EC numbers were removed from the table and the table was sorted by the EC numbers. The total number of reads mapping to each EC number was calculated

using the custom Perl script ECcounter.pl (http://github.com/khturner/metaRNA-seq), to produce a table containing number of reads mapping to each EC number in each sample. Differential expression of EC enzymes was determined using the custom UNIX shell script Pairwise_edgeR.sh (http://github.com/khturner/metaRNA-seq), which depends on the R package edgeR [222] and the supporting R script Pairwise_edgeR.r (http://github.com/khturner/metaRNA-seq).

Chapter 6. Discussion

The beginning of this dissertation focused on characterizing genetic regulatory mechanisms and examining the evolutionary history of the opportunistic pathogen *A. actinomycetemcomitans*. Using directed molecular probing of non-coding RNAs and whole transcriptome profiling, I described numerous genetic regulatory mechanisms employed by *A. actinomycetemcomitans* to respond to its environment. I also showed that the evolutionary history of *A. actinomycetemcomitans* is tightly linked to natural transformation and CRISPR genome defense. Through transcriptional profiling, I provided the first comprehensive view of *A. actinomycetemcomitans* transcription and determined the importance of oxygen-independent metabolism for survival *in vivo*.

This dissertation work culminated in the examination of community behavior of the oral microbiome in health and disease. This study showed for the first time that the microbiome differentially regulates gene expression during disease. Transcriptional profiling of individual community members showed that metabolic gene expression changes in a single species might prolong disease. Together these data suggest that keystone species may not influence polymicrobial disease through "traditional" virulence factors, like toxins, but do so through basic metabolic processes that are impacted by other community members.

6.1 *A. ACTINOMYCETEMCOMITANS* GENE REGULATION AND EVOLUTION

6.1.1 Switching off expression: knowing when enough is enough

Using Northern blot analysis I discovered the first ncRNAs in *A. actinomycetemcomitans*. During these studies I characterized a new class of bacterial lysine-specific transporter, encoded by the widespread *lysT* gene. Northern blot analyses showed that *lysT* is transcriptionally regulated by a lysine riboswitch. The riboswitch functions by shutting down *lysT* expression at a threshold intracellular lysine concentration. This mechanism theoretically allows *A. actinomycetemcomitans* and other organisms to divert resources from *lysT* expression to other necessities once enough lysine has been acquired from the environment. This study also opens up the possibility that other putative NhaC-type transporters have roles beyond sodium and hydrogen transport.

6.1.2 A CRISPR view of *A. actinomycetemcomitans* evolution

Phylogenetic characterization of *A. actinomycetemcomitans* showed that its evolution is linked to natural transformation. These analyses provide strong evidence that, in the case of *A. actinomycetemcomitans,* natural transformation is under selection due to its effect on genetic expansion. This hypothesis is based on the loss of competence genes throughout *A. actinomycetemcomitans* evolution, followed by the inactivation of CRISPR genome defense systems. Non-competent strains were consequently more vulnerable to minor genetic expansion by acquiring mobile genetic elements. In contrast, competent strains experienced continued genetic variation through both genome expansion and gene rearrangement.

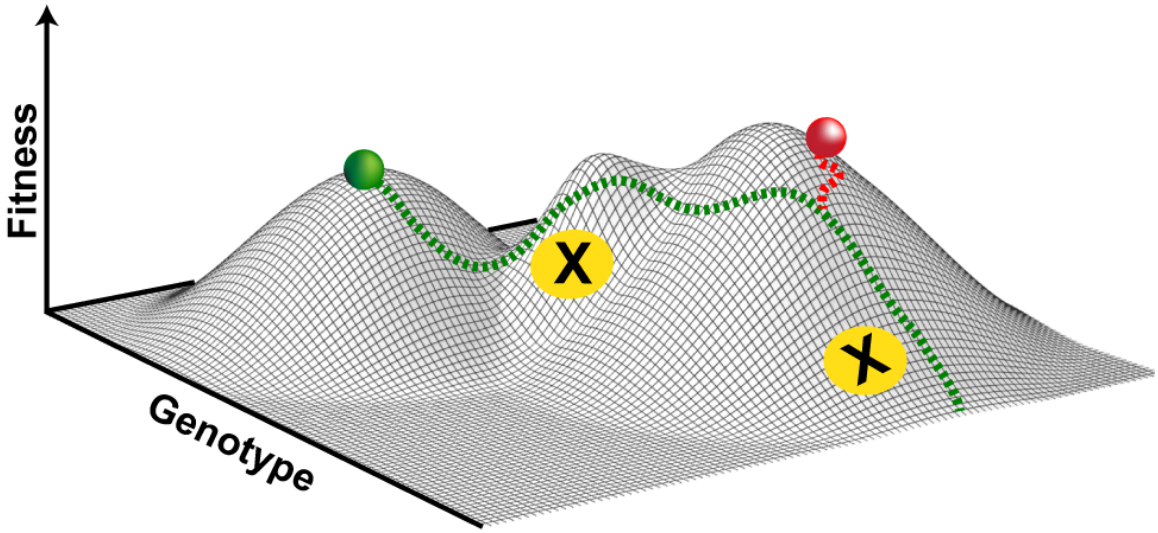


Figure 6.1 Navigating the genomic fitness landscape through genome expansion and reduction.

Non-competent *A. actinomycetemcomitans* strains constantly arise in competent populations throughout evolution; however, non-competent strains only persist when near fitness peaks (red circle) and are otherwise outcompeted by competent strains away from fitness peaks (yellow X's). The evolution of competent strains (green circle, green line) through genomic rearrangements and acquisition of new genes (Fig. 3.2) leads to speciation, as competent strains can explore the fitness landscape to a greater extent than non-competent strains. The loss of CRISPR-*cas* allows non-competent bacteria to acquire new genes through bacteriophage and plasmids to ascend local fitness peaks (red line). In this context, genotype refers to genome content including mutations, new genes, and rearrangements.

The evolutionary connection between natural transformation and CRISPR immunity highlights a balancing act between genome expansion and reduction (Fig. 6.1). We propose a model wherein natural transformation permits exploration of a genomic fitness landscape through the acquisition of new genetic material. Near fitness peaks, selection for exploration through transformation may be weakened, and competence genes can be lost to inactivating mutations. However, due to new selective pressures, genome expansion through mobile genetic elements is favored, and inactivating CRISPR mutations arise. Long-term *in vitro* evolution studies could be designed to test this theory.

6.1.3 *A. actinomycetemcomitans* behavior *in vivo*

Transcriptional profiling revealed that fermentation and anaerobic respiration impact *A. actinomycetemcomitans* survival in the murine abscess infection. While *A. actinomycetemcomitans* possesses multiple fermentative and anaerobic respiratory pathways, two specific operons, *frdABCD* and *fdhF1F2*, were up-regulated *in vivo* and mutants in these operons had impaired fitness (Fig. 4.5). This study provided a comprehensive transcriptional map for *A. actinomycetemcomitans in vivo* and in a single species biofilm, as I identified 691 transcriptional start sites and 210 ncRNAs. Among the putative ncRNAs identified, 80 were differentially regulated *in vivo*, suggesting that they may play important roles in the host environment. In addition to providing insight into the genetic regulatory mechanisms mediating persistence *in vivo,* this study also lays an essential foundation for studying gene expression changes impacting synergistic polymicrobial infections.

