UNIVERSITY OF CALIFORNIA
RIVERSIDE

Identification and Characterization of Mobile Genetic Elements in Gut Bacteroidota

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Microbiology

by

Jericho Resabal Ortañez

June 2024

Dissertation Committee:
Dr. Patrick H. Degnan, Chairperson
Dr. Ansel Hsiao
Dr. Jason Stajich

The Dissertation of Jericho Resabal Ortañez is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgments

Thank you, Dr. Patrick Degnan. We have come a long way since our days at Webber Hall. I deeply appreciate your patience and clarity in explaining complex concepts, experimental design, and direction in writing manuscripts. Your mentorship was invaluable as I delved into the world of horizontal gene transfer, shaping me into the scientist I am today.

To my dissertation committee members - Thank you, Dr. Jason Stajich. Your belief in me and the opportunities you have provided have been fundamental in my success. Joining your lab as an undergraduate was the catalyst for my pursuit of a doctorate degree. Dr. Ansel Hsiao, your feedback, suggestions, and unwavering support throughout my program have been indispensable. Also, I am eternally grateful to you and Patrick for treating us out to hotpot.

To my fellow lab mates in the Degnan lab – Andrew Crago, Disha Dutta, Samson Avena, and William Hofstedt. Thank you for all the great memories and panda updates. Let's plan a trip to the Getaway Café soon. To my former housemates – Dr. Biagio DiSalvo, thanks for taking the risk of deciding to be my housemate via email. Also, what color is a carrot? To Dr. Jessica Trinh, thanks for being the glue that brought our cohort together. To Benjamin Hoyt, I appreciate our late-night venting sessions and I forgive you for spoiling Naruto. To my friends/colleagues at UCR – Elyza Do, thanks for the emotional support and the cat-sitting service. To Varadh Piamthai, Dr. John Macbeth and

Dr. Jennifer Cho, thank you for all the advice and interesting conversations. To Talyssa Topacio, thanks for making me laugh – you have the funniest Instagram stories.

To Dr. Derreck Carter-House, your informal mentorship was a turning point in my decision to pursue a doctorate, despite your cautionary advice against grad school. To Dr. Jesus Peña, I appreciate all the advice you have given me up to this point. I have always admired your beautiful PowerPoints and presentation skills – you were made to be an educator. To Julia Adams, thank you for giving me the opportunity to work with you as a co-author. I also appreciate the time and effort you spent to help me apply to graduate programs and the continual support throughout my program.

To my parents, Rogelio Ortañez and Dr. Maria Asuncion Ortañez, thank you for having the courage to embark on our journey to the United States and fostering my early interest in science. Ma, thanks for helping me win the county science fair in 6th grade through my "Which detergent is the most economical to use?" project. To my brothers/cousins in the United States– Gian-Carlo Ortañez, Dr. Ryan-Carlo Ortañez, Pierre-Carlo Ortañez, Melissa Nano, Melissa Nano: thank you for nurturing my interest in science by watching Crocodile Hunter and "The Most Extreme" on Animal Planet with me.

To my childhood friends from Moreno Valley, the "Day Ones," your enduring support and friendship are cherished. Spending time with you all gave me the mental break I needed from work.

Thank you to my 9th grade biology teacher, Mrs. Barbara Sataki. Enrolling in your class ignited a lifelong fascination with science for me.

To my beloved cats, Meelo and Bumi, your presence brings solace in difficult times. If I had the power to give away some years of my life to you both, I would do it in a heartbeat. Thank you to all the other pets I had the pleasure of spending time with: Binx DiSalvo, Mochi, Cholla, Ziggy, Scout, Saetbyeol, Toby, Kairo, Sadie, and Riyu.

To my wife, Jezreel Janelle Ortañez. Your unwavering support has been my guiding light. As always, I'm eternally grateful to live in the same timeline as you.

Dedication

To my wife, Jezreel. And to our future children (maybe).

ABSTRACT OF THE DISSERTATION

Identification and Characterization of Mobile Genetic Elements in Gut Bacteroidota

by

Jericho Resabal Ortañez

Doctor of Philosophy, Graduate Program in Microbiology
University of California, Riverside, June 2024
Dr. Patrick H. Degnan, Chairperson

Mobile genetic elements (MGEs) are important drivers of bacterial evolution, facilitating the exchange of fitness determinant genes such as antibiotic resistance and virulence factors. Although various computational methods exist for identifying potential MGEs, confirming their ability to transfer requires additional experimental approaches. Here, we introduce the transposon (Tn) mutagenesis mobilization method (TMMM) as a means to confirm mobilization without the need for targeted mutations. Using this method, we identified two MGEs, one being a novel CTn labeled *Pv*CTn, present in *Phocaeicola vulgatus*. Through Tn mutagenesis and subsequent gene deletion, we discovered that a helix-turn-helix motif gene, BVU3433, negatively regulated the conjugation efficiency of *Pv*CTn *in vitro*. Furthermore, our transcriptomics data revealed that BVU3433 plays a crucial role in the repression concouof *Pv*CTn genes, including genes involved in forming complete conjugation machinery (Type IV Secretion System (T4SS)). Finally, analysis of individual strain genomes and community metagenomes identified the widespread prevalence of *Pv*CTn-like elements with putative BVU3433 homologs among human gut-associated bacteria. Tn mutagenesis detection of MGEs

enables *in vitro* observation of transfer events. Further this method has potential for *in vivo* identification of additional functional MGEs and characterization of the fitness determinants they encode. For example, there is an increasing threat of antibiotic resistance in clinically relevant bacteria, often exacerbated by MGEs. We specifically investigated CTnDOT, a key contributor to tetracycline resistance in 80% of clinical and community isolates of Bacteroides. In this study, we observed the *in vitro* mobilization of a novel CTnDOT-like element (PmDOT), from *Parabacteroides merdae* into *Bacteroides thetaiotaomicron*. Subsequent genome screens revealed the presence of intact CTnDOT-like elements in ~50% of the 134 human-gut associated Bacteroidota genomes examined. Moreover, through the analysis of human fecal metagenomes, we determined the widespread prevalence of CTnDOT-like elements across diverse global human populations. Our findings confirm CTnDOT-like elements are a diverse and significant group of MGEs within the human gut microbiome.

Table of Contents

List of Figures

Chapter 2

Chapter 3

List of Tables

Chapter 2

Chapter 3

CHAPTER 1: General Introduction

Human Gut Microbiome

The human gastrointestinal tract is host to a diverse microbial community comprised primarily of Bacteria, but also including fungi, Archaea, and viruses. While the gut microbiome has been associated with diseases like inflammatory bowel disease, irritable bowel syndrome, obesity, and diabetes, it also provides numerous positive benefits to human health, including the metabolization of non-digestible components of foods, safeguarding against pathogens, and modulation of the immune system (Bull & Plummer 2014). Therefore, efforts to describe microbiome compositional differences among patient groups ultimately lead to an ability to understand mechanisms contributing to these outcomes and lead to translational medical interventions.

Studies using data from the METAgenomics of the Human Intestinal Tract (MetaHIT) and the Human Microbiome project identified 2172 species, where 93.5% of those species are classified into four bacterial phyla: Bacillota, Pseudomonadota, Actinomycetota, and Bacteroidota (formerly named Firmicutes, Proteobacteria, Actinobacteria, and Bacteroidetes respectively)(Hugon et al. 2015). While healthy gut microbiota composition varies individually, specific compositions are often linked to diseases, like obesity in pediatric patients is associated with elevated proportions of species from the Bacillota and depleted proportions of Bacteroidota species (Riva et al. 2017).

Bacteroidota are gram-negative microbes with numerous anaerobic members that are adapted to living in the gastrointestinal tracts of mammals. Bacteroidota excel in polysaccharide fermentation, which supplement the host with extra energy in the form of short chain fatty acids (Wexler 2007). Additionally, the presence of Bacteroidota in the colon acts as a preventative measure against infections by opportunistic pathogens like *C. difficile* (Deng et al. 2018). Individual microbes like *Bacteroides fragilis* have been shown to play critical roles in the transformation of conjugated to deconjugated bile salts through the production of bile salt hydrolases, leading to further metabolism of secondary bile acids (Wexler 2007). Further, there is a direct link between the protection from recurrent *Clostridioides difficile* infection and increased bile salt hydrolase activity (Mullish et al. 2019). However, in some instances, members of the Bacteroidota may exhibit negative effects on hosts. For instance, *B. fragilis* has been associated with abdominal infections, leading to an increased risk colon cancer from these infections, while *B. thetaiotaomicron* enhances disease progression of enterohemorrhagic *Escherichia coli* (Brook 1989, Curtis et al. 2014). Nevertheless, the advantages of the vast majority of the commensal Bacteroidota outweigh the drawbacks. In contrast, the host immune system can modulate the species diversity of gut microorganisms (Guzman-Bautista et al. 2020). For example, immunocompromised hosts have a distinct microbiota composition when compared to healthy hosts (Zhang et al. 2015). Overall, the symbiotic relationship critical for host health is molded by the evolutionary history between humans and their microbiota (Ochman et al. 2010).

Over 20 years ago, Abigail Salyers recognized strain diversity in the microbiome with a collection of over 200 *Bacteroides* isolated from community and clinical isolates (Shoemaker et al. 2001). Her laboratory demonstrated, employing classical methods like DNA hybridizations, the significant role of HGT in spreading antibiotic resistance genes, shedding light on the potential impact of human antibiotic overuse (Shoemaker et al. 2001). Advancements in sequencing technology have accelerated in gut microbiome research (Wensel et al. 2022), with next-generation sequencing of the gut microbiome revealing potential for disease diagnosis, such as detecting colorectal cancer risk (Wensel et al. 2022, Wirbel et al. 2019). A recent study showed an ever-increasing catalog of over 154 million nonredundant genes assembled from over 10,000 metagenomes (Xie et al. 2021). Metagenomic sequencing unveils significant genetic diversity, characterized by diverse rearrangement including inversions, duplications, and insertions, with HGT emerging as the key driver of this diversity (Pasolli et al. 2019, Carrow et al. 2020).

Horizontal Gene Transfer (HGT)

HGT refers to the process by which genetic information is shared between or acquired by organisms during their lifetimes. This contrasts with vertical inheritance, where genetic information is received directly from the parent organism(s). Among microbes, mobile genetic elements (MGEs) are one mechanism for the transfer of this genetic information. As their name suggests, MGEs are sequences of genetic material that can transfer horizontally (i.e., plasmids, integrative and conjugative elements, conjugative transposons, phages, and transposons) (Khedkar et al. 2022).

MGEs play a critical role in bacterial evolution, significantly shaping fitness though the acquisition of advantages phenotypes, including evasion of host immune responses (Avery et al. 1944), acquisition of virulence genes (Herold et al. 2004), acquisition of scarce resources in competitive environments (Frye et al, 2021), and the development of antibiotic resistance (Shoemaker et al. 2001, Partridge et al. 2018). However, HGT may also pose adverse effects if the transferred genes lack functionality, are incompatible, or self-replicate (Vogan & Higgs 2011).

There are three typical routes for HGT in microbes: transformation, transduction, and conjugation. Transformation involves competent bacteria take in external DNA from the environment and incorporating it into their existing replicons (e.g., chromosome or plasmid). This external DNA can originate from lysed bacterial cells, live bacteria releasing DNA, or be artificially induced *in vitro* by combining bacteria and DNA (Lorenz & Wackernagel 1994). Transduction occurs when bacterial cells receive bacterial DNA from bacteriophages (phages), viruses that infect bacteria. During infection, genes that have recombined into the phage genome or are adjacent to the bacteriophage attachment sites may be packaged with the bacteriophage DNA, allowing the transfer of these new bacterial genes during subsequent infection cycles (Chiang et al. 2019). Conjugation involves a bacterial cell directly transferring genetic information to another bacterial cell by replicating its DNA, forming a sex pilus, and facilitating the migration and integration of the newly generated DNA copy into a recipient cell chromosome or plasmid.

Conjugative Transposons (CTns)

Originally discovered in gram-positive cocci, CTns represent hybrid elements combining characteristics of transposons, plasmids, and bacteriophages (Salyers et al. 1995). Like transposons, CTns excise from and integrate into DNA, but they differ because they do not duplicate their target site during integration (Bedzyk et al. 1992). Resembling plasmids, CTns possess a covalently closed circular transfer intermediate and are transferred via conjugation; however, unlike plasmids, CTns do not replicate independently of chromosomes (Rice & Carias 1994). Additionally, CTns demonstrate phage-like characteristics in their excision and integration processes, reminiscent of bacteriophages with circular intermediates (Poyart-Salmeron et al. 1990). Sequence analysis on integrases found in some CTns suggests their affiliation with the λ integrase family (Poyart-Salmeron et al. 1990). Despite these similarities, CTns do not form viral particles and are transferred solely via conjugation, not transduction. In summary CTns are integrated elements transferred through conjugation, forming a covalently closed circular intermediate after excision (Salyers et al. 1995).

The scientific community has undergone several changes in terminology regarding CTns. In the 1990s, there were only a few genomic islands encoding self-conjugative transfer and integration into a recipient cell genome were described, therefore the classifications for these elements were mostly limited to CTns and integrative plasmids (Bellanger et al. 2014). However, the term integrative and conjugative mobile elements (ICEs) was later proposed as an umbrella term for all self-transmissible

integrative and conjugative elements, regardless of their integration or conjugation mechanisms (Wozniak & Waldor 2010, Burrus et al. 2002). While CTns and ICEs are often used interchangeably in literature, ICEs encompass a broader range of elements, comprising of CTns such as CTnDOT from *Bacteroides thetaiotaomicron* and Tn*916* from *Enterococcus faecalis*, which integrate into the host chromosome with minimal sequence specificity, alongside elements like SXT from *Vibrio cholerae* that have more restrictive integration sites (Franke & Clewell 1981, Burrus & Waldor et al. 2003, Wozniak & Waldor 2010).

Detecting Mobile Genetic Elements (MGEs)

In 1928, Frederick Griffith discovered the first recorded instance of HGT, where a "transforming principle" that observed the transfer of virulence factors from a virulent strain of *Diplococcus pneumoniae* to a non-virulent one (Griffith 1928). Nearly two decades later, the Avery-MacLeod-McCarty experiment demonstrated that DNA is the substance that causes the change in *D. pneumoniae* virulence phenotype (Avery et al. 1944). Over time, the impact of HGT in driving bacterial evolution has been proven time and time again (Boto 2009, Jain et al. 2003, Niehus et al. 2015). As a result, various computational and experimental methods have been developed to detect them in order to track these elements that can be a major force in affecting bacterial fitness, diversity, and ecology.

Before the widespread adoption of whole genome sequencing, researchers had already identified pathogenicity islands. Coined in 1990, the term was first referred to

specific DNA sequences encoding fimbrial adhesins and hemolysins, which significantly contribute to the pathogenicity of *Escherichia coli* strains associated with urinary tract infections, sepsis, and neonatal meningitis (Hacker et al. 1990). Subsequent genome alignments of enterohaemorrhagic EDL933, uropathogenic CFT073, and non-pathogenic MG1655 revealed that only 39.2% of their collective set of proteins were shared across all three strains (Welch et al. 2002). Furthermore, the analysis revealed that while various *E. coli* strains share a similar chromosomal sequence backbone, the backbone is interrupted by islands, likely due to horizontal gene transfer (Welch et al. 2002). As sequencing costs continue to decrease and sequencing technologies advance, the growing accessibility of genomes increases the potential for new MGEs to be found.

To identify new MGEs, computational methods such as identifying protein family domains (Pfams) associated with MGEs (i.e. Type IV Secretion Systems) or dedicated tools like MobileElementFinder (Johansson et al. 2021) and *PhiSpy* (Akhter et al. 2012). These dedicated tools use algorithms that leverage characteristics associated with MGEs (e.g., AT/GC base composition, multi-gene operons, inverted repeats/direct repeats) to effectively detect and characterize MGEs within microbial genomes.

MobileElementFinder can detect various types of integrating MGEs (iMGEs) in assembled sequence data and annotates their relationship to antimicrobial resistance, virulence genes and plasmids (Johansson et al. 2021). These iMGEs include miniature inverted repeats (MITEs), insertion sequences (ISs), composite transposons (ComTns), integrative and conjugative elements (ICEs), integrative mobilizable elements (IMEs),

and cis-mobilizable elements (CIMEs) (Johansson et al. 2021). MobileElementFinder was designed as a user-friendly tool that does not require as much bioinformatics expertise to operate and is not limited to a specific MGE type (e.g. ICEberg, ISfinder, Plasmidfinder) (Liu et al. 2019, Siguier 2006, Carattoli 2014).

The bioinformatics tool, *PhiSpy*, specializes in detecting prophages, which are phages integrated into host bacterial genomes. *PhiSpy* analyses are done in the absence of homology to known phage proteins, allowing for the discovery of new prophages by assessing protein length, transcription strand directionality, customized AT and GC skew, and the abundance of unique phage DNA sequence words (Akhter et al. 2012). Additionally, *PhiSpy* employs similarity-based approaches for comprehensive prophage identification, evaluating each predicted region through duplicate *att* site identification and phage protein similarity (Akhter et al. 2012). Remarkably, *PhiSpy* achieved a 94% prophage detection rate with only a 6% false-negative rate and a 0.66% false-positive rate across 50 bacterial genomes (Akhter et al. 2012).

Computational methods are effective at identifying potential MGEs in bacterial genomes (Akhter et al. 2012, Johansson et al 2021, Ozer et al. 2014, Roux et al. 2015). However, relying solely on predictions cannot definitively confirm the mobilization capability of a MGE. Often, additional experimental methodologies such as targeted mutations or pulse field gel electrophoresis are necessary to validate these predictions, or researchers must rely on capturing chance transfer events in genomic data (Coyne et al. 2014). To bypass the need for initial computational analyses and targeted mutations,

employing untargeted experimental methods appears to be a promising strategy for discovering novel and functional MGEs.

For instance, the transposon-aided capture (TRACA) allows for the preferential capture of circular extrachromosomal MGEs (i.e., plasmids) by shearing chromosomal DNA with a plasmid-safe DNAse, tagging the remaining circular DNA with a Tn5 transposon, and transforming it into *E. coli* (Jones and Marchesi 2006). This method can detect MGEs that lack selectable markers due to the antibiotic resistance phenotype carried with the Tn5 transposon. Another method akin to TRACA is multiple displacement amplification (MDA). Similar to TRACA, DNA for MDA protocols is digested using a similar DNAse. Then the undigested plasmid DNA is amplified using random hexamer primers and phi29 polymerase (Delaney et al. 2018). Phi29 polymerase extends from the bound primer of the adjacent fragments, generating a branched network of amplified DNA. The amplified plasmids can then undergo high-throughput sequencing or be introduced into a bacterial host such as *E. coli* (Delaney et al. 2018). Additionally, entrapment vectors like pBACpAK detect MGEs through the random insertion of into a repressor gene (*cI*), leading to the depressesion of a promoter (P_R) associated with a tetracycline-resistance gene (*tet*(A)), thereby granting tetracycline resistance to transconjugants (Tansirichaiya et al. 2022).

The TRACA and MDA experimental methods have the proven ability to identify extrachromosomal elements and can be used to test host range or compatibility of the detected elements. However, neither of these methods are designed to specifically detect self-transmissible MGEs or *in vivo* HGT events. In contrast, entrapment vectors can

detect *in vivo* mobilization events, although the MGE must be able it insert/integrate into a specific target region (e.g., *cI*). Many MGEs encode site-specific integrases that may not be able to integrate into the target region and therefore their mobilization events would go undetected. As such additional approaches that can detect a wide range of *in vivo* MGE transfer events would contribute greatly to this field of study.

Bacterial Conjugation – Mechanisms

CTns are found integrated in host chromosomes and contain the genes needed not only for replication and conjugation but for integration and excision as well. Conjugative plasmids are generally categorized separately as they exist as extrachromosomal elements that replicate independently from the host chromosome. However, both CTns and conjugative plasmids are MGEs that harbor genes that produce the multi-protein complex required for conjugation (Johnson & Grossman 2015).

Conjugative transfer is a well-studied phenomenon that starts with cell-to-cell contact, followed by mating pair formation (Mpf) systems, and then the transfer of DNA through a conjugative pilus. The components of Mpf systems consist of a set of conserved proteins that form a membrane protein complex and a sex pilus (Schröder & Lanka 2005). A separate set of genes are then responsible for the DNA replication and system transfer that replicates and processes of the CTn ssDNA that will be mobilized through the sex pilus (Schröder & Lanka 2005). These conjugative pili are members of the Type-IV Secretion System (T4SS) family. However, not all T4SSs are used for conjugation. Divergent T4SSs have been identified that are involved in the uptake and

release of DNA during transformation and the secretion of toxic effector proteins in pathogen-host interactions (Wallden et al. 2010, Stingl et al. 2010, Ramsey et al. 2011).

Normally inactive CTns remain dormant within the host genome. However, upon induction, such as with CTnDOT, a CTn carrying genes for tetracycline and erythromycin resistance (Tet^R & Erm^R), the CTn excises itself from the chromosome, forming a circular, plasmid-like dsDNA molecule (Whittle et al. 2006). Subsequently, the Mpf genes initiate the formation of a mating pore, facilitating the transfer of ICE DNA. Host and ICE-encoded proteins identify the origin of transfer (*oriT*) and process the ICE DNA into a linear ssDNA-protein complex. This complex is then transferred into the recipient via a formed T4SS. Within the recipient, the ICE re-circularizes, then the DNA polymerase synthesizes the second strand returning it to a dsDNA molecule, and then the ICE recombines into the chromosome utilizing an ICE-encoded recombinase, referred to as an integrase (Johnson & Grossman 2015).

Bacterial Conjugation – History

Bacterial conjugation is a widely conserved and effective strategy to horizontally transfer genes in microbial populations. Interestingly, bacterial conjugation plays a pivotal role genetic exchange that extends beyond the bacterial kingdom. Instances of conjugative DNA transfer from bacteria to eukaryotic organisms such as yeast and mammalian cells have been documented (Heinemann & Sprague 1989, Waters 2001). Conjugation was first discovered in *Escherichia coli* in 1946 by Joshua Lederberg and Edward Tatum. They observed evidence of a sexual process and found that bacteria could

change genetic information through a unidirectional transfer of DNA (Lederberg & Tatum 1946). The MGE that was subsequently identified in these studies by Esther Lederberg, Luigi Cavalli-Sforza, and Joshua Lederberg was coined "F-plasmid", and became the first plasmid discovered (Cavalli-Sforza et al. 1953). The term "plasmid" was proposed by Joshua Lederberg to describe any extrachromosomal hereditary element (Lederberg 1952). Note that future researchers modified this definition to not include bacterial viruses by defining plasmids as autonomously replicating genetic elements that exist only or mainly outside of the chromosome. In the decades since the discovery of the F-plasmid numerous conjugative elements have been discovered, including CTns.

CTns were first discovered in gram-positive *Enterococcus faecalis* called Tn*916* (Franke & Clewell 1981) and in *Streptococcus pneumoniae* with Tn*5253* (Ayoubi et al. 1991). Both Tn*916* and Tn*5253* differ in size by ~41.5kb, with Tn*5253* being identified as a combined structure comprising of two smaller elements, Tn*5251* resembling Tn*916*, that have integrated into Tn*5252* (Ayoubi et al. 1991). As subsequent studies identified additional CTns, researchers routinely found that many CTns had striking similarities to one another but often contained significant gene content differences. For example, Tn*1545* is nearly identical in sequence to Tn*916*, apart from a ~25kb sequence consisting of kanamycin and erythromycin resistance genes encoded by Tn*1545* (Clewell et al. 1995). This pattern began to shape an understanding of the importance of CTns for their susceptibility to variable gene gains and losses as well as mobilizing antibiotic resistance.

At the time Clewell et al. 1995 was published, the Tn*916*/Tn*1545* CTn family was considered to have a broad host range with the Tn*916*/Tn*1545* being found in 24 genera of Bacteria. However, with more recent advancements in sequencing technologies, studies have shown the number of genera increasing to at least 36 (Ciric et al. 2013). Interestingly, Tn*916* shows such a broad host range that some gram-negative bacteria including known human pathogens *Neisseria gonorrhoeae* and *Kingella denitrificans*, have them (Knapp et al. 1988). The ability of CTns like the Tn*916*/Tn*1545* family to transfer among such diverse hosts and bring along various antibiotic resistance genes is of great concern for the clinical treatment of infectious diseases.

With the hundreds of thousands of bacterial genomes sequenced, CTns have been discovered in an enormous diversity of species, including known human pathogens (e.g. *Pseudomonas aeruginosa, Streptomyces* spp., *Yersinia pestis,* and *Escherichia coli*) harboring CTns encoding pathogenicity islands (Mikkelsen et al. 2013, Kers et al. 2005, Schubert et al. 2004). In gut associated Bacteroidota, there are a wide array of MGEs (e.g. plasmids, mobilizable transposons, CTns), with many of them harboring antibiotic resistance genes (Whittle et al. 2002).

CTnDOT in Bacteroidota

Discovered in the chromosome of a *Bacteroides* spp. clinical isolate, CTnDOT became one of the most extensively characterized CTns in *Bacteroides* is CTnDOT (Wood & Gardner 2015). CTnDOT is a model CTn among gut Bacteroidota, as it plays a pivotal role in the dissemination of tetracycline and erythromycin resistance (Tet^R and

Erm^R) (Cheng et al. 2000). Early studies using Southern hybridizations and PCR of marker genes determined that >80% of natural *Bacteroides* clinical and community isolates in 1990 were Tet^R, compared to 20–30% in the 1970s (Shoemaker et al. 2001). The massive increase in Tet^R was largely attributed to the increased mobilization of CTnDOT among these gut microbes likely due to the common prescription of tetracycline during this time (Shoemaker et al. 2001). Like other ICEs, CTnDOT encodes the genes necessary for its excision and transfer to recipient cells via conjugation and the ability to integrate into the recipient cell chromosome.

CTnDOT consists of the *intDOT* gene, coding for the tyrosine recombinase necessary for the integration and excision of CTnDOT. The excision operon contains *xi2c*,*xis2d*, *orf3*, and *exc* genes, where all genes except *orf3* are known to be associated with excision. Next is the *mob* operon (*mobA*, *mobB*, *mobC*) that encodes the relaxase and coupling proteins needed for mobilization and the *tra* region genes involved in conjugation. Finally, there is the *tetQ-rteA-rteB* operon with *tetQ* encoding a ribosomal protection protein for tetracycline resistance and a predicted two-component regulatory system with *rteA* and *rteB*, an *ermF* region, encoding resistance to erythromycin, unique from a similar CTn found in *Bacteroides*, CTnERL, *tetX1*/*tetX2* also conferring tetracycline resistance and *aaDs* for streptomycin resistance (Moon et al. 2007, Whittle et al. 2001).

Mechanism of CTnDOT

Horizontal transfer of CTnDOT requires a series of key steps including excision, mobilization, transfer, replication, and integration. (Waters & Salyers 2013). While the *tetQ-rteA-rteB* operon is constitutively expressed, translation only happens from exposure to tetracycline (Wang et al. 2005). Once exposed to tetracycline, translation resumes, enabling RteB to initiate RteC transcription (Cheng et al. 2001, Jeters et al. 2009). RteC is then regulates the transcription of the excision operon (*xis2c*, *xis2d*, *orf3*, *exc*), which facilitates the removal of CTnDOT from the chromosome (Keeton et al. 2013). Xis2c and Xis2d regulate the *tra* operon, while Xis2d and Exc stimulate the transcription of the *mob* region, allowing CTnDOT to transfer to the recipient cell (Waters & Salyers 2013). However, tetracycline induction to this multilayered regulatory cascade can be overridden by a series of negative regulators that prevent the premature transfer of CTnDOT. CTnDOT contains sequence for the 90nt transcript of small RNA, RteR (Waters & Salyers 2012). Evidence indicates that the presence of *rteR* on pLYL72, a plasmid that contains an 18kb-region of CTnDOT that can self-transfer, *in trans* results in many of the *tra* genes minimally detectable, suggesting the potential role RteR in initiating premature transcription termination within the *tra* operon (Waters & Salyers 2012). Interestingly, when *intDOT* is present on pLYL72 *in trans* CTnDOT transfer is undetectable regardless of tetracycline induction (Waters & Salyers 2012).

Other regulatory mechanisms of CTns

CTns have complex regulatory networks that control integration excision and transfer. As a model, CTnDOT exhibits multi-layered regulatory mechanisms involving transcriptional and translational regulation. However, there is tremendous diversity of regulatory mechanisms among other CTns.

CTns combine features of other classes of MGEs such as phages that integrate into and excise from the host chromosome but are obviously different because phages are not transmitted by conjugation like CTns (Wozniak et al. 2010). Another characteristic shared by CTns and phages is critical role of regulators encoding helix-turn-helix motifs. Pioneering work in phage λ identified the key transcriptional regulators the Cro and CI repressors. In summary, the CI protein is expressed and maintains the repression of the integrated phage (lysogeny/prophage), while the Cro repressor when expressed turns off CI and enables the lytic cycle to begin (Takeda et al. 1977). Both Cro and CI repressors encode a helix-turn-helix structural DNA-binding motif consisting of two alpha helical regions of a protein joined by an amino acid sequence that allows the alpha helices to turn (Brennan & Matthews 1989). Subsequent studies have concluded that the helix-turn-helix domain is highly common in transcription factors found in all three domains of life, including the widespread CTn originally found in *Enterococcus* faecalis, Tn*916* (Aravind et al. 2005, Wright & Grossman 2016).

Like Tn*916,* the CTn SXT has the SetR repressor, another homologue of the phage λ CI repressor, regulates transfer by binding to operators upstream of *setR* to

inhibit the expression of an operon encoding master transcriptional activators (*setD* and *setC*) of genes required for SXT transfer (*int* and *tra* operons) (Burrus & Waldor et al. 2003). However, other CTns have distinct means of activation for instance ICE*Bs*1 uses the SOS response and the response regulator aspartate phosphatase I (RapI) to regulate the ImmR-ImmA operon (Auchtung et al. 2005, Bose et al. 2008). ImmR is inactivated by the proteolytic inactivation of ImmA, causing a derepression of an operon encoding the conjugation machinery and excisionase (Bose et al. 2008). Further CTns like ICE*clc*^{B13} use an integrase regulator (InrR) and certain environmental stimuli (i.e. CBA, a compound degradable by ICE*clc*^{B13} proteins) to upregulate the expression from a weak promoter (P_{int}) which induces ICE*clc*^{B13} excision (Sentchilo et al. 2003). Afterwards, the circularization of ICE*clc*^{B13} places a strong constitutive promoter (P_{circ}) is brought near *int*, allowing for the increase in *int* expression (Sentchilo et al. 2003). Therefore, the circularization of ICE*clc*^{B13} overall increases *int* expression, allowing for the integration of the element into the recipient or donor cell (Sentchilo et al. 2003). We note that in the vast majority of instances, regulation of computationally identified CTns are not understood.

Core/Pan Genomes

The first bacterial genome was sequenced less than thirty years ago (Fleischman et al. 1995). As subsequent genome sequences were compared to one another, it became clear that there were significant sequence differences, even between strains within a bacterial species (Welch et al. 2002). Bacterial genomes can be dynamic and have a

complex evolutionary history due to the various factors that can change it, such as gene loss/duplications, selfish DNA, homologous recombination, and HGT (Daubin & Ochman 2004, Wozniak et al. 2010). In 2005, the term "pangenome" was introduced to represent the entire set of genes within a group of analyzed genomes (Tettelin et al. 2005). Typically, the genomes being analyzed represent strains of the species although they can also represent other relevant groups of microbes. Within the pangenome genes may be classified as "Core" that is present in all genomes or "Variable" genes present individually or among subsets of the analyzed genomes (e.g., the accessory genome) (Tettelin et al. 2005). Generally, a significant portion of core genes consists of essential housekeeping genes vital for cell survival and are often constitutively expressed, such as genes involved in DNA replication, transcription, and translation (Chintan et al. 2002). However, when identifying core genes within narrow subsets, such as within a species (Tantoso et al. 2022, Wallace et al. 2022) or a group of CTns (Ortañez & Degnan 2024b), intriguing interrelationships within the group can be discovered.

Transposon mutagenesis

Transposable elements (TEs) are DNA sequences that are capable of changing their position within a genome (Bourque et al. 2018). TEs were first characterized by Barbara McClintock's groundbreaking discoveries that illuminated the dynamic nature of genomes and pioneered the concept of epigenetics. Discoveries that predated the elucidation of the molecular structure of DNA (McClintock 1931, McClintock 1950, Watson & Crick 1953). In the decades since their discovery and the diverse types of TEs

identified, TEs remain broad class of DNA sequences that capable of changing position within genomes (Bourque et al. 2018).

Transposon mutagenesis is a method that leverages the ability of TEs to insert into DNA sequences to generate randomized mutations in bacterial genomes. This approach often yields high mutation frequencies, single-hit mutations, enables the recovery of mutated genes post-mutagenesis, and facilitates the introduction of selectable markers into a genome for strain construction (Seifert et al. 1986). Mutant cells can then be assayed for expected or unexpected changes in phenotype caused by Tn gene disruptions. Mutant screens can also reveal non-essential genes that have no effect on a strain's phenotype (Hamer et al. 2001). Notably, the initial methodologies for identifying the Tn insertion locations leading to mutant phenotypes were challenging and often low throughput.

Despite the occasional complications, the experimental value of Tn mutagenesis led to the development of applications that work among all domains of life (Lampe et al. 1996, Goodman et al. 2009, Seifert et al. 1986, Mazurkiewicz et al. 2006, Kilijunen et al. 2017). However, through the application of next generation sequencing the identification of Tn insertion locations was radically improved (Goodman et al. 2009, van Opijnen et al. 2009, Cain et al. 2020). Goodman et al. (2009) devised a method of generating a transposon cassette that through a process of arbitrary PCR could produce, amplified fragments that could be sequenced directly on the Illumina sequencing platform. These sequenced fragments capture the insertion location of each individual Tn and can be used

on large populations of mutants enabling studying tens to hundreds of thousands of mutants simultaneously (INSeq/TnSeq). This method utilizes a sequencing-adapted *mariner* transposon plasmid (pSAM) carrying the *mariner* transposase initially isolated from the horn fly, *Haematobia irritans* (Lampe et al. 1996). The transposon cassette was engineered to no longer encode the transposase and instead encodes an antibiotic resistance cassette flanked by MmeI-modified inverted repeats. Additionally, the vector includes the necessary apparatus for replication in the donor strain and subsequent transfer via conjugation (Goodman et al. 2009). This methodology is highly successful in a number of human gut microbes (Cullen et al. 2015) and has great promise for additional experimental applications.

Conclusion

Many studies have proved the importance of gut microbe CTns in shaping the microbial diversity of the human gastrointestinal tract, which affects the overall health of the host they reside in. While there have been major strides in identifying and characterizing these CTns since the first CTn was discovered in 1981, most CTns are not well characterized in terms of mobility or possible fitness effects. Further experimental studies are the key to more clearly understanding the complex systems that define the gut microbiome.