6.1.4 Future directions

One of the primary goals of this dissertation was to develop a foundation in *A. actinomycetemcomitans* gene regulation that could be used to define genes and regulatory networks underlying polymicrobial synergy. Because *A. actinomycetemcomitans* and *S. gordonii* exhibit synergy in a mixed species infection, the obvious next step in these studies is to use RNA-seq to transcriptionally profile *A. actinomycetemcomitans* and *S. gordonii* murine abscess co-infections. Since *apiA* and *katA* genes were up-regulated by hydrogen peroxide and *S. gordonii in vitro* [16], we anticipate that these genes will be up-regulated in co-infections as well. These genes should serve as controls. Using RNA-seq we will identify global changes in gene expression in addition to *katA* and *apiA* that likely contribute to polymicrobial fitness.

To begin defining gene regulation mediating synergy in co-infections, I have analyzed *A. actinomycetemcomitans* gene expression in co-culture biofilms and murine abscess co-infections with *S. gordonii*. In these preliminary analyses *A. actinomycetemcomitans* up-regulates expression of both *apiA* and *katA* in the presence of *S. gordonii*, as expected (Fig. 6.2AB). However, these genes were up-regulated to a lesser extent *in vivo* than *in vitro*. In a single species infection, *A. actinomycetemcomitans* expressed genes involved in anaerobic metabolism, suggesting limited oxygen availability. Since hydrogen peroxide is produced by *S. gordonii* in an oxygen-dependent manner, and *katA* and *apiA* are induced by *S. gordonii*-produced hydrogen peroxide, these results could be explained by the reduced availability of oxygen in the abscess relative to the biofilm.

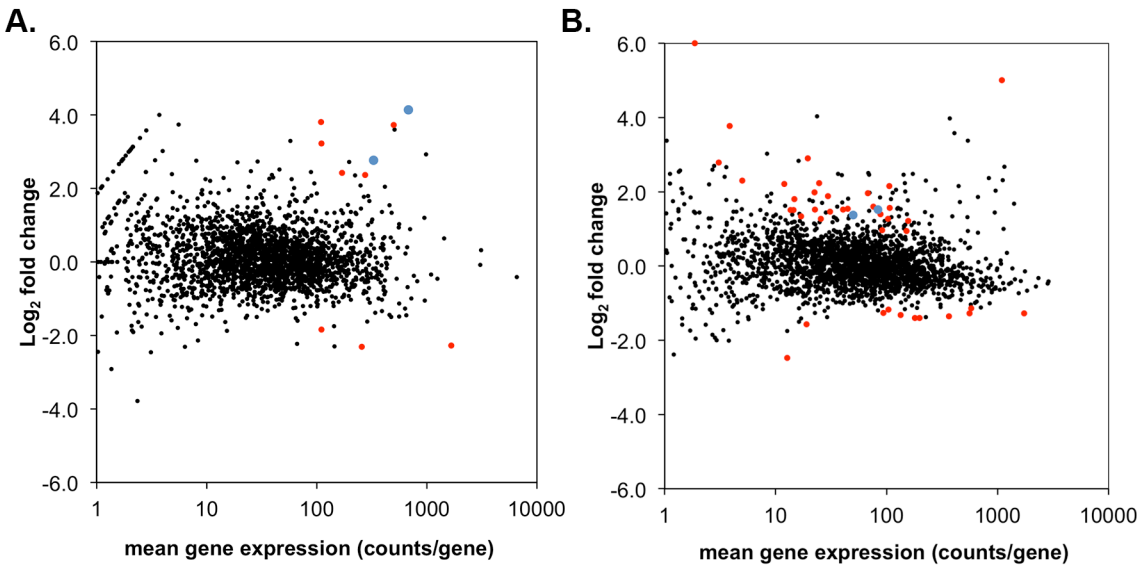


Figure 6.2 Differential gene regulation of *A. actinomycetemcomitans in vitro* and *in vivo* during growth with *S. gordonii*.

(A) Gene expression of *A. actinomycetemcomitans* during co-culture biofilm growth with *S. gordonii in vitro*. *A. actinomycetemcomitans* was grown with *S. gordonii* in colony biofilms grown on a solid glucose defined medium and gene expression was compared to single species biofilms using RNA-seq as described in Chapter 4 (n = 2). **(B)** Gene expression of *A. actinomycetemcomitans* during murine abscess co-infection with *S. gordonii*. *A. actinomycetemcomitans* and *S. gordonii* were co-incoulated into abscesses as described previously [22]. RNA was isolated from abscesses at 3 days post-infection and gene expression was compared to single species infections using RNA-seq, as described in Chapter 4 (n = 3). In both A and B, blue points indicate *katA and apiA*, and red points indicate significant gene expression changes in co-culture ($P < 0.05$, DESeq).

Only one other protein-coding gene, the putative xylose transport gene *xylF*, was commonly up-regulated during growth with *S. gordonii in vitro* and *in vivo*. This observation is interesting considering that the *A. actinomycetemcomitans* strain used in our RNA-seq experiments was a serotype A strain, and serotype A strains are incapable of fermenting xylose [223]. Furthermore, xylose is a sugar commonly found in plants and not produced in the oral cavity. This result has several implications. First, it may indicate that xylose fermentation was once important for polymicrobial interactions in ancestral environmental relatives of *A. actinomycetemcomitans*, but throughout the course of evolution selective pressures maintaining active xylose fermentation genes have weakened. Second, it suggests that certain strains have experienced selective pressure to maintain xylose fermentation systems. If so, these strains would only encounter xylose from dietary sources. Either way, these results suggest that in the presence of streptococci, *A. actinomycetemcomitans* increases xylose transport. *S. gordonii* does not ferment xylose, so this may represent another way that *A. actinomycetemcomitans* avoids competition for carbon sources with faster-growing community members.

RNA-seq provides a global view of gene expression and can be used to identify genes impacting fitness. However, the results obtained from RNA-seq are dependent on the choice of control conditions, and genes impacting fitness that are not differentially regulated may be overlooked by this technique. Transposon sequencing (Tn-seq) is a complementary approach to RNA-seq that can be used to define genome-wide fitness determinants [224]. The technique utilizes a saturating transposon mutant library and high-throughput sequencing to identify mutants that are less fit under a chosen selective condition. Tn-seq requires a saturating transposon mutant library. This library should

contain a pool of individual insertion mutants, representing unique mutants for every 20 bp in the genome. An *A. actinomycetemcomitans* Tn-seq library should have ~100,000 unique mutants. The abundance of mutants in this library can be determined by sequencing transposon adjacent DNA before and after exposure to a selective condition. RNA-seq can be used to identify differentially expressed genes in co-infection, while Tn-seq can be used to measure mutant fitness independent of expression changes. Combining Tn-seq with RNA-seq can therefore provide powerful comprehensive information about a polymicrobial infection.

Studying pairwise interactions will inform us about genes impacting synergy in polymicrobial infections, but ultimately it will be important to examine behavior within more complex communities. One study used metatranscriptomics to examine transcription within defined *in vitro* oral microbial communities composed of "commensals" and "commensals and pathogens;" however, this study was limited by the experimental design, which focused only on "commensal" gene expression [225]. Here I propose a step-wise additive approach for studying polymicrobial communities: microbes should first be studied in pairwise combinations, and then more complex interactions can be explored by adding organisms one at a time to the community. This approach will allow the precise definition of the effects each organism has on the community.

6.2 IMPLICATIONS FOR GENE EXPRESSION CHANGES IN THE MICROBIOME

6.2.1 Gene expression of the oral microbiome

Using qrRNA-seq and metaRNA-seq, I demonstrated that periodontitis is characterized by changes in microbial population structure, as well as differential gene expression at the community level and within individual species (Fig. 5.1-5.6, Table 5.5.). Specifically, diseased microbial populations adopted a similar composition that was less diverse than healthy oral microbial populations (Fig. 5.2). A few highly abundant organisms dominated diseased populations, whereas many low abundance species comprised healthy populations. As a community, gene expression changes revealed a shift toward amino acid consumption. This metabolic shift was paired with increases in the expression of proteases, which are important virulence factors. Many of the community gene expression changes could be attributed to *T. forsythia* and *F. nucleatum*. While *T. forsythia* ribosomal content increased during disease, *F. nucleatum* remained equally abundant within the community. This suggests that *F. nucleatum* expression changes were due to differential gene regulation rather than mere changes in abundance. A consequence of differential gene expression in *F. nucleatum* was the increase in lysine fermentative gene expression. Lysine fermentation by *F. nucleatum* produces the oral inflammatory molecule butyrate [226]. Since butyrate arrests fibroblast cell growth and likely delays wound healing [218], we propose that *F. nucleatum* is a keystone species in the diseased periodontal microbial community.

6.2.2 Future directions

These pilot studies revealed microbial behaviors that likely play important roles in oral disease progression. However, a key question remains unanswered. Are changes in microbial behavior a cause or consequence of periodontitis? While this question is difficult to answer, careful experimental design may provide a way to get at the root of the problem in periodontal disease. A multifacteted approach combining bacterial ecology, bacterial genetics, and host genetics showed that *P. gingivalis*-induced murine periodontitis was dependent on specific host and *P. gingivalis* genes, as well as the commensal microbiota [214]. A similar approach could be used to test the role of *F. nucleatum* butyrate production in periodontal disease. An alternate approach would be to use metaRNA-seq to characterize different stages of periodontal disease. A variety of patient plaque samples from healthy, moderate, and severe periodontal pockets would need to be collected for analyses. If gene microbial gene expression changes are observed in moderate pockets before the onset of fully developed periodontitis, one could conclude that microbial gene expression changes are leading to disease. However, if the opposite is true, and fully developed periodontitis precedes changes in the microbiota, then other factors likely lead to disease.

It will be interesting to compare microbial community compositions and behavior in different types of periodontal disease. Periodontitis is a general term that describes several different diseases. We used metaRNA-seq to analyze gene expression in aggressive periodontitis in adults. Aggressive periodontitis in adolescents is one major form of periodontal disease and is typically associated with *A. actinomycetemcomitans*. Because we have studied *A. actinomycetemcomitans* polymicrobial interactions *in vitro*

and in murine abscess infections, metaRNA-seq provides a complementary tool to test the hypotheses we developed about polymicrobial infections using streptococci. In contrast, chronic periodontitis is more commonly associated with *P. gingivalis, T. denticola*, and *T. forsythia* [19]. Applying metaRNA-seq to study these infections will help test hypotheses about these disease processes, specifically how they relate to each other, how they differ, and the potential roles the microbes play in these diseases.

Finally, metaRNA-seq is a powerful tool that can be applied to study any microbial community. Potential applications for metaRNA-seq include studying other microbiome-mediated diseases, diseases caused by invasive pathogens, or microbial communities in the environment. The gut microbiome has been of extreme interest due to population shifts that have been associated with obesity, diabetes, and Crohn's disease [199-202]. Using metaRNA-seq to study gut microbial gene expression in healthy and sick patients will help determine which specific microbial processes are actively contributing to these ailments. A similar approach to metaRNA-seq has been used to study microbial gene expression patterns in ocean sediment, revealing microbial contributions to biogeochemical cycles [227]. This tool could also be used to study the impacts of pollution on different ecosystems. For instance, differences have been observed in microbiota populating healthy and diseased coral [228], and metaRNA-seq analyses of these microbial communities may lend insight into reef bleaching. Thoughtful experimental design will allow researchers to use metaRNA-seq to address many open questions in the fields of microbial ecology and pathogenesis.

References

1. Darveau, R.P., *Periodontitis: a polymicrobial disruption of host homeostasis*. Nat Rev Microbiol, 2010. **8**(7): p. 481-90.
2. Vos, T., et al., *Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010*. Lancet, 2012. **380**(9859): p. 2163-96.
3. Kolenbrander, P.E., et al., *Oral multispecies biofilm development and the key role of cell-cell distance*. Nat Rev Microbiol, 2010. **8**(7): p. 471-80.
4. Kroes, I., P.W. Lepp, and D.A. Relman, *Bacterial diversity within the human subgingival crevice*. Proc Natl Acad Sci U S A, 1999. **96**(25): p. 14547-52.
5. Bik, E.M., et al., *Bacterial diversity in the oral cavity of 10 healthy individuals*. ISME J, 2010. **4**(8): p. 962-74.
6. Griffen, A.L., et al., *Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing*. ISME J, 2012. **6**(6): p. 1176-85.
7. Loesche, W.J., et al., *Relationship between oxygen tension and subgingival bacterial flora in untreated human periodontal pockets*. Infect Immun, 1983. **42**(2): p. 659-67.
8. Biswas, S., D.F. Duperon, and F.S. Chebib, *Study of crevice fluid in relation to periodontal disease in children. II. Effect of age, sex and gingival inflammation on crevice fluid protein, carbohydrate, total calcium, phosphate and nitrogen*. J Periodontal Res, 1977. **12**(4): p. 265-78.
9. Attstrom, R., *Presence of leukocytes in crevices of healthy and chronically inflamed gingivae*. J Periodontal Res, 1970. **5**(1): p. 42-7.
10. Shillitoe, E.J. and T. Lehner, *Immunoglobulins and complement in crevicular fluid, serum and saliva in man*. Arch Oral Biol, 1972. **17**(2): p. 241-7.
11. Thurlow, L.R., et al., *Staphylococcus aureus biofilms prevent macrophage phagocytosis and attenuate inflammation in vivo*. J Immunol, 2011. **186**(11): p. 6585-96.
12. Takahashi, N., et al., *Susceptibility of Actinobacillus actinomycetemcomitans to six antibiotics decreases as biofilm matures*. J Antimicrob Chemother, 2007. **59**(1): p. 59-65.
13. Tolker-Nielsen, T. and S. Molin, *Spatial Organization of Microbial Biofilm Communities*. Microb Ecol, 2000. **40**(2): p. 75-84.
14. Brown, S.A. and M. Whiteley, *A novel exclusion mechanism for carbon resource partitioning in Aggregatibacter actinomycetemcomitans*. J Bacteriol, 2007. **189**(17): p. 6407-14.
15. Grenier, D. and D. Mayrand, *Nutritional relationships between oral bacteria*. Infect Immun, 1986. **53**(3): p. 616-20.