CHAPTER 2: Tracking and Characterization of a Novel Conjugative Transposon Identified by Shotgun Transposon Mutagenesis

Abstract

The horizontal transfer of mobile genetic elements (MGEs) is an essential process determining the functional and genomic diversity of bacterial populations. MGEs facilitate the exchange of fitness determinant genes like antibiotic resistance and virulence factors. Various computational methods exist to identify potential MGEs, but confirming their ability to transfer requires additional experimental approaches. Here, we apply a transposon (Tn) mutagenesis technique for confirming mobilization without the need for targeted mutations. Using this method, we identified two MGEs, including a previously known conjugative transposon (CTn) called *Bo*CTn found in *Bacteroides ovatus* and a novel CTn, *Pv*CTn, identified in *Phocaeicola vulgatus*. In addition, Tn mutagenesis and subsequent genetic deletion enabled our characterization of a helix-turn-helix motif gene, BVU3433 which negatively regulates the conjugation efficiency of *Pv*CTn *in vitro*. Furthermore, our transcriptomics data revealed that BVU3433 plays a crucial role in the repression of *Pv*CTn genes, including genes involved in forming complete conjugation machinery (Type IV Secretion System (T4SS)). Finally, analysis of individual strain genomes and community metagenomes identified the widespread prevalence of *Pv*CTn-like elements with putative BVU3433 homologs among human gut-associated bacteria. In summary, this Tn mutagenesis mobilization method (TMMM) enables observation of transfer events *in vitro* and can ultimately be applied *in vivo* to

identify a broader diversity of functional MGEs that may underly the transfer of important fitness determinants.

Introduction

Mobile Genetic Elements (MGEs) are important drivers of bacterial evolution by promoting gene acquisitions that can profoundly affect a bacterial host's fitness. Known fitness determinants that can be mobilized include those that enable evasion of host immune responses (Rendueles et al. 2018) and antibiotics (Partridge et al. 2018), as well as the ability to intoxicate hosts (Schmidt et al. 1999) and acquire scarce resources in competitive environments (Frye et al. 2021). Furthermore, sequence evidence indicates that extensive interspecies transfer of MGEs among the Bacteroidota, a Gram-negative phylum of bacteria (formerly Bacteroidetes) that is common in the human gut and can represent as much as 80% of the microbiome of some individuals (HMP 2012, Coyne et al. 2014, Smillie et al. 2010, Wexler 2007, Nayfach et al. 2019). As a result, computational methods to identify, classify, and determine the prevalence of potential MGE activity from DNA sequences have become increasingly sophisticated (e.g., Akhter et al. 2012, Johansson et al. 2021, Ozer et al. 2014, Roux et al. 2015). However, given the structural and sequence diversity of MGEs (Osborn and Böltner 2002), the possibility of inactivating mutations and the presence of cooperative mobilizing elements (Duerkop et al. 2012), confirming the activity of a computationally predicted MGEs generally requires experimental investigation.

Many confirmed MGEs in Bacteroidota have obvious phenotypes that readily enabled the characterization of their functions (e.g., antibiotic resistance) (Wood & Gardner 2015, Waters & Salyers 2013, Schlesinger et al. 2007). In contrast, many predicted MGEs lack evidence for obvious functions that would be amenable to a genetic screen (Durrant et al. 2020). Further, the number of predicted MGEs per genome can be quite high, where among *Bacteroides* spp. 20–50% have at least one plasmid (Wexler 2007) and 80% encode at least one conjugative transposon (CTn)(Shoemaker et al. 2001). As such, strategies that avoid 10s or 100s of targeted mutations to enable screening for mobilization have the potential to accelerate experimental validation of MGE activity and functions.

One such attempt to implement an untargeted approach to identify functional MGEs in *Enterobacteriaceae* was developed by Tansirichaiya et al. (2022). The authors constructed a type of entrapment vector (Gay et al. 1985), pBACpAK, that expresses its tetracycline resistance (Tet^R) allele when a *cI* repressor gene is disrupted by a MGE (Tansirichaiya et al. 2022). This approach enabled the detection of MGEs integrating into a focal species, or smaller elements replicating within the focal species (e.g., IS elements). However, this system is limited to MGEs that can integrate into the ~0.6kb *cI* repressor gene and to species within the *Enterobacteriaceae*. These limitations suggest the opportunity for additional untargeted approaches to identify and track functional MGEs. Transposon (Tn) mutagenesis is a commonly used method to generate untargeted genomic mutations and characterize gene functions among all domains of life (Kiljunen

et al. 2017, Lampe et al. 1996). We propose combining Tn mutagenesis with screens for horizontal gene transfer (HGT) in bacteria to efficiently assess MGE activity.

Here we evaluate the effectiveness of capturing MGEs by combining *mariner* Tn mutagenesis and HGT screens for a group of human gut-associated bacteria. Through this untargeted Tn mutagenesis mobilization method (TMMM), we successfully tracked the mobilization of two conjugative transposons (CTns), including a novel CTn in *P. vulgatus* ATCC 8482 that we have named *Pv*CTn. Further, this work provides insight into the regulatory mechanisms of *Pv*CTn, most notably the helix-turn-helix motif encoding gene *BVU3433*. We also provide computational evidence for the prevalence of *Pv*CTn-like elements among a panel of human gut-associated bacterial genomes and metagenomes from globally distributed patient samples. The observed mobilization of *Pv*CTn and the widespread presence of genetically diverse *Pv*CTn-like elements in Bacteroidota reinforce the importance of MGEs shaping the human microbiome.

Materials and Methods

Strains, Plasmids, Culture Conditions

All Bacteroidota cultures were grown on Brain Heart Infusion (BHI) agar supplemented with 10% defibrinated horse blood (BHI-HB; Quad Five, Ryegate, MT) and tryptone-yeast extract-glucose broth (TYG) medium with and without agar using an anaerobic chamber (Coy Laboratory Products, Grass Lake, MI) filled with 73% N_2, 20% CO_2, and 7% H_2. *E. coli* with the appropriate plasmid(s) (pSAM, pNBU2, pExchange, pLGB13) were grown on LB. All strains and plasmids are listed in Table 2.1. Ampicillin (100

μg/mL), gentamicin (200 μg/mL), erythromycin (25 μg/mL), tetracycline (2 μg/mL), cefoxitin (20 μg/mL), 5'-fluorodeoxyuridine (FUdR; 20 μg/mL), and anhydrotetracycline (aTC; 100 ng/mL) were added to the media when appropriate.

During conjugations, donors with conjugative plasmids (pNBU2, pExchange, pLGB13) or conjugative transposons (*Pv*CTn) and Bacteroidota recipients were grown overnight in 5 mL LB and TYG medium respectively with the appropriate antibiotic(s). Overnight stationary phase *E. coli* strains were then used to inoculate 10mL subcultures of LB at the following dilutions: 1:200, 1:500, and 1:750 dilutions. *B. thetaiotaomicron*, *B. ovatus*, and *P. vulgatus* strains were similarly inoculated into TYG but at lower dilutions: 1:25, 1:50, and 1:100. Subcultured *E. coli* S-17 were shaken aerobically at 250 rpm at 37ºC. Bacteroidota were subcultured in an anaerobic chamber, stoppered, removed from the chamber, and incubated statically at 37ºC. Growth was monitored and cells were pelleted (4,000 x g for 5 min) when they reached an OD_{600} of ~0.4. The supernatants were removed, and cells were resuspended with 1mL TYG medium and pelleted again. After removing the supernatant, 1mL TYG medium was used to combine donors and recipients. The combined cells were then spread onto BHI-HB agar. Conjugation plates with *E. coli* donors were incubated aerobically, while dual-Bacteroidota conjugations were incubated anaerobically. Both aerobic and anaerobic conjugations were incubated at 37ºC for 24 hrs. Conjugation masses were then scraped and resuspended in 5 mL TYG medium. The resuspended conjugation masses were then plated onto LB or BHI-HB agar with serial 10-fold dilutions and antibiotic supplement(s). Conjugation efficiencies were calculated using the following equation:

$$\frac{\text{Recipient } \frac{\text{CFUs}}{\text{mL}}}{\text{Transconjugant } \frac{\text{CFUs}}{\text{mL}}}$$

Effects of Peroxide Stress on Conjugation

Conjugation experiments were performed as described above, however, once donor and recipient Bacteroidota cells were mixed and pelleted after reaching an OD_{600} of ~0.4, the cells were plated on TYG agar where 880μM H_2O_2 was surface spread 30 min before plating when appropriate. The resuspended conjugation masses were plated on BHI-HB agar with serial 10-fold dilutions and antibiotic supplement(s). Conjugation efficiencies were calculated as described above.

Transposon Mutagenesis Based Identification of MGEs

Generating Transposon (Tn) mutant libraries - *B. thetaiotaomicron*, *P. vulgatus*, and *B. ovatus* were mutagenized through conjugation with *E. coli* S17-1 containing the sequencing-adapted *mariner* transposon plasmid (pSAM) harboring an erythromycin resistance (Erm^R) cassette (Goodman et al., 2009). Conjugations were carried out as above, however, 100 μL aliquots of the entire conjugation masses were plated on 50 BHI-HB plates supplemented with erythromycin and incubated anaerobically at 37ºC for 48 hrs. Colonies were collected and pooled in TYG + 20% glycerol and stored at -80ºC in 500 μL aliquots.

Screening Tn libraries for MGEs - Aliquots of Tn mutant library donors (Erm^R) were then conjugated with Tet^R recipient strains of *B. thetaiotaomicron* VPI-5482, *B. thetaiotaomicron* 3731, *Parabacteroides merdae* ATCC 43184, *B. thetaiotaomicron* VPI-5482 ΔCPS, and *B. uniformis* ATCC 8492. To screen for MGE transfers, the conjugation masses were plated onto BHI-HB supplemented with tetracycline and erythromycin. After incubation for 48 hrs at 37ºC, colonies were re-struck for colony purification of the putative transductants or transconjugants. Individual colonies were then grown overnight in liquid TYG supplemented with tetracycline and erythromycin at 37ºC for 24 hrs.

Arbitrary PCR - To map pSAM's integration sites, arbitrary PCR assay was used as described previously (Goodman et al. 2009). Amplicons were cleaned using a QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and submitted for Sanger sequencing through the University of California, Riverside Institute for Integrative Genomic Biology (UCR IIGB) core. The sequences were then used as BLASTn search queries against the *B. thetaiotaomicron* (Refseq:NC_004663), *B. ovatus* (Refseq:NZ_AAXF00000000), and *P. vulgatus* (Refseq:NC_009614) reference genomes.

Cloning and Mutagenesis (pExchange, pLGB13, pNBU2)

Multiple vector systems were employed to generate the required strains. First, deletion of *BVU3433* mutants was achieved using the *P. vulgatus* specific thymidine kinase (*tdk*) allelic exchange system (Campbell et al., 2020). Briefly, the ~1kbp regions flanking *BVU3433* were individually amplified, then combined and amplified in the splicing by

overlap extension (SOE) reaction. All products were amplified using KAPA HiFi Taq MasterMix (KAPA BIOsystems, Wilmington, MA) with the primers listed in Table 2.1. The purified SOE product was then subjected to restriction digestion (NEB) and ligated into pExchange-*tdk*BV using T4 DNA Ligase (NEB).

Second, targeted insertional mutations of tetracycline resistance (*tetQ*) was achieved with pLGB13 which uses erythromycin-aTC counter selection (García-Bayona & Comstock 2019). We generated a ~2kb SOE product encompassing the intergenic spacer between two co-transcribed hypothetical proteins *BVU3417* and *BVU3418*. However, at the SOE junction within this intergenic spacer, we included a SpeI and a XmaI restriction sites (Table 2.1). In a two-step process, the SOE product was then digested, cleaned and ligated into pLGB13 as described above. Then a *tetQ* (Martens et al. 2008) cassette was amplified and ligated into the SpeI/XmaI restriction sites in the intergenic spacer.

Third, the *BVU3433* complementation construct was generated using pNBU2-*bla-CfxA* (Campbell et al., 2020). We amplified the gene and its native promoter (351 bp upstream of the start codon) and cloned it as described above into the multiple cloning site of the integrative plasmid pNBU2-*bla-CfxA*.

All ligation products were transformed into *E. coli* S-17 with electroporation. Individual ampicillin resistant colonies were isolated and purified plasmids were confirmed through PCR and Sanger sequencing. The confirmed allelic exchange and complementation vectors were then conjugated into the appropriate parent stains (e.g., *P. vulgatus* Δ*tdk*) and recombinant merodiploids were selected for on erythromycin. For the

deletion and insertion mutants, merodiploids were allowed to loop out. Resolved merodiploids resistant to 20μg/mL FUdR or 100ng/μL aTC were isolated and screened by PCR desired mutations.

Measuring growth kinetics of P. vulgatus

Wild-type and mutant *P. vulgatus* were grown and washed in TYG media. The cells were normalized and diluted to OD_{600} of 0.002 in TYG and dispensed in triplicate into a 96-well plate. Cell growth was measured every 30 min, over 36 hrs using a BMG Labtech CLARIOstar plate reader. Doubling times were calculated using the least-squares method for growth between 0.05 and 0.12 OD_{600} ($n = 3$).

RNA isolation and RNA-Seq

Samples for RNA-Seq of *P. vulgatus, P. vulgatus* Δ*BVU3433*, and *B. thetaiotaomicron Pv*CTn:*tetQ*::Δ*BVU3433* were grown overnight in 5mL TYG medium. Overnight cells were used to inoculate cultures at 10mL TYG at a final dilution of 1:25 in biological triplicate. Cell growth was monitored and harvested at an OD_{600} of ~0.4. Total RNA was extracted using a lysis buffer (Degnan et al., 2014) and prepared with a Qiagen RNeasy kit and treated with DNA-*free*™ DNA Removal Kit (Invitrogen).

Total RNA was also extracted from *B. thetaiotaomicron* and *P. vulgatus* during conjugation. Conjugation of the two strains was carried out as above. However, after 24

hrs of growth, the conjugation mass was scraped from the plate surface into TYG and the cells were immediately pelleted and total RNA was extracted as described above.

Library preparation of RNA was completed by following the Illumina (San Diego, CA) Stranded Total RNA Prep, Ligation with Ribo-Zero protocol and using a starting material of 500 ng of total RNA. The only variation from the Illumina protocol is that the volume used for each reagent was reduced by half. The library was submitted to the UCR IIGB core for quality analysis of the multiplexed samples and sequencing on an Illumina NextSeq mid output 75 bp paired end platform.

RNA-Seq read quality was determined through FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed using trimmomatic (SLIDINGWINDOW:4:15 LEADING:2 TRAILING:2 MINLEN:70)(Bolger et al, 2014). Transcript expression was calculated using Rockhopper with default parameters (McClure et al., 2013) and trimmed reads were mapped to the *P. vulgatus* ATCC 8482 reference genome (Refseq: NC_009614.1), *B. thetaiotaomicron* VPI-5482 (Refseq:NC_004663), and/or *Pv*CTn when appropriate.

cDNA preparation, qPCR

P. vulgatus Δ*tdk* strains were grown overnight in 5mL TYG medium. Overnight cells were used to inoculate cultures at 10mL TYG at a final dilution of 1:25 in biological triplicate. Cell growth was monitored and harvested at an OD_{600} of ~0.4. When appropriate, cells were perturbed by 880μM H_2O_2 and immediately harvested after 30

minutes of incubation at 37ºC. Total RNA was extracted using a lysis buffer (Degnan et al., 2014) and prepared with a Qiagen RNeasy kit and treated with DNA-*free*™ DNA Removal Kit (Invitrogen).

Complementary DNA (cDNA) was generated from 500 ng total RNA using SUPERase•In™ RNase Inhibitor (Invitrogen) and Superscript-II RT (Invitrogen), where the RNA template was eventually degraded by inoculating 1N NaOH for 30 min at 65°C and neutralized with 1N HCl. DNA was isolated using a QIAquick spin column (Qiagen) and eluted in 10 mM Tris-Cl, pH 8.5.

Quantification of *BVU3433* gene expression was measured through real-time quantitative PCR (qPCR) using Bio-Rad CFX96 Touch Real-Time PCR Detection System and SYBR Green (KAPA Biosystems) fluorescent dye. The CFX Maestro Software and $\Delta\Delta C_q$ method (Bookout et al., 2006) were used to process and calculate differences in *BVU3433* and *16s rRNA* (Table 2.2) expression. *BVU3366* and *BVU3378* were used as candidate genes to confirm RNA-Seq expression profiles.

Identification of PvCTn family among gut Bacteroidota

The boundaries of *Pv*CTn were predicted using functional gene annotations of the regions surrounding *BVU3433* and alignments of the genomic region with a panel of related *P. vulgatus* (*n*=13) and *P. dorei* (*n*=10) strains using Mauve (Darling et al. 2004) (Table 2.3). This included functionally characterizing the genomic region using HMMER v3 (hmmer.org) with trusted cutoffs to search the PFAM v35 (El-Gebali et al., 2019) and

TIGRFAM v15 (Haft 2001) databases (as in Frye et al., 2021). The CTn attachment sites (*attL* and *attR*) were determined using a combination of Mauve alignment inspection, BLASTn and manual sequence alignment.

The putative negative regulator BVU3433 was subsequently used to screen a total of 134 gut Bacteroidota genomes by BLASTp to identify homologs (bit score ratio ≥ 0.3) (Table 2.3). The process to identify the boundaries of *Pv*CTn was repeated for each homolog detected as described above, determining its genomic context and if the homolog was a part of an MGE. Identified genomic regions corresponding to putative CTns were then compared using pairwise BLASTn (E value ≤ 0.0001), filtered for ≥20% percent length aligned (PLA) (https://doi.org/10.5281/zenodo.5138177) and then clustered using MCL (Inflation = 20) (Enright 2002). Clusters were visualized in Cytoscape (Shannon 2003). Alignments of individual genes were performed using Muscle v3.8.1551 (Edgar 2004) and maximum likelihood phylogenies were reconstructed using FastTree v2.1.11 (Price et al 2010).

Regulatory protein and Promoter conservation analysis

Examination of the CTn clusters identified a conserved three gene regulatory region including *BVU3433* and genes *BVU3432* and *BVU_RS21835*. We extracted the intergenic region between *BVU_RS21835* and *BVU3433* (and homologs) for alignment and assessment for conserved sequence features. In addition, 250 nt upstream regions of differentially regulated operons were retrieved from *Pv*CTn and conserved genes in related CTns. MEME (Bailey et al. 2015) was used to analyze these regions for

conserved sequence patterns that may be involved as regulator binding sites. Searches were performed iteratively using different combinations of promoter regions, number of patterns retrieved (n= 5 – 6), maximum motif widths (n= 20 – 30), and motif distributions ('zero or one occurrence per sequence' or 'any number of repetitions').

Detection of PvCTn in human gut metagenomes

To identify the frequency of *Pv*CTn-like elements in human gut metagenome samples we employed a marker gene approach. First, nucleotide sequences for two sets of species-specific markers were retrieved from 46 representative genomes in our panel (Table 2.3) corresponding to universally conserved 30S ribosomal protein S5 (*rpsE*) and the *attB* site in pyruvate phosphate dikinase (*ppdK*). Second, 500 bp regions centered on the *attL* and *attR* of known *Pv*CTn-like elements were retrieved. Finally, the terminal 500 bp of *Pv*CTn-like integrases and the entire sequences of *BVU3433* homologs were retrieved. In the end, all markers were ~500 bp in size. Then short read metagenomic datasets PRJEB7774 (Feng et al. 2015), PRJEB12449 (Vogtmann et al. 2016) and PRJEB10878 (Yu et al. 2015) from healthy patients and cohorts with colorectal cancer were retrieved from the NCBI SRA database (Table 2.6). The data were quality filtered with Trimmomatic as described above and mapped to the 6 marker gene regions using Bowtie2 with the default settings (Langmead et al 2012). Conservative estimates of read coverage was measured for each sample based solely on the number of reads spanning the central 24 nt of a given marker gene. This ensured reads spanned the *attB*, *attL* and

attR. Read coverages were first normalized by the number of reads per sample, then to the read coverage of *rpsE*. Paired reads were also taxonomically classified using Kraken2 (Wood et al. 2019) using the default parameters and the pluspf database (downloaded 17 May 2021).

To evaluate the robustness of this analysis, simulated short read data were generated from the genomes of *P. vulgatus* ATCC 8483 and 14 additional strains with wgsim (-1 100 -2 100 -d 300 -s 100 -N 4000000) part of the SamTools v1.16 package (Danecek et al 2021) Three groups of ten randomized datasets of 30 million pairs of reads were generated with *P. vulgatus* ATCC 8483 comprising 10%, 1% or 0.1% of the sample. The remaining reads in each dataset were randomly selected but divided equally among the other 14 strains. The datasets were then mapped with Bowtie2 against the marker genes and observed and expected read coverages were evaluated as described above.

Results

Tn mutagenesis enabled the identification of functional MGEs

Computationally predicted mobile genetic elements (MGEs) are common among Bacteroidota genomes, with over 97% of genomes examined having 1 or more predicted type of elements (Figure 2.1A)(Frye et al., 2021). To determine the activity and functionality of these predicted MGEs, we employed an untargeted transposon (Tn) mutagenesis approach. Three prominent gut microbe species with computationally predicted MGEs were selected as donors: *B. thetaiotaomicron* VPI-5482, *P. vulgatus* ATCC 8482, and *B. ovatus* ATCC 8483. Each species has a distinct repertoire of

predicted MGEs including CTns and integrated prophages (Figure 2.1B). Some of these MGEs have been demonstrated to be functional, allowing them to act as positive controls (Frye et al., 2021, Reyes et al., 2013, Campbell et al., 2020).

The first step in the untargeted Tn mutagenesis mobilization method (TMMM) (Figure 2.1C) is to generate Tn libraries using the pSAM INseq vector (Goodman et al., 2009) for the three species. We isolated and pooled ~40,000 independent colonies per Tn library. These Erm^R mutant library pools were then used as donors and mated with five Tet^R Bacteroidota recipients; *B. thetaiotaomicron* VPI-5482, *B. thetaiotaomicron* 3731, *Parabacteroides merdae* ATCC 43184, *B. thetaiotaomicron* VPI-5482 ΔCPS, and *B. uniformis* ATCC 8492. After mating and selective plating, dually resistant Erm^R and Tet^R colonies represent putative MGE transfer events (Figure 2.1C).

The success of the TMMM strategy was mixed. In the mating growth conditions tested here, none of the nine predicted MGEs in *B. thetaiotaomicron* were found to be transferable. However, one *B. ovatus* and one *P. vulgatus* MGE were observed to be mobilizable (Figure 2.1B). Dually resistant colonies were recovered from matings of *B. ovatus* Tn with *B. thetaiotaomicron* VPI-5482 and *B. thetaiotaomicron* VPI-5482 ΔCPS and colony purified. Using arbitrary PCR (Goodman et al., 2009) we found that all strains screened were transconjugants that acquired *Bo*CTn (16/16), a CTn that encodes a vitamin B_{12} transport locus (Frye et al., 2021). Thus, confirming this strategy for identifying MGE transfer events. For the *P. vulgatus* Tn donor library we initially only isolated dually resistant colonies when *B. thetaiotaomicron* VPI-5482 ΔCPS was used as

the recipient. However, all the isolates (16/16) harbored a single newly acquired and uncharacterized *P. vulgatus* CTn (*Pv*CTn).

Fortuitous detection of a putative repressor of PvCTn activity

*Pv*CTn is ~75 kb with 77 total genes including T4SS machinery, a putative ADP-ribosylglycohydrolase, and a Type I restriction-modification system (Figure 2.2A, Table 2.2). Examination of the Tn insertion locations for the *B. thetaiotaomicron* VPI-5482 *Δ*CPS *Pv*CTn transconjugants revealed that they were all inserted into gene *BVU3433*. This gene encodes a putative 149 amino acid helix-turn-helix (HTH) DNA binding domain protein (PF01381). *BVU3433* has extensive Phyre2 predicted structural homology with diverse transcriptional regulators (e.g., PlcR [2QFC] 99.5% confidence, 91% coverage) (Kelley et al. 2015, Declerck et al., 2007). Further, alignments of the 16 arbitrary PCR products identified a total of 4 independent Tn insertion sites within *BVU3433,* all of which are expected to disrupt its function (Figure 2.2B). Together, this suggests that the inactivation of *BVU3433* through pSAM mutagenesis de-represses *Pv*CTn conjugation genes, thereby increasing the frequency of *Pv*CTn mobilization.

To directly test the function of BVU3433, we hypothesized that deleting *BVU3433* from *P. vulgatus* should increase the conjugation efficiency of *Pv*CTn. We first generated a clean deletion of *BVU3433*, followed by marking *Pv*CTn in the wildtype and deletion background with an antibiotic resistance cassette using allelic exchange resulting in *P. vulgatus ΔBVU3433 Pv*CTn:*tetQ* and *P. vulgatus Δtdk Pv*CTn:*tetQ*. We then carried out conjugations with these new donor strains using *B. thetaiotaomicron* VPI-5482

pNBU2_*ermG* and *B. thetaiotaomicron* VPI-5482 *Δ*CPS pNBU2_*ermG* as the recipient(s). The conjugation efficiencies of *Pv*CTn ranged from 1.71 x 10^{-8} – 2.11 x 10^{-7} CFUs/mL for *P. vulgatus Pv*CTn:*tetQ* and 8.77 x 10^{-7} – 1.28 x 10^{-6} CFUs/mL for *P. vulgatus Pv*CTn:*tetQ ΔBVU3433* depending on the recipient used (Figure 2.3A). Even though *P. vulgatus ΔBVU3433* has a longer lag phase compared to WT (Figure 2.9A), there were consistently 4.4 – 51.3-fold higher conjugation efficiencies of the *P. vulgatus ΔBVU3433* mutant. We attempted to complement *BVU3433* in *trans* using pNBU2-*bla-CfxA* however, no significant difference in conjugation efficiency compared to *P. vulgatus* with only pNBU2-*bla-CfxA* was detected ($P = 0.12$)(Figure 2.3B). Regardless, the *P. vulgatus ΔBVU3433* results recapitulate the phenotype detected with the initial Tn mutants.

BVU3433 alters the PvCTn transcriptome

After observing the consistent increase in *Pv*CTn conjugation efficiency of the *BVU3433* mutant we investigated differences in *Pv*CTn WT and *ΔBVU3433* transcription through RNA-Seq. We hypothesized that the increase in the mutant *Pv*CTn conjugation efficiency is due to BVU3433 no longer repressing the putative conjugation genes in *Pv*CTn. Therefore, we expected the conjugation genes in *Pv*CTn to be upregulated in the mutant due to the lack of *BVU3433* repression.

Analysis of the RNA-Seq data revealed that there was indeed an increase in expression throughout *Pv*CTn (Figure 3.4). In fact, 60% of *Pv*CTn genes (44/73) have a 2-fold or greater upregulation in the mutant. This included operons 4 and 5 that encode

the majority of the *Pv*CTn conjugation genes which were expressed an average of 11-fold greater in the *BVU3433* mutant. Using qPCR, we confirmed the upregulation of two candidate operon 5 conjugation genes, one predicted to code for a putative DNA partitioning protein (*BVU3366*) and the other for a conserved protein found in CTns (*BVU3378*). We measured a significant 42 ($P = 1.58 \times 10^{-5}$) and 23-fold ($P = 2.28 \times 10^{-8}$) increase in expression in *BVU3366* and *BVU3378*, respectively, in the mutant when compared to WT. This validates our RNA-Seq results (Figure 2.10).

In addition to genes directly linked to the conjugation apparatus, we identified nine other *Pv*CTn-encoded, putative regulatory genes that were differentially regulated by the absence of *BVU3433* (Figure 2.4, Table 2.4). This includes six putative HTH domain genes: *BVU3415, BVU3420, BVU3426, BVU3429, BVU_RS21735, and BVU3432*. These genes were expressed ~3.5 – 31.6 fold greater in the *BVU3433* mutant when compared to WT. It is likely that one or more of these HTH genes are involved in the regulation of *Pv*CTn however, the increased expression in the mutant suggests that BVU3433 is the primary regulator repressing these genes. The other three non-HTH regulators include an SOS-response transcriptional repressor (*BVU3423*), a HNH endonuclease (*BVU3427*), and a putative bacterial DNA-binding protein (*BVU3428*). Like the HTH genes, the *BVU3433* mutant has increased gene expression for all three, with a fold change increase ranging from 2.6 to 51.3-fold more compared to WT, further indicating the key role of BVU3433 in *Pv*CTn regulation.

To better understand the effect of BVU3433 on conjugation, RNA from two additional samples were sequenced. The first was from *B. thetaiotaomicron* with a

mobilized *Pv*CTn:*tetQ*::Δ*BVU3433* from *P. vulgatus* integrated into its genome. Remarkably, we determined that the expression profile of *Pv*CTn:*tetQ*::Δ*BVU3433* in *B. thetaiotaomicron* was more like that of the WT *P. vulgatus* than the *BVU3433* mutant. If anything, the expression is lower than WT, with 19% of *Pv*CTn genes (14/73) are at least 2-fold more upregulated in than WT. Expression of operons 4, 5, 7, and 8 are not detected in *B. thetaiotaomicron Pv*CTn:*tetQ*::Δ*BVU3433* (Figure 2.4). The lack of conjugation gene expression needed to produce the T4SS structures encoded in operons 4 and 5 suggests that *Pv*CTn:*tetQ*::Δ*BVU3433* in a *B. thetaiotaomicron* recipient likely has reduced conjugation efficiency. Therefore, despite lacking BVU3433, the significant de-repression of genes observed in the mutant *Pv*CTn in one host background (*P. vulgatus*) does not translate to a different host background (*B. thetaiotaomicron*). One possible explanation for this change in expression profile based on the host background may be due to other existing chromosomal or MGE-encoded regulators in *B. thetaiotaomicron* (Figure 2.2B).

The final sample we sequenced for RNA-Seq, unlike the previous RNA-Seq samples from mid-log phase cells growing in liquid TYG medium, was extracted directly from a mixed lawn of WT *P. vulgatus* and *B. thetaiotaomicron* growing on BHI-HB agar. Since conjugation does not generally occur in liquid cultures, we evaluated the expression of *Pv*CTn during relevant conditions on solid medium. We hypothesized that the conjugation genes on operons 4 and 5 will be upregulated due to the increased likelihood of T4SS structures being formed in these conditions. RNA-Seq data of the *B. thetaiotaomicron* and *P. vulgatus* conjugation mass shows that even when only a small

fraction of cells was likely directly conjugating in the mixed population, a slight upregulation of some operon 4 genes and DNA replication genes in operon 8 is observed. However, unlike the WT and mutant *P. vulgatus*, there is a slight downregulation of a cluster of restriction-modification genes in operon 17. Aside from the slight differences in the three operons mentioned, the expression values are like WT *Pv*CTn in *P. vulgatus* (Figure 2.4A).

H_2O_2 decreases BVU3433 expression and increases conjugation efficiency

Due to the observed increase in conjugation and de-repression of *Pv*CTn genes in *P. vulgatus* Δ*BVU3433*, we attempted to identify conditions that might influence the expression of *BVU3433*. We hypothesized that stress may be an important driver for the repression of *BVU3433*, therefore, we tested *BVU3433* expression via qPCR of *P. vulgatus* exposed to various sub-inhibitory stress conditions including H_2O_2, salinity, antibiotics, taurocholic acid, heat, and UV. We discovered a significant ($P = 0.015$) ~4.4-fold decrease in *BVU3433* expression of H_2O_2-exposed *P. vulgatus* when compared to the control group (Figure 2.5A). Considering this decrease in *BVU3433* expression, we tested how H_2O_2-exposure affects conjugation efficiency. We observed an insignificant ($P = 0.42$) 1.73-fold increase in the conjugation efficiency of the H_2O_2-exposed WT *P. vulgatus* (Figure 2.5B) and the *P. vulgatus* Δ*BVU3433* strain followed the same trend with an insignificant ($P = 0.36$) 1.45-fold increase in conjugation efficiency for the H_2O_2-exposed mutant (Figure 2.5B). Together, the qPCR and conjugation efficiency data

suggest that subinhibitory H_2O_2 stress can increase conjugation efficiency. However, since the conjugation efficiency of both WT and Δ*BVU3433 P. vulgatus* increased with H_2O_2 stress, it is possible that H_2O_2 not only influences the expression of BVU3433 but is likely to have additional effects on the cell physiology of the donor and/or recipient cells.

PvCTn-like MGEs are common among other Bacteroidota

To determine the prevalence of BVU3433-mediated regulation among gut microbes, we screened 133 other gut microbial genomes for BVU3433 homologs. This search identified 22 putative BVU3433 homologs, 19 of which were encoded by intact CTns found in eight additional species (Table 2.5). The three remaining homologs reside in genomic loci that appear to have experienced deletions or rearrangements leading to the loss of nearly all genes essential for conjugation. Examination of the shared DNA among the 19 CTns content generated four CTn clusters (Figure 2.6A, Figure 3.7A). Like *Pv*CTn, most encode restriction-modification systems, while others encode putative metal (tellurite) resistance genes (PF02342, PF05099, PF15616), UV protection genes (PF00817, PF11700) and/or protein phosphatases (PF13672).

Despite the divergence in their overall gene content, all 19 CTns share a common integration site at the 3' end of a conserved three gene cluster *rluA-rmuA-ppdK*. Specifically, we identified the *attB* as the 20bp motif 5'-GYS GCN CAR GCK GCH RTH GA-3' within the 3' end of *ppdK* itself (Figure 2.6B). Integration of the CTn produces two imperfect direct repeats (*attL* and *attR)* and in each instance regenerates a

stop codon within three amino acids of the typical *ppdK* stop codon. Our re-examination of the RNA-Seq read data from our novel *B. thetaiotaomicron Pv*CTn:*tetQ*::Δ*BVU3433* transconjugant detected transcriptional readthrough across both the novel *attL* and *attR* sites (Figure 2.11B). These data confirm the preference for this integration location by *Pv*CTn. The conservation of the *attB* is somewhat surprising due to the sequence diversity observed among the predicted integrases (mean = 78.3% ± 9.9% amino acid identity; Figure 2.11C). On further inspection of the *attB* site in the panel of gut microbial genomes we identified another five distinct MGEs of various sizes in 10 genomes that lack *BVU3433* but do encode related integrases (Figure 2.11C).

In addition, we found that all 22 *BVU3433* homologs encode adjacent homologs of *BVU3432* (105AA, *merR*-like HTH domain PF13411) and the small ORF *BVU_RS21835* (63AA, no conserved domains) (Figure 2.7A). The divergent transcriptional organization of these HTH domain proteins is akin to that of *cI* and *cro* in phage λ and similar phages (Degnan et al. 2007). By examining the RNA-Seq transcriptional coverage we were able to identify putative transcriptional start sites and 5' untranslated regions (5'UTRs) for *BVU3433, BVU_RS21835,* and several differentially regulated operons in *Pv*CTn. We subsequently found that the RNA polymerase binding sites (-10 and -35) of *BVU3433* and *BVU_RS21835* occur within the 5'UTR of the other gene owing to the small size of this intergenic region and are highly conserved among the related CTns (Figure 2.7A).

Given the broad conservation of the *BVU3432 – BVU3433* regulatory region among the CTns as well as genes involved in the conjugative apparatus, we

computationally searched for conserved motifs that may be involved in regulating conjugation activity. We examined the upstream regions of five genes differentially expressed in Δ*BVU3433* that represent the first gene in the operon and are conserved in ≥16 of the 19 related CTns. This analysis identified conserved features including RNA polymerase binding sites for all genes as well as potential hairpin structures that might act as transcriptional terminators (Figure 2.7A, Figure 2.11D). The analysis also identified 2 conserved motifs with imperfect dyad symmetry that we observed in all five upstream regions (Figure 2.7B & 2.7C, Figure 2.11D). It is possible that one or both emotifs act as binding sites for BVU3433 or another *Pv*CTn encoded regulator (e.g., BVU3432) and be responsible for activation or repression of the conjugative apparatus.