16. Ramsey, M.M. and M. Whiteley, *Polymicrobial interactions stimulate resistance to host innate immunity through metabolite perception*. Proc Natl Acad Sci U S A, 2009. **106**(5): p. 1578-83.
17. Petersen, P.E., et al., *The global burden of oral diseases and risks to oral health*. Bull World Health Organ, 2005. **83**(9): p. 661-9.
18. Pihlstrom, B.L., B.S. Michalowicz, and N.W. Johnson, *Periodontal diseases*. Lancet, 2005. **366**(9499): p. 1809-20.
19. Socransky, S.S., et al., *Microbial complexes in subgingival plaque*. J Clin Periodontol, 1998. **25**(2): p. 134-44.
20. Henderson, B., et al., *Molecular pathogenicity of the oral opportunistic pathogen Actinobacillus actinomycetemcomitans*. Annu Rev Microbiol, 2003. **57**: p. 29-55.
21. Bostanci, N. and G.N. Belibasakis, *Porphyromonas gingivalis: an invasive and evasive opportunistic oral pathogen*. FEMS Microbiol Lett, 2012. **333**(1): p. 1-9.
22. Ramsey, M.M., K.P. Rumbaugh, and M. Whiteley, *Metabolite cross-feeding enhances virulence in a model polymicrobial infection*. PLoS Pathog, 2011. **7**(3): p. e1002012.
23. Ebersole, J.L., et al., *Comparative virulence of periodontopathogens in a mouse abscess model*. Oral Dis, 1995. **1**(3): p. 115-28.
24. Kesavalu, L., S.C. Holt, and J.L. Ebersole, *Virulence of a polymicrobic complex, Treponema denticola and Porphyromonas gingivalis, in a murine model*. Oral Microbiol Immunol, 1998. **13**(6): p. 373-7.
25. Armbruster, C.E., et al., *Indirect pathogenicity of Haemophilus influenzae and Moraxella catarrhalis in polymicrobial otitis media occurs via interspecies quorum signaling*. MBio, 2010. **1**(3).
26. Helmann, J.D. and M.J. Chamberlin, *Structure and function of bacterial sigma factors*. Annu Rev Biochem, 1988. **57**(1): p. 839-872.
27. Zheng, M., F. Åslund, and G. Storz, *Activation of the OxyR transcription factor by reversible disulfide bond formation*. Science, 1998. **279**(5357): p. 1718-1722.
28. Escolar, L., J. Pérez-Martín, and V. de Lorenzo, *Opening the iron box: transcriptional metalloregulation by the Fur protein*. J Bacteriol, 1999. **181**(20): p. 6223-6229.
29. Gottesman, S., *Micros for microbes: non-coding regulatory RNAs in bacteria*. Trends Genet, 2005. **21**(7): p. 399-404.
30. Møller, T., et al., *Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction*. Molecular cell, 2002. **9**(1): p. 23-30.
31. Winkler, W.C. and R.R. Breaker, *Regulation of bacterial gene expression by riboswitches*. Annu Rev Microbiol, 2005. **59**: p. 487-517.
32. Garneau, J.E., et al., *The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA*. Nature, 2010. **468**(7320): p. 67-71.
33. Godde, J.S. and A. Bickerton, *The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes*. J Mol Evol, 2006. **62**(6): p. 718-29.

34. Mojica, F.J., et al., *Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements*. J Mol Evol, 2005. **60**(2): p. 174-82.
35. Stern, A., et al., *Self-targeting by CRISPR: gene regulation or autoimmunity?* Trends in genetics : TIG, 2010. **26**(8): p. 335-40.
36. Henderson, B., et al., *Actinobacillus actinomycetemcomitans*. J Med Microbiol, 2002. **51**(12): p. 1013-20.
37. Rompikuntal, P.K., et al., *Perinuclear localization of internalized outer membrane vesicles carrying active cytolethal distending toxin from Aggregatibacter actinomycetemcomitans*. Infect Immun, 2012. **80**(1): p. 31-42.
38. Matangkasombut, O., et al., *Cytolethal distending toxin from Aggregatibacter actinomycetemcomitans induces DNA damage, S/G2 cell cycle arrest, and caspase- independent death in a Saccharomyces cerevisiae model*. Infect Immun, 2010. **78**(2): p. 783-92.
39. Fong, K.P., et al., *Aggregatibacter actinomycetemcomitans leukotoxin is post-translationally modified by addition of either saturated or hydroxylated fatty acyl chains*. Mol Oral Microbiol, 2011. **26**(4): p. 262-76.
40. Damek-Poprawa, M., et al., *Localization of Aggregatibacter actinomycetemcomitans cytolethal distending toxin subunits during intoxication of live cells*. Infect Immun, 2012. **80**(8): p. 2761-70.
41. Perez, B., et al., *Genetic analysis of the requirement for flp-2, tadV, and rcpB in Actinobacillus actinomycetemcomitans biofilm formation*. J Bacteriol, 2006. **188**(17): p. 6361-6375.
42. Planet, P.J., et al., *The widespread colonization island of Actinobacillus actinomycetemcomitans*. Nature genetics, 2003. **34**(2): p. 193-198.
43. Gregory, R., et al., *Immunoglobulin−degrading enzymes in localized juvenile periodontitis*. J Periodontal Res, 1992. **27**(3): p. 176-183.
44. Kittichotirat, W., et al., *Identification of the pangenome and its components in 14 distinct Aggregatibacter actinomycetemcomitans strains by comparative genomic analysis*. PLoS One, 2011. **6**(7): p. e22420.
45. Asikainen, S., et al., *Distribution of Actinobacillus actinomycetemcomitans serotypes in periodontal health and disease*. Oral Microbiol Immunol, 1991. **6**(2): p. 115-118.
46. Zambon, J., J. Slots, and R. Genco, *Serology of oral Actinobacillus actinomycetemcomitans and serotype distribution in human periodontal disease*. Infect Immun, 1983. **41**(1): p. 19-27.
47. Brogan, J.M., et al., *Regulation of Actinobacillus actinomycetemcomitans leukotoxin expression: analysis of the promoter regions of leukotoxic and minimally leukotoxic strains*. Infect Immun, 1994. **62**(2): p. 501-508.
48. Fujise, O., et al., *Clonal distribution of natural competence in Actinobacillus actinomycetemcomitans*. Oral Microbiol Immunol, 2004. **19**(5): p. 340-2.
49. Maughan, H. and R.J. Redfield, *Extensive variation in natural competence in Haemophilus influenzae*. Evolution, 2009. **63**(7): p. 1852-66.

50. Wang, Y., et al., *Type IV pilus gene homologs pilABCD are required for natural transformation in Actinobacillus actinomycetemcomitans*. Gene, 2003. **312**: p. 249-55.
51. Tarry, M., et al., *The extra-membranous domains of the competence protein HofQ show DNA binding, flexibility and a shared fold with type I KH domains*. J Mol Biol, 2011. **409**(4): p. 642-53.
52. Mullen, L.M., et al., *Pasteurellaceae ComE1 proteins combine the properties of fibronectin adhesins and DNA binding competence proteins*. PLoS One, 2008. **3**(12): p. e3991.
53. McCarthy, D., *Cloning of the rec-2 locus of Haemophilus influenzae*. Gene, 1989. **75**(1): p. 135-43.
54. Karudapuram, S., X. Zhao, and G.J. Barcak, *DNA sequence and characterization of Haemophilus influenzae dprA+, a gene required for chromosomal but not plasmid DNA transformation*. J Bacteriol, 1995. **177**(11): p. 3235-40.
55. Gwinn, M.L., et al., *A new transformation-deficient mutant of Haemophilus influenzae Rd with normal DNA uptake*. J Bacteriol, 1998. **180**(3): p. 746-8.
56. Bhattacharjee, M.K., D.H. Fine, and D.H. Figurski, *tfoX (sxy)-dependent transformation of Aggregatibacter (Actinobacillus) actinomycetemcomitans*. Gene, 2007. **399**(1): p. 53-64.
57. Chandler, M.S., *The gene encoding cAMP receptor protein is required for competence development in Haemophilus influenzae Rd*. Proc Natl Acad Sci U S A, 1992. **89**(5): p. 1626-30.
58. Maughan, H. and R.J. Redfield, *Tracing the evolution of competence in Haemophilus influenzae*. PLoS One, 2009. **4**(6): p. e5854.
59. Wylie, C.S., et al., *Optimal strategy for competence differentiation in bacteria*. PLoS Genet, 2010. **6**(9).
60. Redfield, R.J., *Do bacteria have sex?* Nat Rev Genet, 2001. **2**(8): p. 634-9.
61. Periasamy, S. and P.E. Kolenbrander, *Aggregatibacter actinomycetemcomitans builds mutualistic biofilm communities with Fusobacterium nucleatum and Veillonella species in saliva*. Infect Immun, 2009. **77**(9): p. 3542-51.
62. Suci, P. and M. Young, *Selective killing of Aggregatibacter actinomycetemcomitans by ciprofloxacin during development of a dual species biofilm with Streptococcus sanguinis*. Arch Oral Biol, 2011. **56**(10): p. 1055-1063.
63. Yuan, A., et al., *Actinobacillus actinomycetemcomitans pneumonia with chest wall involvement and rib destruction*. Chest, 1992. **101**(5): p. 1450-2.
64. Stepanovic, S., et al., *Brain abscess due to Actinobacillus actinomycetemcomitans*. APMIS, 2005. **113**(3): p. 225-8.
65. Paturel, L., et al., *Actinobacillus actinomycetemcomitans endocarditis*. Clin Microbiol Infect, 2004. **10**(2): p. 98-118.