Detection of PvCTn in human gut metagenomes

Using a marker gene approach, evidence for *Pv*CTn-like elements was detected among Bacteroidota species from both healthy patients and those with colorectal cancer (CRC) in 3 geographically distinct studies (Austria, USA, China) (Figure 2.8A, Figure 2.12A & S4B). We found sequence evidence for *Pv*CTn-like elements in ~90% of all the patient samples analyzed (300/335). Metagenome sequence reads semi-quantitatively represent both the diversity and abundance of species and strains present in a community sample of isolated cells. As such we can conservatively estimate that ~5–14% of Bacteroidota cells encode a *BVU3433* homolog among the patient sample groups screened. We detected similar coverage ranges indicating the integration of *Pv*CTn-like elements (e.g., ~7–12% *attL* and ~2–7% *attR*; Figure 2.8A, Figure 2.12A & S4B). The higher prevalence of the

integrase (~14–19% *BVU3359*) than *BVU3433* is consistent with our identification of MGEs that share a *BVU3359*-like integrase, but do not encode the *BVU3432–BVU3433* regulatory region (Figure 2.12C). Further, despite the previously described differences in community composition including a greater proportion of Bacteroidota cells (Figure 2.8B) and immunological status between the healthy patients and those with colorectal cancer (Feng et al. 2015 , Vogtmann et al. 2016, Yu et al. 2015) we saw no difference in the frequency of *Pv*CTn-like elements after normalizing for Bacteroidota abundance (Figure 2.8A).

Analysis of simulated datasets suggests reliable detection of *Pv*CTn when strains with integrated elements account for ≥1% of the population, however, when using the described marker genes and *Pv*CTn encoding strains approach 0.1% detection becomes unpredictable (~50:50)(Figure 2.12C). It is likely that such unpredictability could be ameliorated by deeper sequencing coverage. The simulated data, like the actual patient data, also detected a slightly lower than expected coverage of *attL* and *attR* sequences (Figure 2.12C). This is possibly due to a read mapping conflict with the sequence similar *attB* region, but in the patient data it may also be affected by the existence of novel *attL* and *attR* junctions not represented in our marker gene dataset (e.g., novel *Pv*CTn-like MGE integration events).

Discussion

Bacterial MGEs are known to be significant drivers of bacterial evolution. Therefore, identifying functional MGEs is important for understanding the distribution and exchange of fitness determinants in bacteria. Our study used an untargeted TMMM as a novel approach for tracking functional MGEs in Bacteroidota (Figure 2.1C). Moreover, this method can be applied to non-Bacteroidota models with established means of Tn mutagenesis including well-known human pathogens (e.g., *Pseudomonas aeruginosa*, *Mycobacterium tuberculosis*, *Vibrio cholerae*) and symbionts (*E. coli*, *Bifidobacterium breve*) (Cain et al. 2020, Dempwolff et al. 2020).

Previous research has established the importance of MGEs in altering bacterial phenotypes, which can have considerable effects on human health (Durrant et al. 2020, Panwar et al. 2023). For instance, the emergence and dissemination of a single carbapenem-hydrolyzing gene (bla_{NDM}) carried by diverse transposable elements have put the effectiveness of carbapenems, broad-spectrum antibiotics used for treating various bacterial infections, at risk on a global scale (Acman et al. 2022). Moreover, virulence genes have been observed to mobilize from pathogenic to non-pathogenic bacteria (Messerer et al. 2017), underscoring the potential of MGEs to enhance pathogenicity. As the acquisition of MGEs can immediately impact bacterial fitness, discovering novel MGEs could aid the development of techniques to modify microbial communities for improved human health.

Constraints of existing approaches for identifying functional MGEs

Computational methods are effective at identifying potential MGEs in bacterial genomes (Akhter et al. 2012, Johansson et al 2021, Ozer et al. 2014, Roux et al. 2015). However, predictions alone cannot confirm the ability of an MGE to mobilize, which often requires targeted mutations to confirm or the reliance on capturing chance transfer events in genomic data (Coyne et al. 2014). Our untargeted MGE capture method bypasses the need for directed mutations (Saak et al. 2020). Although a previous study developed transposon-aided capture (TRACA), which allows for the preferential capture of circular extrachromosomal MGEs (i.e., plasmids) (Jones and Marchesi 2006), it is unlikely to capture elements like *Pv*CTn. On the other hand, entrapment vectors like pBACpAK (Tansirichaiya et al. 2022), if adapted for Bacteroidota hosts, may be able to capture *Pv*CTn. However, this method is reliant upon MGEs inserting into a small region of the recipient pBACpAK vector, which may or may not have sufficient sequence homology for site-specific integrases. In contrast, our method mutagenizes the donor and enables MGEs to integrate at a preferred attachment site within recipient cells.

Tn mutagenesis mobilization method provides insight into CTn regulation

In this study, we successfully identified the mobilization of *Bo*CTn and *Pv*CTn using this Tn mutagenesis method. Curiously, all the *Pv*CTn transconjugants harbored pSAM-mediated disruptions of *BVU3433*. Our subsequent analyses demonstrated a consistent increase in the conjugation efficiency of *P. vulgatus* Δ*BVU3433* mutants compared to WT. While additional Tn insertions likely occurred in T4SS genes or activators of *Pv*CTn mobilization the data suggest that our method may be able to

preferentially capture MGEs by disrupting negative regulators of mobilization. However, this is clearly not always the case, as we successfully captured *Bo*CTn which encodes a vitamin B_{12} transporter (Frye et al. 2021) and the Tn cassette integrated into an uncharacterized gene (*BACOVA0479*). Regardless, the range of MGEs captured for each strain will vary based on the functional and regulatory differences in genes essential for mobilization.

Regulation of mobilization of CTns is frequently subject to muti-layer, tight regulation (Johnson and Grossman 2015, Salyers et al 1995). CTnDOT, a well-characterized CTn in Bacteroidota, requires tetracycline exposure to trigger a four-step process. This process involves seven regulatory proteins and RNAs in its regulatory cascade for the excision and mobilization of CTnDOT to complete (Waters & Salyers 2013). Our results identify the critical role of BVU3433 in repressing many *Pv*CTn genes and as a result significantly reducing mobilization. Yet our findings also indicate that *BVU3433* is likely only the first component of the *Pv*CTn regulatory cascade, as nine other *Pv*CTn-encoded regulatory-related genes were upregulated in the absence of *BVU3433*. However, the specific regulatory mechanisms connecting these proteins and the DNA-binding sites of BVU3433 on *Pv*CTn remain to be elucidated. Given the broad upregulation of *Pv*CTn Δ*BVU3433* genes compared to WT *Pv*CTn (Figure 2.4), we hypothesize that *BVU3433* binds to one or more *Pv*CTn intergenic spacers. Further, we propose that the conserved sequence motifs we identified in the upstream regions of five of the differentially regulated operons represent possible binding sites.

Finally, adding to the complexity of possible regulatory systems within *Pv*CTn, operon 2 of *Pv*CTn contains *BVU3362*, a gene predicted to encode ADP-ribosylglycohydrolase, a class of enzymes commonly involved in post-translational modifications (Mikolčević et al. 2021). In Bacteria, ADP-ribosylation is often linked to the post-translational modification of eukaryotic proteins (Simon et al. 2014). For instance, the Diphtheria toxin produced in *Corynebacterium diphtheriae* inhibits cellular protein synthesis in the host and promotes pathogenicity through the ADP-ribosylation of eukaryotic elongation factor 2 (Bachran et al. 2007). Presently, the regulatory mechanism(s) of BVU3362 within *Pv*CTn is unknown. However, given that ADP-ribosylation has been shown to be involved in *Legionella pneumophila* T4SS effector translocation, it is plausible that BVU3362 may also be involved in the post-translational regulation of the *Pv*CTn T4SS (Amor et al. 2005). This hypothesis will require further investigation to fully understand the role and targets of BVU3362. Overall, this study provides a preliminary characterization of *Pv*CTn's regulatory mechanism, which is likely to be complex and involve additional layers of transcriptional and translational regulation.

Pv*CTn-like elements are diverse and globally distributed*

Our characterization of *Pv*CTn and *BVU3433* enabled our detection of a diverse group of *Pv*CTn-like elements in human gut-associated bacterial genomes. These conjugative transposons all share a common site-specific integrase and attachment site (*attB)* along with the *BVU3432-BVU3433* regulatory region, and genes involved in the

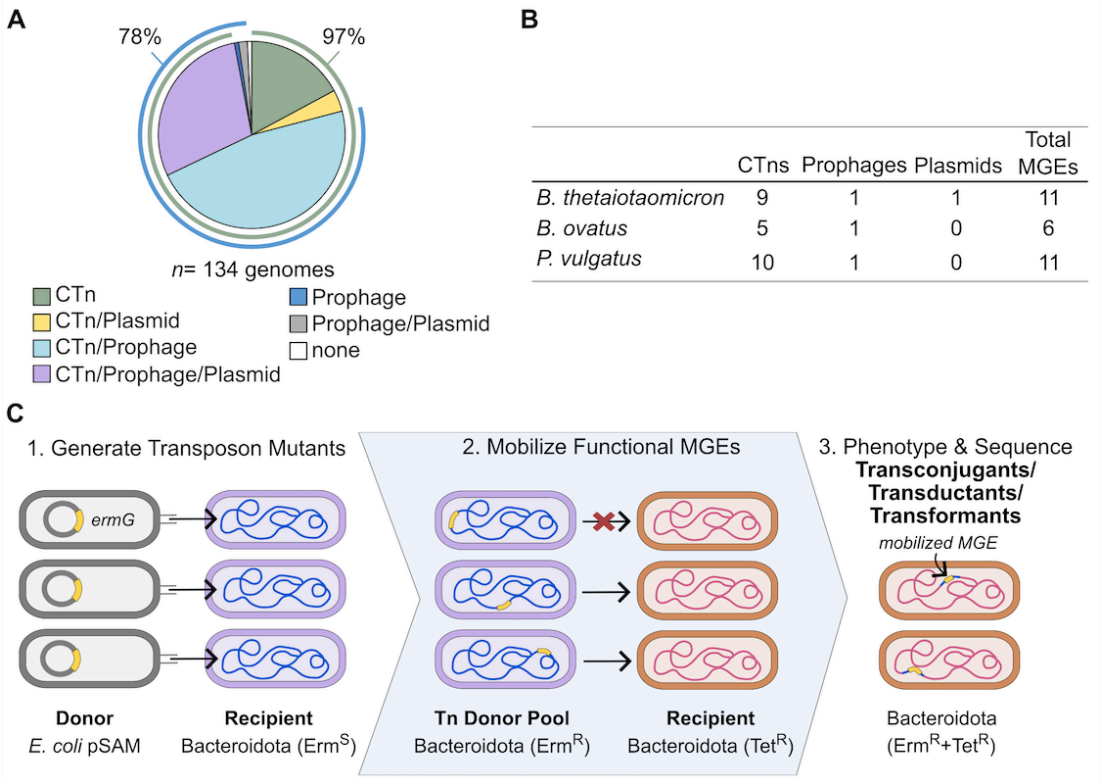
T4SS. While none encode known antibiotic resistance genes, several encode metal and UV resistance genes that may contribute to the fitness of their hosts. Further, virtually all the CTns encode restriction modification systems which may have consequences for further mobile DNA acquisitions by their hosts. Individual bacterial genomes are generally crucial for identifying host-MGE pairs, but they represent only a fraction of species and strain diversity that exists in human guts globally. As such we interrogated the prevalence of *Pv*CTn-like elements in human gut metagenomes (Feng et al. 2015, Vogtmann et al. 2016, Yu et al. 2015), and found they were widespread in ~90% of patient samples. Together these results suggest *Pv*CTn-like elements contribute to genetic and functional diversity of human gut microbes.

Conclusion

In this study, we have successfully demonstrated a new and effective method for capturing MGEs from gut microbes. Although our approach did not detect mobilization of all the predicted MGEs, it expands the currently available methods for MGE identification and offers a potential strategy for numerous other bacteria. Our method enabled us to confirm *Pv*CTn functionality and identify a conserved conjugation repressor protein. Further, *Pv*CTn represents one member of a diverse family of elements that can be detected in patient samples from around the globe. Further, these *Pv*CTn-like elements may indirectly contribute to human health as studies have highlighted the distribution of *P. vulgatus* and its sister species *P. dorei* in the gut microbiome as a critical determinant in the efficacy of immune checkpoint blockade therapy in advanced

melanoma patients (Usyk et al. 2021) and the development of coronary artery disease (Yoshida et al. 2018). And strains of these two species encode 8 of the 20 *Pv*CTn-like elements we identified. Overall, our findings demonstrate the potential of our method for the discovery of novel MGEs and provide insights into the prevalence and distribution of *Pv*CTn-like elements in human gut-associated bacteria.

FIGURES



	CTns	Prophages	Plasmids	Total MGEs
B. thetaiotaomicron	9	1	1	11
B. ovatus	5	1	0	6
P. vulgatus	10	1	0	11

Figure 2.1. MGEs are common in human gut-associated Bacteroidota. (**A**) The proportions of 134 human gut-associated Bacteroidota with one or more of the computationally predicted MGE element classes (prophage, CTn and plasmid). (**B**) Predicted MGEs counts for the three Bacteroidota species used to generate Tn mutant libraries. (**C**) Schematic of untargeted Tn mutagenesis mobilization method (TMMM) for functional MGE detection.

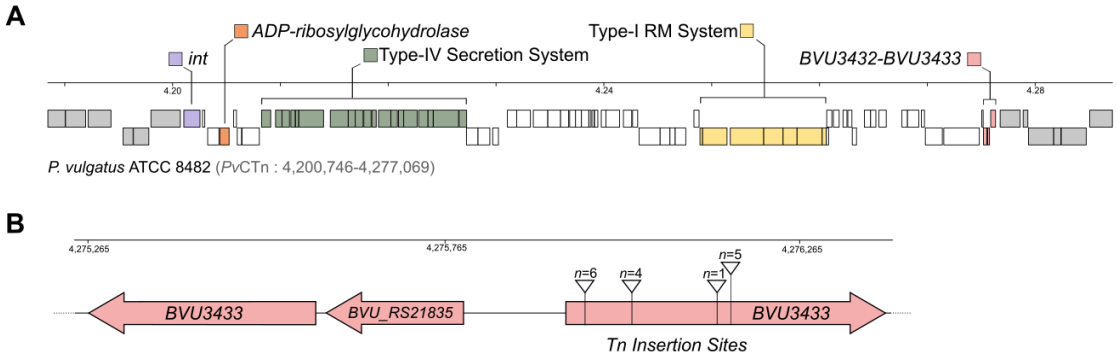


Figure 2.2. Detection of a mobilizable conjugative transposon *Pv*CTn using Tn mutagenesis. (**A**) Scaled diagram of *P. vulgatus* ATCC 8482 genomic region encoding *Pv*CTn and its immediate upstream and downstream regions (genes in grey). Key gene regions described in the text are indicated. Scale bar indicated in megabases (Mb). (**B**) Schematic of the *BVU3432–BVU3433* regulatory region and the four Tn cassette insertion locations identified in *BVU3433* for the 16 screened transconjugants.

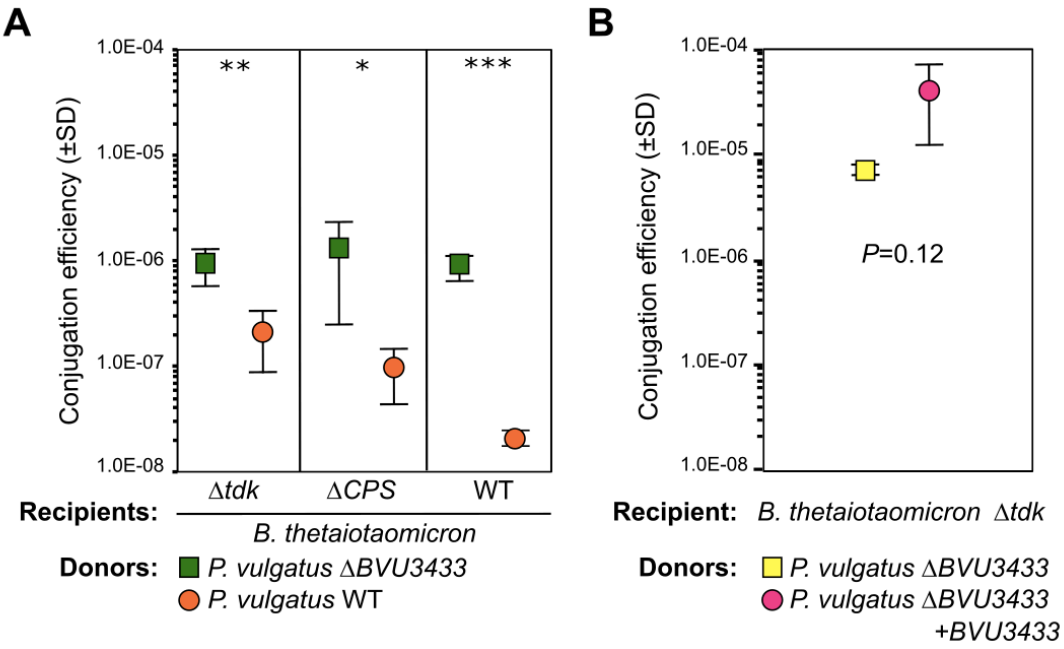


Figure 2.3. Conjugation efficiency of *Pv*CTn is increased in the *BVU3433* mutant. (**A**) Conjugation efficiencies of *Pv*CTn from WT and Δ*BVU3433 P. vulgatus* donor strains were calculated using three *B. thetaiotaomicron* recipient strains with two technical replicates for each conjugation combination. This included mutant strains *B. thetaiotaomicron* Δ*tdk* and *B. thetaiotaomicron* Δ*tdk* ΔCPS and their WT parent *B. thetaiotaomicron* VPI-5482. (**B**) Complementation of *BVU3433* in *trans* using pNBU2-*bla-CfxA* did not restore lower conjugation efficiencies to *ΔBVU3433+BVU3433* donors. Conjugation efficiencies of donor *P. vulgatus ΔBVU3433* with an empty pNBU2-*bla-CfxA* vector or *P. vulgatus ΔBVU3433* pNBU2-*bla-CfxA_BVU3433* were calculated with *B. thetaiotaomicron* Δ*tdk* recipient cells. Conjugation efficiencies were compared between relevant strains using a one-tailed homoscedastic *t*-test significance (***, $P < 0.001$; **, $P < 0.01$; *, $P < .05$). All measures represent means ± standard deviation 2-3 technical replicates.