66. Lamster, I.B. and M.J. Novak, *Host mediators in gingival crevicular fluid: implications for the pathogenesis of periodontal disease*. Crit Rev Oral Biol Med, 1992. **3**(1-2): p. 31-60.
67. Paster, B.J., et al., *The breadth of bacterial diversity in the human periodontal pocket and other oral sites*. Periodontol 2000, 2006. **42**: p. 80-7.
68. de Soet, J.J., B. Nyvad, and M. Kilian, *Strain-related acid production by oral streptococci*. Caries Res, 2000. **34**(6): p. 486-90.
69. Repoila, F., N. Majdalani, and S. Gottesman, *Small non-coding RNAs, co-ordinators of adaptation processes in Escherichia coli: the RpoS paradigm*. Mol Microbiol, 2003. **48**(4): p. 855-61.
70. Legewie, S., et al., *Small RNAs establish delays and temporal thresholds in gene expression*. Biophys J, 2008. **95**(7): p. 3232-8.
71. Grundy, F.J., S.C. Lehman, and T.M. Henkin, *The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes*. Proc Natl Acad Sci U S A, 2003. **100**(21): p. 12057-62.
72. Kwon, M. and S.A. Strobel, *Chemical basis of glycine riboswitch cooperativity*. RNA, 2008. **14**(1): p. 25-34.
73. Socransky, S.S., J.L. Dzink, and C.M. Smith, *Chemically defined medium for oral microorganisms*. J Clin Microbiol, 1985. **22**(2): p. 303-5.
74. Walters, M.C., 3rd, et al., *Contributions of antibiotic penetration, oxygen limitation, and low metabolic activity to tolerance of Pseudomonas aeruginosa biofilms to ciprofloxacin and tobramycin*. Antimicrob Agents Chemother, 2003. **47**(1): p. 317-23.
75. Ausubel, F.M., *Short protocols in molecular biology : a compendium of methods from Current protocols in molecular biology*. 5th ed2002, New York: Wiley. 2 v. (various pagings).
76. Baehni, P., et al., *Interaction of inflammatory cells and oral microorganisms. VIII. Detection of leukotoxic activity of a plaque-derived gram-negative microorganism*. Infect Immun, 1979. **24**(1): p. 233-43.
77. Mintz, K.P. and P.M. Fives-Taylor, *impA, a gene coding for an inner membrane protein, influences colonial morphology of Actinobacillus actinomycetemcomitans*. Infect Immun, 2000. **68**(12): p. 6580-6.
78. Hayes, W., *Genetic recombination in Bact. coli K12; analysis of the stimulating effect of ultra-violet light*. Nature, 1952. **169**(4311): p. 1017-8.
79. Baba, T., et al., *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection*. Mol Syst Biol, 2006. **2**: p. 2006 0008.
80. Bakaletz, L.O., et al., *Frequency of fimbriation of nontypable Haemophilus influenzae and its ability to adhere to chinchilla and human respiratory epithelium*. Infect Immun, 1988. **56**(2): p. 331-5.
81. Lloyd, A.L., B.J. Marshall, and B.J. Mee, *Identifying cloned Helicobacter pylori promoters by primer extension using a FAM-labelled primer and GeneScan analysis*. J Microbiol Methods, 2005. **60**(3): p. 291-8.

82. Palmer, G.C., et al., *Characterization of the Pseudomonas aeruginosa transcriptional response to phenylalanine and tyrosine*. J Bacteriol. **192**(11): p. 2722-8.
83. Danner, D.B., et al., *An eleven-base-pair sequence determines the specificity of DNA uptake in Haemophilus transformation*. Gene, 1980. **11**(3-4): p. 311-8.
84. Smith, H.O., D.B. Danner, and R.A. Deich, *Genetic transformation*. Annu Rev Biochem, 1981. **50**: p. 41-68.
85. Kovach, M.E., et al., *Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes*. Gene, 1995. **166**(1): p. 175-6.
86. Ho, S.N., et al., *Site-directed mutagenesis by overlap extension using the polymerase chain reaction*. Gene, 1989. **77**(1): p. 51-9.
87. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction*. Nucleic Acids Res, 2003. **31**(13): p. 3406-15.
88. Garst, A.D., et al., *Crystal structure of the lysine riboswitch regulatory mRNA element*. J Biol Chem, 2008. **283**(33): p. 22347-51.
89. Serganov, A., L. Huang, and D.J. Patel, *Structural insights into amino acid binding and gene control by a lysine riboswitch*. Nature, 2008. **455**(7217): p. 1263-7.
90. 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.
91. Thompson, J.D., T.J. Gibson, and D.G. Higgins, *Multiple sequence alignment using ClustalW and ClustalX*. Curr Protoc Bioinformatics, 2002. **Chapter 2**: p. Unit 2 3.
92. Zhai, Y., J. Tchieu, and M.H. Saier, Jr., *A web-based Tree View (TV) program for the visualization of phylogenetic trees*. J Mol Microbiol Biotechnol, 2002. **4**(1): p. 69-70.
93. Chang, T.H., et al., *Computational identification of riboswitches based on RNA conserved functional sequences and conformations*. RNA, 2009. **15**(7): p. 1426-30.
94. Spurio, R., et al., *The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending*. EMBO J, 1997. **16**(7): p. 1795-805.
95. Blouin, S. and D.A. Lafontaine, *A loop loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control*. RNA, 2007. **13**(8): p. 1256-67.
96. Phan, T.T. and W. Schumann, *Transcriptional analysis of the lysine-responsive and riboswitch-regulated lysC gene of Bacillus subtilis*. Curr Microbiol, 2009. **59**(4): p. 463-8.
97. Blount, K.F., et al., *Antibacterial lysine analogs that target lysine riboswitches*. Nat Chem Biol, 2007. **3**(1): p. 44-9.