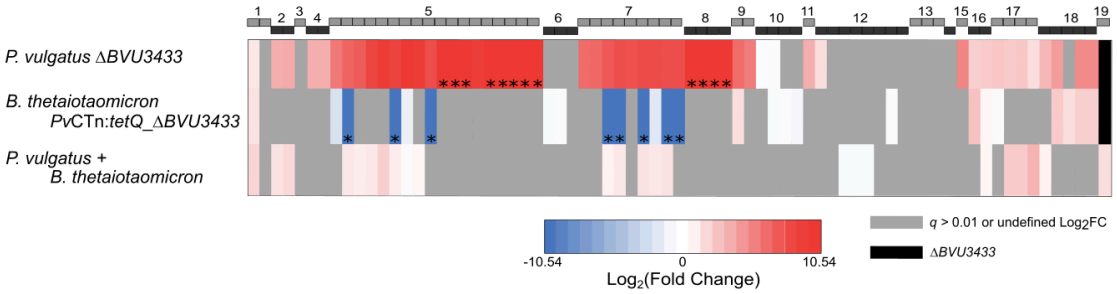


Figure 2.4. Loss of BVU3433 leads to the upregulation of *Pv*CTn genes. Heatmap represents the $\log_2$ fold change ($\log_2$FC) of RNA-Seq gene expression values for *Pv*CTn genes from (i) *P. vulgatus ΔBVU3433*, (ii) *B. thetaiotaomicron Pv*CTn:*tetQ_ΔBVU3433*, and (iii) *P. vulgatus* WT x *B. thetaiotaomicron* when compared to *P. vulgatus* WT (only genes with a *q*-value <.01 are shown, genes with *q*-value > 0.01 and/or undefined $\log_2$FC values are gray). Black bars represent gene deletions. RNA-Seq for all samples was carried out in biological triplicate. The corresponding gene names for each numbered operon are listed in Table 2.2.

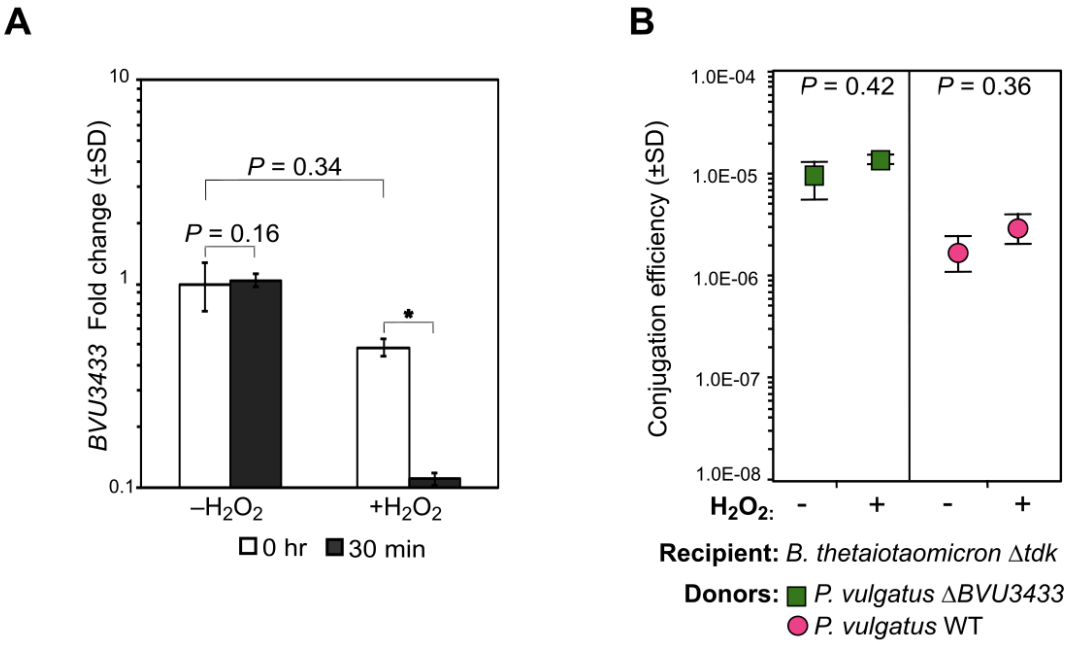


Figure 2.5. H_2O_2 exposure downregulates *BVU3433* expression and increases conjugation efficiency. (**A**) Relative fold change of *BVU3433* was measured in mid-log phase *P. vulgatus* WT treated with H_2O_2 or not immediately prior to exposure and after 30 minutes of incubation at 37ºC. RT-qPCR was performed on RNA extracted from biological triplicate cultures for the two treatment groups. (**B**) Conjugation efficiencies of *Pv*CTn from WT and Δ*BVU3433 P. vulgatus* donor strains were calculated using *B. thetaiotaomicron* Δ*tdk* recipients while exposed to H_2O_2. Conjugation efficiencies were compared between relevant strains using a one-tailed homoscedastic *t*-test significance from 2-3 technical replicates.

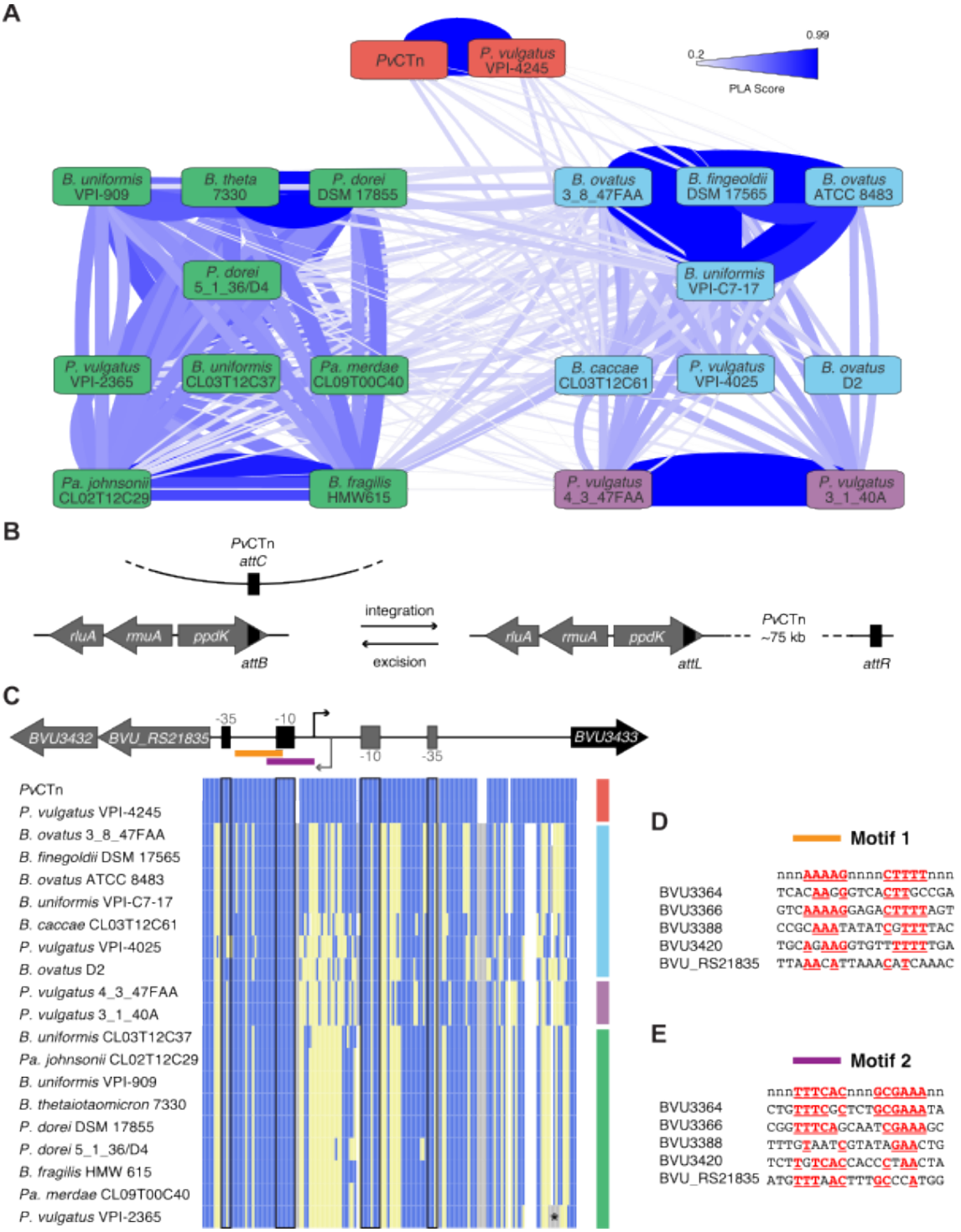

A
PvCTn
P. vulgatus VPI-4245
0.99
0.2
PLA Score
B. uniformis VPI-909
B. theta 7330
P. dorei DSM 17855
B. ovatus 3_8_47FAA
B. fingeoldii DSM 17565
B. ovatus ATCC 8483
P. dorei 5_1_36/D4
B. uniformis VPI-C7-17
P. vulgatus VPI-2365
B. uniformis CL03T12C37
Pa. merdae CL09T00C40
B. caccae CL03T12C61
P. vulgatus VPI-4025
B. ovatus D2
Pa. johnsonii CL02T12C29
B. fragilis HMW615
P. vulgatus 4_3_47FAA
P. vulgatus 3_1_40A
B
PvCTn
attC
rluA
rmuA
ppdK
attB
integration
excision
attL
PvCTn
~75 kb
attR
C
BVU3432
BVU_RS21835
-35
-10
-10
-35
BVU3433
PvCTn
P. vulgatus VPI-4245
B. ovatus 3_8_47FAA
B. finegoldii DSM 17565
B. ovatus ATCC 8483
B. uniformis VPI-C7-17
B. caccae CL03T12C61
P. vulgatus VPI-4025
B. ovatus D2
P. vulgatus 4_3_47FAA
P. vulgatus 3_1_40A
B. uniformis CL03T12C37
Pa. johnsonii CL02T12C29
B. uniformis VPI-909
B. thetaiotaomicron 7330
P. dorei DSM 17855
P. dorei 5_1_36/D4
B. fragilis HMW 615
Pa. merdae CL09T00C40
P. vulgatus VPI-2365
D
Motif 1
nnnAAAAGnnnnCTTTTnnn
BVU3364 TCACAAGGTCACTTGCCGA
BVU3366 GTCAAAAGGAGACTTTTAGT
BVU3388 CCGCAAATATATCGTTTTAC
BVU3420 TGCAGAAGGTGTTTTTTTGA
BVU_RS21835 TTAAACATTAAACATCAAAC
E
Motif 2
nnnTTTCACnnnGCGAAAnn
BVU3364 CTGTTTCGCTCTGCGAAATA
BVU3366 CGGTTTCAGCAATCGAAAGC
BVU3388 TTTGTAATCGTATAGAACTG
BVU3420 TCTTGTCACCACCCTAACTA
BVU_RS21835 ATGTTTAACTTTGCCCATGG

Figure 2.6. Genetic similarities among *Pv*CTn-like MGEs from human gut microbes. (**A**) Network analysis of DNA sequence similarity among BVU3433-homolog encoding CTns from the human gut microbiome. Four clusters of genome encoded CTn-like elements were identified using MCL based on the percentage of length aligned (PLA) scores. Clusters are indicated by the colored nodes labeled with the strain of origin and edge thickness and color correspond to PLA scores shown in the key (20%–99%). (**B**) Schematic of site-specific *Pv*CTn-like MGE insertions in the 3' end of the strain's *ppdK* gene. In each case, a stop codon of *ppdK* is regenerated within 2-3 amino acids of expected location after MGE integration.

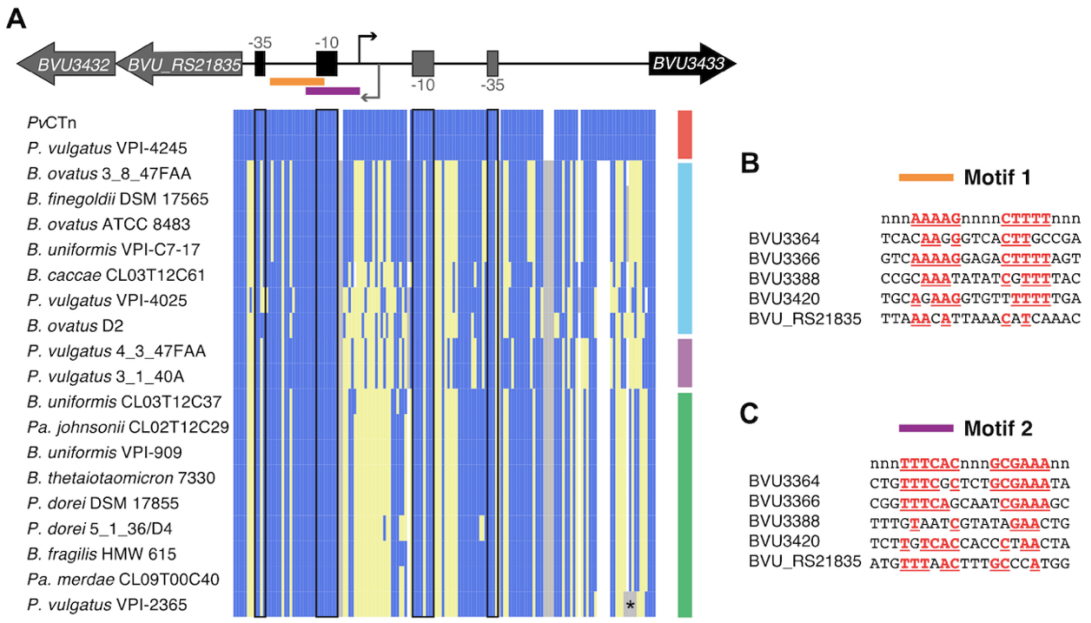


Figure 2.7. *Pv*CTn-like MGEs encode conserved regulatory regions. (**A**) Sequence alignment of *BVU_RS21835–BVU3433* intergenic region from all 20 *Pv*CTn-like MGEs starting and stopping with the first codon of the indicated gene. The overlapping and conserved RNA polymerase binding sites and transcription start sites are indicated for each divergently transcribed gene. Sites in blue correspond to nucleotides identical to *Pv*CTn, yellow represents alternate nucleotides, white indicates gaps in the alignment, and gray represents inserted nucleotides. The asterisk indicates that *P. vulgatus* VPI-2365 has a hypothetical gene inserted at this location and encoded in the same orientation as *BVU3433*. (**B** and **C**) Alignments of two conserved motifs with dyad symmetry identified by MEME and found in the upstream regions of five *Pv*CTn operons that are differentially expressed in the Δ*BVU3433* strain and broadly conserved in *Pv*CTn-like MGEs. Majority rule consensus sequence is shown on the top line in red and matching positions are shown for each upstream region.

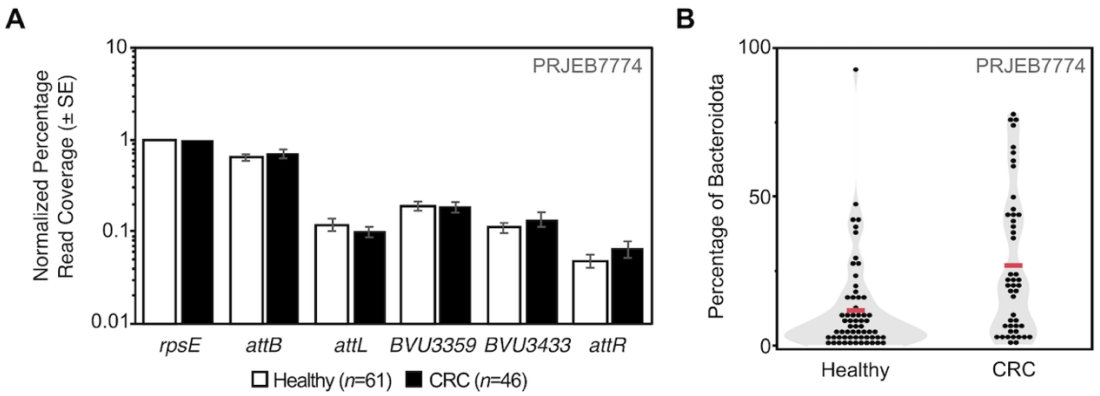


Figure 2.8. Metagenome detection of *Pv*CTn-like elements. (**A**) Metagenomic short reads from a panel of healthy patients and those with colorectal cancer (CRC) were mapped to the indicated marker genes. Graph shows the average read coverage among samples and error bars represent standard error of the mean (SE). Read coverage was normalized for each sample first by the number of quality filtered short reads and then normalized to gut Bacteroidota *rpsE* gene coverage. (**B**) The same data were also subjected to taxonomic classification using Kraken2 to determine the proportion of Bacteroidota per sample. Individual black dots represent proportion of reads classified as Bacteroidota from the total number of reads that were classified per sample. The average proportion of Bacteroidota among each patient cohort is indicated with the red line (Healthy = 11.7% ± 2.0% SE, CRC = 26.8% ± 3.6% SE).