98. Sudarsan, N., et al., *An mRNA structure in bacteria that controls gene expression by binding lysine*. Genes Dev, 2003. **17**(21): p. 2688-97.
99. Bernstein, J.A., et al., *Global analysis of Escherichia coli RNA degradosome function using DNA microarrays*. Proc Natl Acad Sci U S A, 2004. **101**(9): p. 2758-63.
100. Rodionov, D.A., et al., *Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch?* Nucleic Acids Res, 2003. **31**(23): p. 6748-57.
101. Ellis, J., et al., *Topological analysis of the lysine-specific permease of Escherichia coli*. Microbiology, 1995. **141 (Pt 8)**: p. 1927-35.
102. Steffes, C., et al., *The lysP gene encodes the lysine-specific permease*. J Bacteriol, 1992. **174**(10): p. 3242-9.
103. Kang, C.H., et al., *Crystal structure of the lysine-, arginine-, ornithine-binding protein (LAO) from Salmonella typhimurium at 2.7-A resolution*. J Biol Chem, 1991. **266**(35): p. 23893-9.
104. Seep-Feldhaus, A.H., J. Kalinowski, and A. Puhler, *Molecular analysis of the Corynebacterium glutamicum lysI gene involved in lysine uptake*. Mol Microbiol, 1991. **5**(12): p. 2995-3005.
105. Rosen, B.P., *Basic amino acid transport in Escherichia coli*. J Biol Chem, 1971. **246**(11): p. 3653-62.
106. Kelley, L.A. and M.J. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server*. Nat Protoc, 2009. **4**(3): p. 363-71.
107. Ito, M., et al., *Role of the nhaC-encoded Na+/H+ antiporter of alkaliphilic Bacillus firmus OF4*. J Bacteriol, 1997. **179**(12): p. 3851-7.
108. Mesbah, N.M., G.M. Cook, and J. Wiegel, *The halophilic alkalithermophile Natranaerobius thermophilus adapts to multiple environmental extremes using a large repertoire of Na(K)/H antiporters*. Mol Microbiol, 2009. **74**(2): p. 270-81.
109. Wei, Y., et al., *Bacillus subtilis YqkI is a novel malic/Na+-lactate antiporter that enhances growth on malate at low protonmotive force*. J Biol Chem, 2000. **275**(39): p. 30287-92.
110. Altschul, S.F., et al., *Protein database searches using compositionally adjusted substitution matrices*. FEBS J, 2005. **272**(20): p. 5101-9.
111. Foxwell, A.R., J.M. Kyd, and A.W. Cripps, *Nontypeable Haemophilus influenzae: pathogenesis and prevention*. Microbiol Mol Biol Rev, 1998. **62**(2): p. 294-308.
112. Marchler-Bauer, A., et al., *CDD: specific functional annotation with the Conserved Domain Database*. Nucleic Acids Res, 2009. **37**(Database issue): p. D205-10.
113. MacLeod, P.R. and R.A. MacLeod, *Identification and sequence of a Na(+)-linked gene from the marine bacterium Alteromonas haloplanktis which functionally complements the dagA gene of Escherichia coli*. Mol Microbiol, 1992. **6**(18): p. 2673-81.

114. Pulvermacher, S.C., L.T. Stauffer, and G.V. Stauffer, *Role of the sRNA GcvB in regulation of cycA in Escherichia coli*. Microbiology, 2009. **155**(Pt 1): p. 106-14.
115. Pulvermacher, S.C., L.T. Stauffer, and G.V. Stauffer, *The small RNA GcvB regulates sstT mRNA expression in Escherichia coli*. J Bacteriol, 2009. **191**(1): p. 238-48.
116. Urbanowski, M.L., L.T. Stauffer, and G.V. Stauffer, *The gcvB gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in Escherichia coli*. Mol Microbiol, 2000. **37**(4): p. 856-68.
117. Sharma, C.M., et al., *A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites*. Genes Dev, 2007. **21**(21): p. 2804-17.
118. Abe, N., et al., *Roles of Arg- and Lys-gingipains in coaggregation of Porphyromonas gingivalis: identification of its responsible molecules in translation products of rgpA, kgp, and hagA genes*. Biol Chem, 2004. **385**(11): p. 1041-7.
119. Gregory, R.L., et al., *Immunoglobulin-degrading enzymes in localized juvenile periodontitis*. J Periodontal Res, 1992. **27**(3): p. 176-83.
120. Loh, E., et al., *A trans-acting riboswitch controls expression of the virulence regulator PrfA in Listeria monocytogenes*. Cell, 2009. **139**(4): p. 770-9.
121. Redfield, R.J., et al., *Evolution of competence and DNA uptake specificity in the Pasteurellaceae*. BMC Evol Biol, 2006. **6**: p. 82.
122. Levin, B.R. and O.E. Cornejo, *The population and evolutionary dynamics of homologous gene recombination in bacterial populations*. PLoS Genet, 2009. **5**(8): p. e1000601.
123. Wu, M. and J.A. Eisen, *A simple, fast, and accurate method of phylogenomic inference*. Genome Biol, 2008. **9**(10): p. R151.
124. Chen, C., et al., *Genome Sequence of a Serotype b Non-JP2 Aggregatibacter actinomycetemcomitans Strain, ANH9381, from a Periodontally Healthy Individual*. J Bacteriol, 2012. **194**(7): p. 1837.
125. Chen, C., et al., *Genome sequence of naturally competent Aggregatibacter actinomycetemcomitans serotype a strain D7S-1*. Journal of bacteriology, 2010. **192**(10): p. 2643-4.
126. Chen, C., et al., *Genome sequence of Aggregatibacter actinomycetemcomitans serotype c strain D11S-1*. J Bacteriol, 2009. **191**(23): p. 7378-9.
127. Wang, Y., et al., *Natural transformation and DNA uptake signal sequences in Actinobacillus actinomycetemcomitans*. J Bacteriol, 2002. **184**(13): p. 3442-9.
128. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput*. Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
129. J, F., *Confidence limits on phylogenies: An approach using the bootstrap*. Evolution, 1985. **39**: p. 783-791.

130. Tamura, K. and M. Nei, *Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees*. Mol Biol Evol, 1993. **10**(3): p. 512-26.
131. Tamura, K., et al., *MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods*. Mol Biol Evol, 2011. **28**(10): p. 2731-9.
132. Thomas, C.M. and K.M. Nielsen, *Mechanisms of, and barriers to, horizontal gene transfer between bacteria*. Nat Rev Microbiol, 2005. **3**(9): p. 711-21.
133. Richter, M. and R. Rossello-Mora, *Shifting the genomic gold standard for the prokaryotic species definition*. Proc Natl Acad Sci U S A, 2009. **106**(45): p. 19126-31.
134. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes*. Science, 2008. **321**(5891): p. 960-4.
135. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes*. Science, 2007. **315**(5819): p. 1709-12.
136. Makarova, K.S., et al., *Evolution and classification of the CRISPR-Cas systems*. Nat Rev Microbiol, 2011. **9**(6): p. 467-77.
137. Wiedenheft, B., S.H. Sternberg, and J.A. Doudna, *RNA-guided genetic silencing systems in bacteria and archaea*. Nature, 2012. **482**(7385): p. 331-8.
138. Hale, C.R., et al., *RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex*. Cell, 2009. **139**(5): p. 945-56.
139. Marraffini, L.A. and E.J. Sontheimer, *CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA*. Science, 2008. **322**(5909): p. 1843-5.
140. Palmer, K.L. and M.S. Gilmore, *Multidrug-Resistant Enterococci Lack CRISPR-cas*. mBio, 2010. **1**(4).
141. Delaney, N.F., et al., *Ultrafast evolution and loss of CRISPRs following a host shift in a novel wildlife pathogen, Mycoplasma gallisepticum*. PLoS Genet, 2012. **8**(2): p. e1002511.
142. 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.
143. Grissa, I., G. Vergnaud, and C. Pourcel, *CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats*. Nucleic Acids Res, 2007. **35**(Web Server issue): p. W52-7.
144. Mandin, P., et al., *Identification of new noncoding RNAs in Listeria monocytogenes and prediction of mRNA targets*. Nucleic Acids Res, 2007. **35**(3): p. 962-74.
145. Carthew, R.W. and E.J. Sontheimer, *Origins and Mechanisms of miRNAs and siRNAs*. Cell, 2009. **136**(4): p. 642-55.
146. Izano, E.A., et al., *Poly-N-acetylglucosamine mediates biofilm formation and detergent resistance in Aggregatibacter actinomycetemcomitans*. Microb Pathog, 2008. **44**(1): p. 52-60.

147. Bikard, D., et al., *CRISPR Interference Can Prevent Natural Transformation and Virulence Acquisition during In†Vivo Bacterial Infection*. Cell Host Microbe, 2012. **12**(2): p. 177-186.
148. Avery, O.T., C.M. Macleod, and M. McCarty, *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii*. J Exp Med, 1944. **79**(2): p. 137-58.
149. Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors*. Nature, 2005. **437**(7057): p. 376-80.
150. Darling, A.E., B. Mau, and N.T. Perna, *progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement*. PloS one, 2010. **5**(6): p. e11147.
151. Grissa, I., G. Vergnaud, and C. Pourcel, *CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats*. Nucleic Acids Res, 2007. **35**(Web Server issue): p. W52-7.
152. Zhang, Z., et al., *A greedy algorithm for aligning DNA sequences*. J Comp Biol, 2000. **7**(1-2): p. 203-14.
153. Jorth, P. and M. Whiteley, *Characterization of a novel riboswitch-regulated lysine transporter in Aggregatibacter actinomycetemcomitans*. J Bacteriol, 2010. **192**(23): p. 6240-50.
154. Pasteur, L., *La Theorie des Germes*. Comptes Rendus l'Academie des Sciences, 1878. **86**: p. 1037-1043.
155. Xu, Q., M. Dziejman, and J.J. Mekalanos, *Determination of the transcriptome of Vibrio cholerae during intraintestinal growth and midexponential phase in vitro*. Proc Natl Acad Sci U S A, 2003. **100**(3): p. 1286-91.
156. Snyder, J.A., et al., *Transcriptome of uropathogenic Escherichia coli during urinary tract infection*. Infect Immun, 2004. **72**(11): p. 6373-81.
157. Orihuela, C.J., et al., *Microarray analysis of pneumococcal gene expression during invasive disease*. Infect Immun, 2004. **72**(10): p. 5582-96.
158. Mashburn, L.M., et al., *Staphylococcus aureus serves as an iron source for Pseudomonas aeruginosa during in vivo coculture*. J Bacteriol, 2005. **187**(2): p. 554-66.
159. Westermann, A.J., S.A. Gorski, and J. Vogel, *Dual RNA-seq of pathogen and host*. Nat Rev Microbiol, 2012. **10**(9): p. 618-30.
160. Mandlik, A., et al., *RNA-Seq-based monitoring of infection-linked changes in Vibrio cholerae gene expression*. Cell Host Microbe, 2011. **10**(2): p. 165-74.
161. Palmer, K.L., et al., *Cystic fibrosis sputum supports growth and cues key aspects of Pseudomonas aeruginosa physiology*. J Bacteriol, 2005. **187**(15): p. 5267-77.
162. Palmer, K.L., S.A. Brown, and M. Whiteley, *Membrane-bound nitrate reductase is required for anaerobic growth in cystic fibrosis sputum*. J Bacteriol, 2007. **189**(12): p. 4449-55.

163. Palmer, K.L., L.M. Aye, and M. Whiteley, *Nutritional cues control Pseudomonas aeruginosa multicellular behavior in cystic fibrosis sputum*. J Bacteriol, 2007. **189**(22): p. 8079-87.
164. Palmer, G.C., et al., *Characterization of the Pseudomonas aeruginosa transcriptional response to phenylalanine and tyrosine*. J Bacteriol, 2010. **192**(11): p. 2722-8.
165. Brown, S.A. and M. Whiteley, *Characterization of the L-lactate dehydrogenase from Aggregatibacter actinomycetemcomitans*. PLoS One, 2009. **4**(11): p. e7864.
166. Zambon, J.J., *Actinobacillus actinomycetemcomitans in human periodontal disease*. J Clin Periodontol, 1985. **12**(1): p. 1-20.
167. Slots, J., H.S. Reynolds, and R.J. Genco, *Actinobacillus actinomycetemcomitans in human periodontal disease: a cross-sectional microbiological investigation*. Infect Immun, 1980. **29**(3): p. 1013-20.
168. Kozarov, E.V., et al., *Human atherosclerotic plaque contains viable invasive Actinobacillus actinomycetemcomitans and Porphyromonas gingivalis*. Arterioscler Thromb Vasc Biol, 2005. **25**(3): p. e17-8.
169. David, M., et al., *SHRiMP2: sensitive yet practical SHort Read Mapping*. Bioinformatics, 2011. **27**(7): p. 1011-2.
170. Dodt, M., et al., *FLEXBAR—Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms*. Biology, 2012. **1**(3): p. 895-905.
171. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2*. Nat Methods, 2012. **9**(4): p. 357-9.
172. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.
173. Robinson, J.T., et al., *Integrative genomics viewer*. Nat Biotechnol, 2011. **29**(1): p. 24-6.
174. Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov, *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration*. Brief Bioinform, 2013. **14**(2): p. 178-92.
175. Anders, S. and W. Huber, *Differential expression analysis for sequence count data*. Genome Biol, 2010. **11**(10): p. R106.
176. Krzywinski, M., et al., *Circos: an information aesthetic for comparative genomics*. Genome Res, 2009. **19**(9): p. 1639-45.
177. Mintz, K.P., C. Brissette, and P.M. Fives-Taylor, *A recombinase A-deficient strain of Actinobacillus actinomycetemcomitans constructed by insertional mutagenesis using a mobilizable plasmid*. FEMS Microbiol Lett, 2002. **206**(1): p. 87-92.
178. Evguenieva-Hackenberg, E. and G. Klug, *New aspects of RNA processing in prokaryotes*. Curr Opin Microbiol, 2011. **14**(5): p. 587-92.
179. Mitschke, J., et al., *Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in Anabaena sp. PCC7120*. Proc Natl Acad Sci U S A, 2011. **108**(50): p. 20130-5.