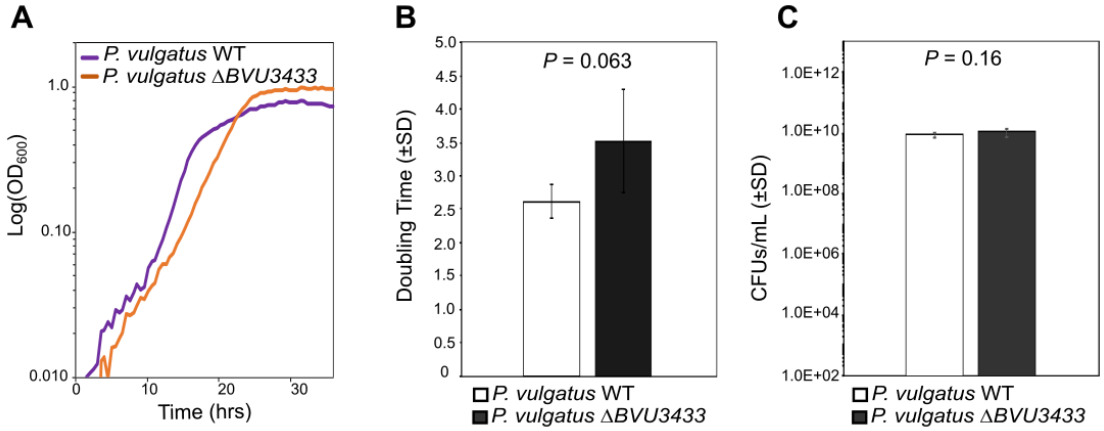


Figure 2.9. *P. vulgatus WT* and *P. vulgatus* ΔBVU3433 growth and doubling times. **(A)** Representative growth curves in TYG media optical densities (OD_{600}) were recorded every 30 minutes. Curves are averages from technical replicates ($n = 3$). **(B)** Doubling times were calculated using the least-squares method for growth between 0.05 and 0.12 OD_{600} ($n = 3$). **(C)** Colony forming units (CFUs) were calculated for the inocula used in **A**. Both **B** and **C** used a one-tailed homoscedastic *t*-test to test significance.

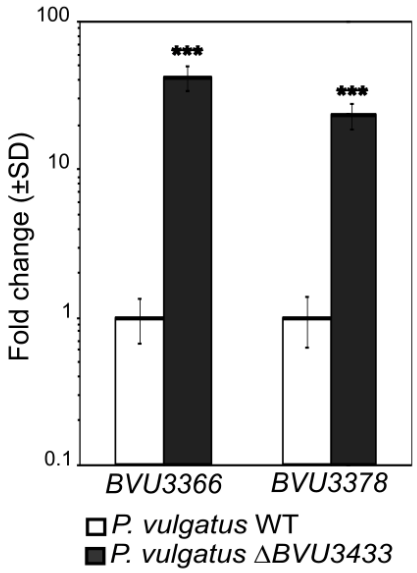


Figure 2.10. Candidate conjugation genes are upregulated in *P. vulgatus ΔBVU3433*. Relative fold change was calculated for representative conjugation genes *BVU3366* and *BVU3378* by RT-qPCR comparing expression in *P. vulgatus* WT versus *P. vulgatus* Δ*BVU3433* during mid-log phase growth in TYG. (***, $P < 0.001$; **, $P < 0.01$; *, $P <$ .05)

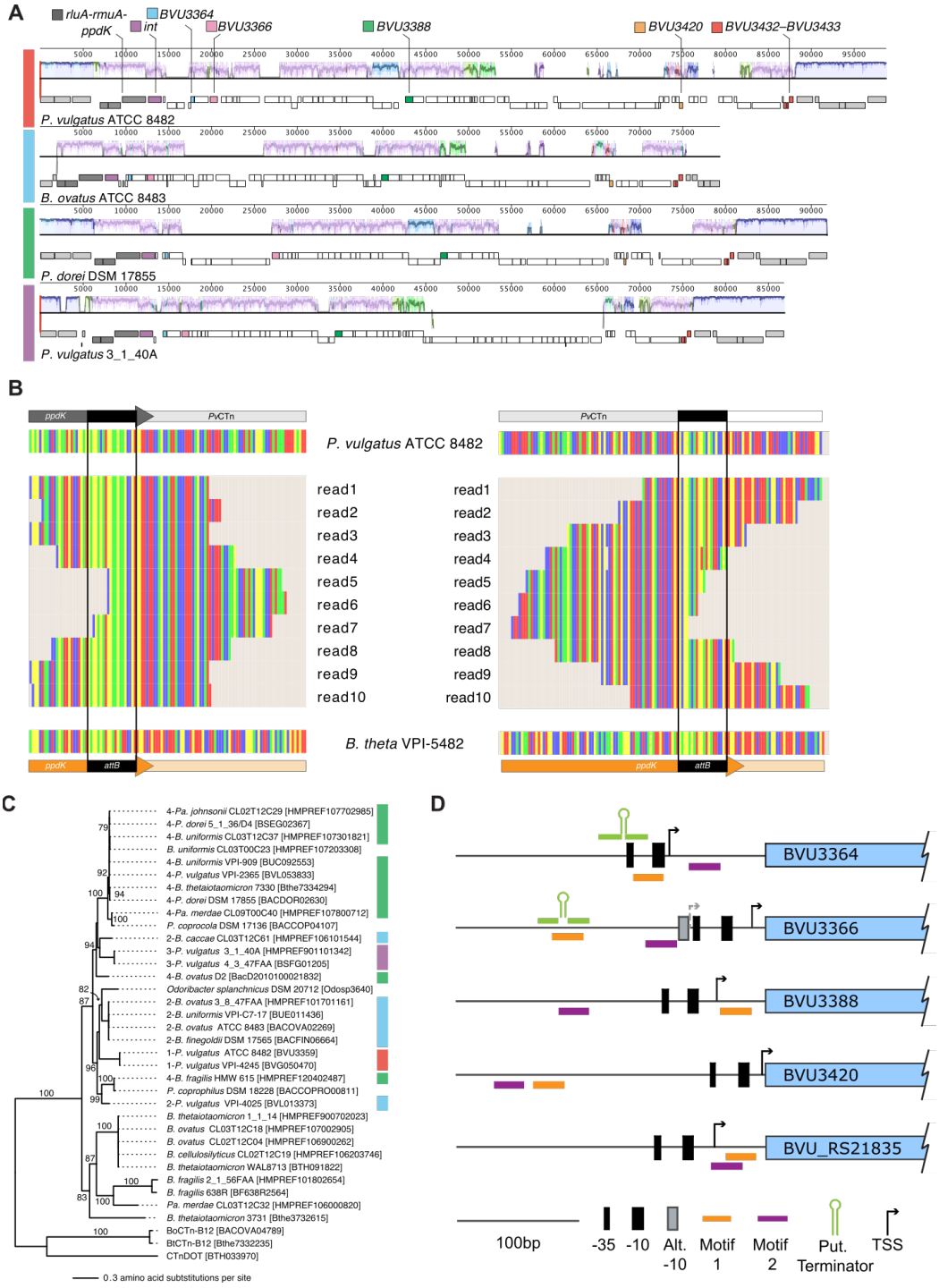

A
rluA-rmuA-ppdK
int
BVU3364
BVU3366
BVU3388
BVU3420
BVU3432–BVU3433
P. vulgatus ATCC 8482
B. ovatus ATCC 8483
P. dorei DSM 17855
P. vulgatus 3_1_40A
B
ppdK
PvCTn
attB
P. vulgatus ATCC 8482
B. theta VPI-5482
read1
read2
read3
read4
read5
read6
read7
read8
read9
read10
C
4-Pa. johnsonii CL02T12C29 [HMPREF107702985]
4-P. dorei 5_1_36/D4 [BSEG02367]
4-B. uniformis CL03T12C37 [HMPREF107301821]
B. uniformis CL03T00C23 [HMPREF107203308]
4-B. uniformis VPI-909 [BUC092553]
4-P. vulgatus VPI-2365 [BVL053833]
4-B. thetaiotaomicron 7330 [Bthe7334294]
4-P. dorei DSM 17855 [BACDOR02630]
4-Pa. merdae CL09T00C40 [HMPREF107800712]
P. coprocola DSM 17136 [BACCOP04107]
2-B. caccae CL03T12C61 [HMPREF106101544]
3-P. vulgatus 3_1_40A [HMPREF901101342]
3-P. vulgatus 4_3_47FAA [BSFG01205]
4-B. ovatus D2 [BacD2010100021832]
Odoribacter splanchnicus DSM 20712 [Odosp3640]
2-B. ovatus 3_8_47FAA [HMPREF101701161]
2-B. uniformis VPI-C7-17 [BUE011436]
2-B. ovatus ATCC 8483 [BACOVA02269]
2-B. finegoldii DSM 17565 [BACFIN06664]
1-P. vulgatus ATCC 8482 [BVU3359]
1-P. vulgatus VPI-4245 [BVG050470]
4-B. fragilis HMW 615 [HMPREF120402487]
P. coprophilus DSM 18228 [BACCOPRO00811]
2-P. vulgatus VPI-4025 [BVL013373]
B. thetaiotaomicron 1_1_14 [HMPREF900702023]
B. ovatus CL03T12C18 [HMPREF107002905]
B. ovatus CL02T12C04 [HMPREF106900262]
B. cellulosilyticus CL02T12C19 [HMPREF106203746]
B. thetaiotaomicron WAL8713 [BTH091822]
B. fragilis 2_1_56FAA [HMPREF101802654]
B. fragilis 638R [BF638R2564]
Pa. merdae CL03T12C32 [HMPREF106000820]
B. thetaiotaomicron 3731 [Bthe3732615]
BoCTn-B12 [BACOVA04789]
BtCTn-B12 [Bthe7332235]
CTnDOT [BTH033970]
0.3 amino acid subtstitutions per site
D
BVU3364
BVU3366
BVU3388
BVU3420
BVU_RS21835
100bp
-35
-10
Alt. -10
Motif 1
Motif 2
Put. Terminator
TSS

Figure 2.11. Structural and functional conservation of *Pv*CTn-like MGEs from human gut microbes. (**A**) Sequence alignment of representative integrated *Pv*CTn-like MGEs from the four identified clusters. Nucleotide similarity plots are shown above each scaled representation of genes present on the leading and lagging strands. Genes flanking each integrated MGE are shown in gray with the *attB* insertion site in *ppdK* shown in dark gray. The integrase gene and the five differentially expressed and conserved genes are marked with their corresponding colors. (**B**) Alignments of RNAseq short reads confirm integration of *Pv*CTn in the *ppdK* gene of a *B. thetaiotaomicron* VPI-5482 transconjugant. Ten representative reads from the *attL* and *attR* region of *Pv*CTn span the indicated attachment site and continue into the *ppdK* chromosomal region of *B. thetaiotaomicron* and not that of the donor *P. vulgatus*. (**C**) Maximum likelihood phylogenetic reconstruction of *ppdK-attB* inserted MGE integrase proteins. Despite the most divergent ingroup sequences having as little as 42% amino acid identity, all appear to site specifically integrate into *ppdK*. Integrases from a *Pv*CTn-like MGE are marked with a colored bar that corresponds to clusters in Figure 2.6. FastTree estimated bootstrap values ≥50% are shown adjacent to the appropriate node. Integrases from CTnDOT and vitamin B_{12} transporter encoding CTns were used as outgroups. (**D**) Diagrams of the upstream regions of the five differentially expressed and conserved genes with identified conserved regulatory motifs. Size and spacing of features indicated in the key are to scale.

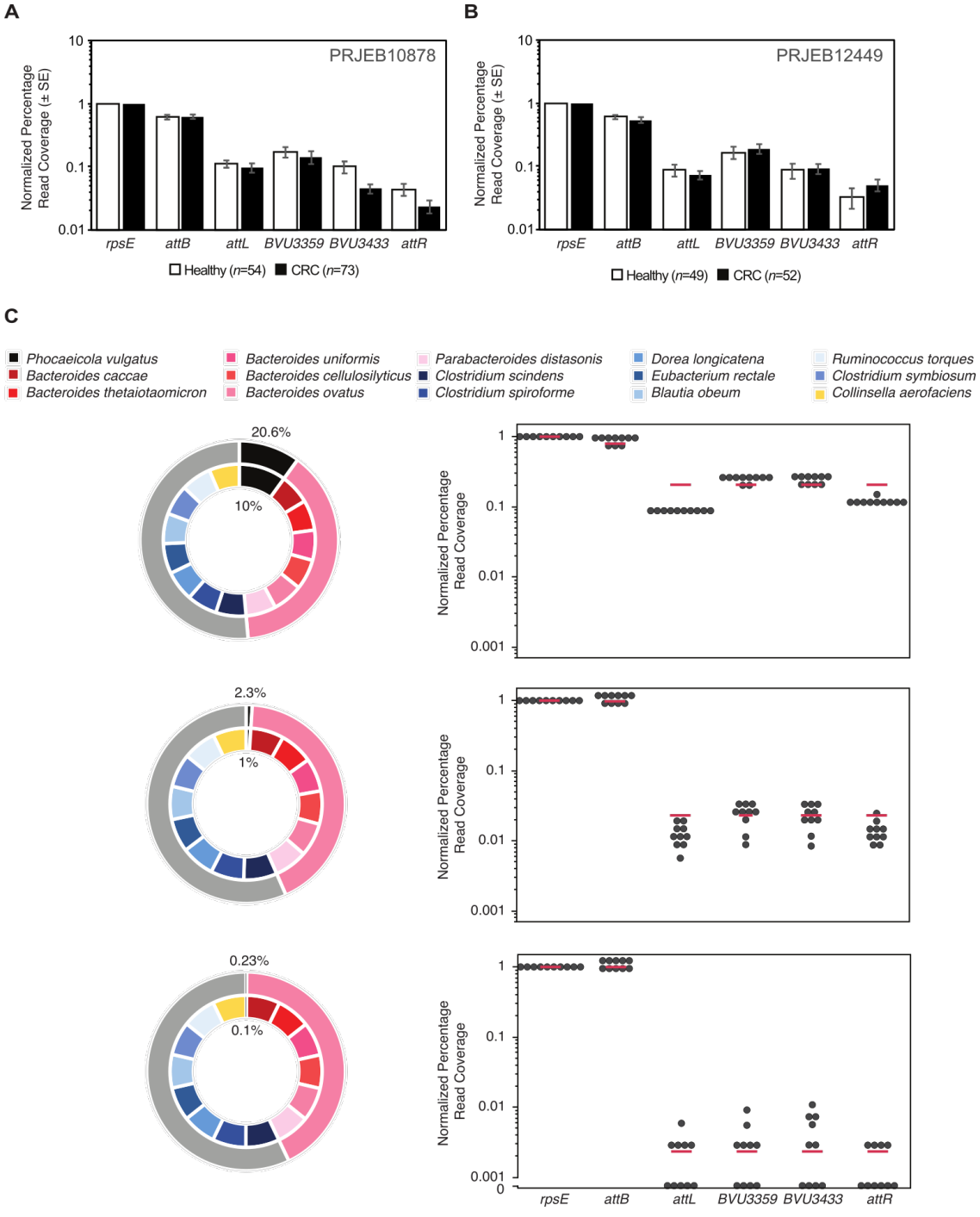

A
PRJEB10878
Normalized Percentage Read Coverage (± SE)
10
1
0.1
0.01
rpsE
attB
attL
BVU3359
BVU3433
attR
Healthy (n=54)
CRC (n=73)
B
PRJEB12449
Normalized Percentage Read Coverage (± SE)
Healthy (n=49)
CRC (n=52)
C
Phocaeicola vulgatus
Bacteroides caccae
Bacteroides thetaiotaomicron
Bacteroides uniformis
Bacteroides cellulosilyticus
Bacteroides ovatus
Parabacteroides distasonis
Clostridium scindens
Clostridium spiroforme
Dorea longicatena
Eubacterium rectale
Blautia obeum
Ruminococcus torques
Clostridium symbiosum
Collinsella aerofaciens
20.6%
10%
2.3%
1%
0.23%
0.1%
Normalized Percentage Read Coverage
1
0.1
0.01
0.001
0

Figure 2.12. Metagenome detection of *Pv*CTn-like elements in clinical and simulated datasets. (**A, B**) Metagenomic short reads from two additional panels of healthy patients and those with colorectal cancer (CRC) were mapped to the indicated marker genes. Graphs show read coverage normalized first by number of quality filtered short reads and then normalized to gut Bacteroidota *rpsE* gene coverage. Standard error bars are shown. (**C**) Simulated read data were generated from 15 indicated genomes, then randomly selected for three groups of 10 datasets with 10%, 1% or 0.1% *P. vulglatus* represented. Inner circles of doughnut plots indicate the proportion of each species according to the key and outer circles indicate the proportion of *P. vulglatus* relative to the other Bacteroidota present in the sample (e.g., expected *rpsE* read coverage). Plots at right show normalized read coverage detected for each marker gene from each randomized dataset as a black point. Expected coverage is shown for each marker as a red line. Genomes for the following bacterial strains were used *P. vulgatus* ATCC 8482, *B. caccae* ATCC 43185, *B. thetaiotaomicron* VPI-5482, *B. uniformis* ATCC 8492, *B. cellulosilyticus* WH2, *B. ovatus* SD_CC_2a, *Pa. distasonis* ATCC 8503, *Clostridium scindens* ATCC 35704, *C. spiroforme* DSM 1552, *Dorea longicatena* DSM 1381, *Eubacterium rectale* ATCC 33656, *Blatuia obeum* ATCC 29174, *Ruminococcus torques* ATCC 27756, *C. symbiosum* ATCC 14940, *Collinsella aerofaciens* ATCC 25986.