180. Sharma, C.M., et al., *The primary transcriptome of the major human pathogen Helicobacter pylori*. Nature, 2010. **464**(7286): p. 250-5.
181. Gardner, P.P., et al., *Rfam: Wikipedia, clans and the "decimal" release*. Nucleic Acids Res, 2011. **39**(Database issue): p. D141-5.
182. Griffiths-Jones, S., et al., *Rfam: an RNA family database*. Nucleic Acids Res, 2003. **31**(1): p. 439-41.
183. Guerrier-Takada, C., et al., *The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme*. Cell, 1983. **35**(3 Pt 2): p. 849-57.
184. Jorth, P. and M. Whiteley, *Characterization of a novel riboswitch-regulated lysine transporter in Aggregatibacter actinomycetemcomitans*. J Bacteriol, 2010. **192**(23): p. 6240-50.
185. Jorth, P. and M. Whiteley, *An evolutionary link between natural transformation and CRISPR adaptive immunity*. MBio, 2012. **3**(5): p. e00309-12.
186. Keiler, K.C., *Biology of trans-translation*. Annu Rev Microbiol, 2008. **62**: p. 133-51.
187. Ray, B.K. and D. Apirion, *Characterization of 10S RNA: a new stable rna molecule from Escherichia coli*. Mol Gen Genet, 1979. **174**(1): p. 25-32.
188. Ulbrandt, N.D., J.A. Newitt, and H.D. Bernstein, *The E. coli signal recognition particle is required for the insertion of a subset of inner membrane proteins*. Cell, 1997. **88**(2): p. 187-96.
189. Wassarman, K.M. and G. Storz, *6S RNA regulates E. coli RNA polymerase activity*. Cell, 2000. **101**(6): p. 613-23.
190. Mann, B., et al., *Control of virulence by small RNAs in Streptococcus pneumoniae*. PLoS Pathog, 2012. **8**(7): p. e1002788.
191. Meibom, K.L., et al., *Hfq, a novel pleiotropic regulator of virulence-associated genes in Francisella tularensis*. Infect Immun, 2009. **77**(5): p. 1866-80.
192. Sampson, T.R., et al., *A CRISPR/Cas system mediates bacterial innate immune evasion and virulence*. Nature, 2013.
193. Childress, C., et al., *Mlc Is a Transcriptional Activator with a Key Role in Integrating Cyclic AMP Receptor Protein and Integration Host Factor Regulation of Leukotoxin RNA Synthesis in Aggregatibacter actinomycetemcomitans*. J Bacteriol, 2013. **195**(10): p. 2284-97.
194. Tatusov, R.L., et al., *The COG database: an updated version includes eukaryotes*. BMC Bioinformatics, 2003. **4**: p. 41.
195. Tatusov, R.L., E.V. Koonin, and D.J. Lipman, *A genomic perspective on protein families*. Science, 1997. **278**(5338): p. 631-7.
196. Fender, A., et al., *RNAs actively cycle on the Sm-like protein Hfq*. Genes Dev, 2010. **24**(23): p. 2621-6.
197. Link, T.M., P. Valentin-Hansen, and R.G. Brennan, *Structure of Escherichia coli Hfq bound to polyriboadenylate RNA*. Proc Natl Acad Sci U S A, 2009. **106**(46): p. 19292-7.

198. Koo, J.T., et al., *Global discovery of small RNAs in Yersinia pseudotuberculosis identifies Yersinia-specific small, noncoding RNAs required for virulence*. Proc Natl Acad Sci U S A, 2011. **108**(37): p. E709-17.
199. Karlsson, F.H., et al., *Gut metagenome in European women with normal, impaired and diabetic glucose control*. Nature, 2013. **498**(7452): p. 99-103.
200. Ley, R.E., et al., *Microbial ecology: human gut microbes associated with obesity*. Nature, 2006. **444**(7122): p. 1022-3.
201. Manichanh, C., et al., *Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach*. Gut, 2006. **55**(2): p. 205-11.
202. Markle, J.G., et al., *Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity*. Science, 2013. **339**(6123): p. 1084-8.
203. Caporaso, J.G., et al., *Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms*. ISME J, 2012. **6**(8): p. 1621-4.
204. Claesson, M.J., et al., *Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions*. Nucleic Acids Res, 2010. **38**(22): p. e200.
205. Das, T., et al., *Role of extracellular DNA in initial bacterial adhesion and surface aggregation*. Appl Environ Microbiol, 2010. **76**(10): p. 3405-8.
206. Flemming, H.C. and J. Wingender, *The biofilm matrix*. Nat Rev Microbiol, 2010. **8**(9): p. 623-33.
207. Miura, A., et al., *Growth-rate-dependent regulation of ribosome synthesis in E. coli: expression of the lacZ and galK genes fused to ribosomal promoters*. Cell, 1981. **25**(3): p. 773-82.
208. DeSantis, T.Z., et al., *Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB*. Appl Environ Microbiol, 2006. **72**(7): p. 5069-72.
209. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data*. Nat Methods, 2010. **7**(5): p. 335-6.
210. Chen, T., et al., *The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information*. Database (Oxford), 2010. **2010**: p. baq013.
211. Kastenmuller, G., et al., *Uncovering metabolic pathways relevant to phenotypic traits of microbial genomes*. Genome Biol, 2009. **10**(3): p. R28.
212. Langille, M.G., et al., *Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences*. Nat Biotechnol, 2013. **31**(9): p. 814-21.
213. Ogata, H., et al., *KEGG: Kyoto Encyclopedia of Genes and Genomes*. Nucleic Acids Res, 1999. **27**(1): p. 29-34.
214. Hajishengallis, G., et al., *Low-abundance biofilm species orchestrates inflammatory periodontal disease through the commensal microbiota and complement*. Cell Host Microbe, 2011. **10**(5): p. 497-506.

215. Potempa, M., et al., *Interpain A, a cysteine proteinase from Prevotella intermedia, inhibits complement by degrading complement factor C3*. PLoS Pathog, 2009. **5**(2): p. e1000316.
216. Saito, T., et al., *Cloning, expression, and sequencing of a protease gene from Bacteroides forsythus ATCC 43037 in Escherichia coli*. Infect Immun, 1997. **65**(11): p. 4888-91.
217. Niederman, R., et al., *Short-chain carboxylic acid concentration in human gingival crevicular fluid*. J Dent Res, 1997. **76**(1): p. 575-9.
218. Chang, M.C., et al., *Butyrate induces reactive oxygen species production and affects cell cycle progression in human gingival fibroblasts*. J Periodontal Res, 2013. **48**(1): p. 66-73.
219. Aronesty, E., *Comparisons of Sequencing Utility Programs*. Open Bioinformatics J, 2013. **7**: p. 1-8.
220. Lozupone, C. and R. Knight, *UniFrac: a new phylogenetic method for comparing microbial communities*. Appl Environ Microbiol, 2005. **71**(12): p. 8228-35.
221. Jorth, P., et al., *Probing bacterial metabolism during infection using high-resolution transcriptomics*. J Bacteriol, 2013.
222. Robinson, M.D. and G.K. Smyth, *Moderated statistical tests for assessing differences in tag abundance*. Bioinformatics, 2007. **23**(21): p. 2881-7.
223. Zambon, J.J., J. Slots, and R.J. Genco, *Serology of oral Actinobacillus actinomycetemcomitans and serotype distribution in human periodontal disease*. Infect Immun, 1983. **41**(1): p. 19-27.
224. van Opijnen, T., K.L. Bodi, and A. Camilli, *Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms*. Nat Methods, 2009. **6**(10): p. 767-72.
225. Frias-Lopez, J. and A. Duran-Pinedo, *Effect of periodontal pathogens on the metatranscriptome of a healthy multispecies biofilm model*. J Bacteriol, 2012. **194**(8): p. 2082-95.
226. Barker, H., J.M. Kahn, and L. Hedrick, *Pathway of lysine degradation in Fusobacterium nucleatum*. J Bacteriol, 1982. **152**(1): p. 201-207.
227. Orsi, W.D., et al., *Gene expression in the deep biosphere*. Nature, 2013. **499**(7457): p. 205-8.
228. Krediet, C.J., et al., *Coral-associated micro-organisms and their roles in promoting coral health and thwarting diseases*. Proc Royal Soc B, 2013. **280**(1755).