TABLES

Table 2.1 Bacterial Strains, Plasmids, and Primers Used in this Study

Strain Name	Short name in text	Notes	Source and/or reference
E. coli S17-1λpir			Simon *et al*. 1983
B. thetaiotaomicron VPI-5482 *Δtdk*	*B. thetaiotaomicron*	"wild-type" deletion background; Used to generate *B. thetaiotaomicron* Tn mutant library	Eggerth and Gagnon, 1933
Bacteroides ovatus ATCC 8483	*B. ovatus*	Used to generate *B. ovatus* Tn mutant library	Eggerth and Gagnon, 1933
Phocaeicola vulgatus ATCC 8482 Δtdk	*P. vulgatus*	"wild-type" deletion background; Used to generate P. vulgatus Tn mutant library and *BVU3433* deletion	Eggerth and Gagnon, 1933
Bacteroides thetaiotaomicron VPI5482 pNBU2_tetQ	*B. thetaiotaomicron VPI-5482* TetR	Recipient strain to screen Tn libraries for MGEs, TetR	This study
Bacteroides thetaiotaomicron 3731	*B. thetaiotaomicron* 3731 TetR	Recipient strain to screen Tn libraries for MGEs, Naturally TetR	Meng *et al*. 2015
Parabacteroides merdae ATCC 43184	*Parabacteroides merdae 43184* TetR	Recipient strain to screen Tn libraries for MGEs, Naturally TetR	
Bacteroides thetaiotaomicron VPI-5482 Δtdk ΔCPS pNBU2_tetQ	*B. thetaiotaomicron* ΔCPS TetR	Recipient strain to screen Tn libraries for MGEs, TetR	This study
Bacteroides uniformis ATCC 8492 pNBU2_tetQ	*B. uniformis* TetR	Recipient strain to screen Tn libraries for MGEs, TetR	This study
Bacteroides thetaiotaomicron VPI-5482 Δtdk pNBU2_tetQ BoCTn*:pSAM_ermG*		Transconjugant from *B. ovatus* Tn library & *B. thetaiotaomicron* conjugation	This study
Bacteroides thetaiotaomicron VPI-5482 Δtdk ΔCPSpNBU2_tetQ PvCTn:*pSAM_ermG*		Transconjugant from *P. vulgatus* Tn library & B. thetaiotaomicron ΔCPS conjugation	This study
*Phocaeicola vulgatus ATCC 8482 Δtdk Pv*CTn*:ΔBVU3433*	*P. vulgatus ΔBVU3433*		This study
*Phocaeicola vulgatus ATCC 8482 Δtdk Pv*CTn*:tetQ*	*P. vulgatus Pv*CTn*:tetQ*	Used as donor strain to calculate conjugation efficiency, TetR	This study
*Phocaeicola vulgatus ATCC 8482 Δtdk Pv*CTn*:tetQ::ΔBVU3433*	*P. vulagatus ΔBVU3433 Pv*CTn*:tetQ*	Used as donor strain to calculate conjugation efficiency, TetR	This study
Bacteroides thetaiotaomicron VPI-5482 *Δtdk pNBU2_ermG*		Used as recipient strain to calculate conjugation efficiency, ErmR	Frye et al. 2022
Bacteroides thetaiotaomicron VPI-5482 pNBU2_ermG		"True wild-type" background with *tdk* intact; Used as recipient strain to calculate conjugation efficiency, ErmR	This study
Bacteroides thetaiotaomicron VPI-5482 Δtdk ΔCPS pNBU2_ermG		"wild-type" *tdk* deletion background with all 8 CPS loci deleted; used as recipient Strain to calculate conjugation efficiency, ErmR	This study
Bacteroides thetaiotaomicron Δtdk *Pv*CTn*:tetQ::ΔBVU3433*		Transconjugant from P. vulgatus ΔBVU3433 PvCTn:tetQ & B. thetaiotaomicron *pNBU2_ermG* conjugation	This study
*Phocaeicola vulgatus ATCC 8482 Δtdk Pv*CTn*:tetQ::ΔBVU3433 pNBU2-bla-CfxA_BVU3433*		*pNBU2_CfxA_BVU3433* complement	This study
*Phocaeicola vulgatus ATCC 8482 Δtdk Pv*CTn*:tetQ::ΔBVU3433 pNBU2-bla-CfxA*		pNBU2_CfxA negative control	This study

Plasmid name		**Notes**	
pSAM_BC2_ermG		Barcoded donor used to generate *B. thetaiotaomicron* Tn library; ermR	Cullen et al. 2015
pSAM_BC3_ermG		Barcoded donor used to generate *B. ovatus* library ; ErmR	Cullen et al. 2015
pSAM_BC4_ermG		Barcoded donor used to generate *P. vulgatus* library ; ErmR	Cullen et al. 2015
pExchange-tdk_Bv		Counterselectable homologous exchange vector for strain construction; AmpR ErmR FUdRS tdk	Campbell *et al*. 2020
pLGB13		Suicide vector used for allelic replacement in Bacteroidota, ErmR and contains aTC-inducible ss-Bfe1 counterselection	Leonor García-Bayona & Comstock 2019
pNBU2-bla-ermGb		Integrates cloned fragments to NBU2 att1 and/or att2 sites, AmpR ErmR	Koropatkin *et al*. 2008
pNBU2-bla-tetQ		Integrates cloned fragments to NBU2 att1 and/or att2 sites, AmpR TetR	Martens *et al*. 2008
pNBU2-bla-CfxA		Integrates cloned fragments to NBU2 att1 and/or att2 sites, AmpR CfxR	Campbell *et al*. 2020
Primer name		**Sequence (5' - 3')***	**Source and/or reference**
Primers for pExchange-tdk_Bv *BVU3433* deletion			
NotI_BVU3433_LF_5'		TTAgcggccgcTGGGAGGATGACGAGGTTA	This work
SpeI_BVU3433_RF_3'		GACactagtCAGCCGGGATACAATTTGAC	This work
BVU3433_LF_SOE_3'		AGGCAACTATGGTTAATACTATGGTTACATAGCATCCGAACGAAATTTTTAAT	This work
BVU3433_RF_SOE_5'		ATTAAAAATTTCGTTCGGATGCTATGTAACCATAGTATTAACCATAGTTGCCT	This work
Primers for cloning SpeI & XmaI restriction sites on a PvCTn intergenic spacer (BVU3419-BVU3420)			
XbaI_Addition_IS4_LF_5'		TGGtctagaGCTCGGTTGATAGGCTTTGG	This work
SalI_Addition_IS4_RF_3'		TTCgtcgacGGCTGTCGCGCAGTTTCTTC	This work
Addition_IS4_LF_SOE_3'_SpeI_XmaI		CCCAATTTGGCACAAATTGTGAAactagtCCGCcccgggAGGCTGCCGGAAAGTACTCC	This work
Addition_IS4_RF_SOE_5'_SpeI_XmaI		GGAGTACTTTCCGGCAGCCTcccgggGCGGactagtTTCACAATTTGTGCCAAATTGGG	This work
Primers for cloning TetR cassete (*tetQ*) onto SpeI & XmaI restriction sites			
tetQ_pNBU2_R_SpeI		GGCactagtTTATTTTGATGACATTGATTTTTTGGAACATG	This work
tetQ_pNBU2_F3_SpeI		AATactagtACGCGTTATCTCCTTAACGTAC	This work
Primers for amplifying *BVU3433 for pNBU2_CfxA* complementation			
pNBU2_BVU3433_5'_LF_BamHI		AGGggatccAGGACTCAGGTTTACTCTACGAAACT	This work

pNBU2_BVU3433_3'_RF_Not I		AGGgcggccgcAGGCATCGTTGAAAA GTTGGC	This work
Primers for RT-qPCR (16S, 3433, 3368, 3378)			
BT 16S rRNA 5		GGTAGTCCACACAGTAAACGATGA A	
BT 16S rRNA 3		CCCGTCAATTCCTTTGAGTTTC	
BVU3433_F1		AGCAGCAGTGTCCATTCATG	
BVU3433_R1		CCTCCAGACGTTTGATCTCG	
BVU3366_F1		CTTTGCCAACCAGAAAGGAG	
BVU3366_R1		TTGTTGGGTTCGTATGACCA	
BVU3378_F1		GTACGGCCCTCAAACAGAAG	
BVU3378_R1		GTACCCCGTGGTAATCATGC	
AmpR, ampicillin resistant; ErmR, erythromycin resistant; FUdRR, FUdR resistant; FUdRS, FUdR sensitive; CfxR, cefoxitin resistant			
*Restriction sites are underlined and lower-case			

Table 2.2 *P. vulgatus* ATCC8482 *Pv*CTn annotation

Start	Stop	Strand	aa_length	Gene	Synonym	COG	Product	Type	nt_length	Operon #	Phage description	PFAM	PFAM_Description
4201098	4202597	+	499	-	BVU3359	-	transposase	CDS	1500	1		PF13102	Phage integrase SAM-like domain
4202835	4203041	+	68	-	BVU3360	-	hypothetical protein	CDS	207	1			
4203311	4204354	-	347	-	BVU3361	-	hypothetical protein	CDS	1044	2			
4204436	4205296	-	286	-	BVU3362	COG1397O	ADP-ribosylglycohydrolase	CDS	861	2		PF03747	ADP-ribosylglycohydrolase
4205787	4206047	-	86	-	BVU3363	-	hypothetical protein	CDS	261	3			
4206054	4206449	+	131	-	BVU3364	-	hypothetical protein	CDS	396	4			
4206503	4208098	+	531	-	BVU3365	-	putative mobilization protein	CDS	1596	4		PF03432	Relaxase/Mobilisation nuclease domain
4208308	4209156	+	282	parA	BVU3366	-	putative ParA-related protein	CDS	849	5		PF01656, PF09140, PF03205, PF02374, PF07015	CobQ/CobB/MinD/ParA nucleotide binding domain , ATPase MipZ , Molybdopterin guanine dinucleotide synthesis prot , Anion-transporting ATPase , VirC1 protein
4209607	4210197	+	196	-	BVU3367	-	hypothetical protein	CDS	591	5			
4210206	4210895	+	229	-	BVU3368	-	hypothetical protein	CDS	690	5			
4211023	4211421	+	132	traE	BVU3369	-	conserved protein found in conjugative transposon TraE	CDS	399	5		PF13572	Domain of unknown function (DUF4134)
4211451	4211741	+	96	-	BVU3370	-	hypothetical protein	CDS	291	5			
4211758	4214058	+	766	traG1	BVU3371	COG3451U	conserved hypothetical protein found in conjugative transposon TraG	CDS	2301	5		PF12846	AAA-like domain
4214158	4214308	+	0	-	BVU_19m	-	RF02012 group-II-D1D4-7	misc_RNA	151	5			
4214682	4216337	+	551	-	BVU3372	COG3344L,COG2801L	putative reverse transcriptase	CDS	1656	5		PF00078, PF13655, PF08388	Reverse transcriptase (RNA-dependent DNA polymera , N-terminal domain of reverse transcriptase , Group II intron; maturase-specific domain
4216340	4216999	+	219	traG2	BVU3373	COG3451U	conserved hypothetical protein found in conjugative transposon TraG	CDS	660	5		PF12846	AAA-like domain
4216413	4216489	+	0	-	BVU_20m	-	RF00029 Intron_gpII	misc_RNA	77	5			
4216987	4217724	+	245	-	BVU3374	-	hypothetical protein	CDS	738	5			

4218541	4218942	+	133	-	BVU3376	-	hypothetical protein	CDS	402	5			
4219152	4220291	+	379	traJ	BVU3377	-	conserved hypothetical protein found in conjugative transposon TraJ	CDS	1140	5			
4220310	4220906	+	198	traK	BVU3378	-	conserved protein found in conjugative transposon TraK	CDS	597	5			
4221008	4221499	+	163	-	BVU3379	-	hypothetical protein	CDS	492	5			
4221496	4222734	+	412	traM	BVU3380	-	conserved protein found in conjugative transposon TraM	CDS	1239	5		PF12508	Protein of unknown function (DUF3714)
4222870	4223742	+	290	traN	BVU3381	-	conserved protein found in conjugative transposon TraN	CDS	873	5		PF13595	Domain of unknown function (DUF4138)
4223746	4224219	+	157	-	BVU3382	-	hypothetical protein	CDS	474	5			
4224242	4226578	+	778	virD4	BVU3383	COG3505U	putative mobilization protein	CDS	2337	5		PF02534, PF12696, PF10412	Type IV secretory system Conjugative DNA transfer , TraM recognition site of TraD and TraG , Type IV secretion-system coupling protein DNA-bind
4226633	4227328	+	231	traX	BVU3384	-	putative conjugative transposon gene TraX	CDS	696	5		PF05857	TraX protein
4227312	4228385	-	357	-	BVU3385	-	hypothetical protein	CDS	1074	6			
4228385	4229479	-	364	-	BVU3386	-	putative ATPase involved in DNA repair	CDS	1095	6			
4229749	4230276	-	175	-	BVU3387	-	hypothetical protein	CDS	528	6			
4231106	4231963	+	285	-	BVU3388	-	hypothetical protein	CDS	858	7		PF13351, PF13101	Protein of unknown function (DUF4099), Protein of unknown function (DUF3945)
4231976	4233244	+	422	-	BVU3389	COG0249L	putative DNA mismatch repair protein mutS	CDS	1269	7		PF01624	MutS domain I
4233512	4234837	+	441	-	BVU3390	COG4227L,COG4643S	putative DNA primase	CDS	1326	7		PF08401	Domain of unknown function (DUF1738)
4234854	4236038	+	394	-	BVU3391	-	hypothetical protein	CDS	1185	7		PF13154, PF13155	Protein of unknown function (DUF3991), Toprim-like
4236043	4236786	+	247	-	BVU3392	-	hypothetical protein	CDS	744	7			
4236884	4237423	+	179	exc	BVU3393	COG0550L	putative DNA topoisomerase I	CDS	540	7			
4237420	4237980	+	186	-	BVU3394	-	hypothetical protein	CDS	561	7		PF02757	YLP motif
4237977	4238654	+	225	-	BVU3395	-	hypothetical protein	CDS	678	7	3rd insertion site		
4238682	4238858	+	58	-	BVU3396	-	hypothetical protein	CDS	177	7		PF07432	Histone H1-like protein Hc1
4238953	4239150	+	65	-	BVU3397	-	hypothetical protein	CDS	198	8			

4239195	4239497	+	100	-	BVU3398	-	hypothetical protein	CDS	303	8			
4239816	4240247	+	143	-	BVU3399	-	hypothetical protein	CDS	432	8		PF14058	PcfK-like protein
4240244	4241518	+	424	-	BVU3400	-	hypothetical protein	CDS	1275	8		PF14284	PcfJ-like protein
4241870	4242634	+	254	-	BVU3401	COG1192D	ATPase involved in chromosome partitioning	CDS	765	9		PF01656, PF09140, PF06564, PF02374, PF13614	CobQ/CobB/MinD/ParA nucleotide binding domain , ATPase MipZ , YhjQ protein , Anion-transporting ATPase , AAA domain
4242720	4243193	+	157	-	BVU3402	-	hypothetical protein	CDS	474	9	1st Insertion Site		
4243322	4245262	-	646	-	BVU3403	-	conserved hypothetical protein, putative ribose phosphate pyrophosphokinase	CDS	1941	10			
4245259	4246176	-	305	-	BVU3404	-	hypothetical protein	CDS	918	10			
4246179	4246676	-	165	-	BVU3405	COG0546R	putative phosphatase	CDS	498	10		PF13419, PF13242, PF00702, PF03031	Haloacid dehalogenase-like hydrolase , HAD-hyrolase-like , haloacid dehalogenase-like hydrolase , NLI interacting factor-like phosphatase
4246673	4247653	-	326	-	BVU3406	COG0758LU	putative DNA processing enzyme DprA	CDS	981	10		PF02481	DNA recombination-mediator protein A
4248411	4248884	+	157	-	BVU3407	-	hypothetical protein	CDS	474	11	89% nt identical to BVU3402 +us region		
4248931	4248973	+	0	-	BVU_288tt	-	RNIE terminator	TransTerm	43				
4248971	4249204	-	77	-	BVU3408	-	hypothetical protein	CDS	234	12			
4249215	4251476	-	753	-	BVU3409	COG0507L,COG3385L	ATP-dependent exonuclease V, alpha subunit - helicase superfamily I member	CDS	2262	12		PF13604, PF14490, PF13538, PF13245, PF14520, PF05127	AAA domain , Helix-hairpin-helix containing domain , UvrD-like helicase C-terminal domain , Part of AAA domain , Helix-hairpin-helix domain , Helicase
4251782	4254865	-	1027	-	BVU3410	COG0610V	type I restriction enzyme HsdR, putative	CDS	3084	12		PF04851, PF04313, PF00270	Type III restriction enzyme; res subunit , Type I restriction enzyme R

													protein N terminus (H , DEAD/DEAH box helicase
4254872	4256677	-	601	-	BVU3411	COG1479S	hypothetical protein	CDS	1806	12		PF03235, PF07510	Protein of unknown function DUF262 , Protein of unknown function (DUF1524)
4256679	4257971	-	430	-	BVU3412	COG0732V	type I restriction-modification system S subunit	CDS	1293	12		PF01420, PF01420	Type I restriction modification DNA specificity d , Type I restriction modification DNA specificity d
4257976	4260291	-	771	-	BVU3413	COG0286V	type I restriction-modification system M subunit	CDS	2316	12		PF02384, PF12161, PF13659, PF01170	N-6 DNA Methylase , HsdM N-terminal domain , Methyltransferase domain , Putative RNA methylase family UPF0020
4260298	4260684	-	128	-	BVU3414	-	hypothetical protein	CDS	387	12			
4260689	4260898	-	69	-	BVU3415	-	Putative Helix-turn-helix domain	CDS	210	12		PF13443, PF01381	Cro/C1-type HTH DNA-binding domain , Helix-turn-helix
4260726	4261202	+	158	-	BVU3416	-	hypothetical protein	CDS	477	13	89% nt identical to BVU3417		
4261345	4261878	+	177	-	BVU3417	-	hypothetical protein	CDS	534	13	4th insertion site		
4262238	4262600	+	120	-	BVU3418	-	hypothetical protein	CDS	363	13			
4262685	4263062	+	125	-	BVU3419	-	hypothetical protein	CDS	378	14	2nd Insertion site		
4263096	4263488	-	130	-	BVU3420	-	Putative Helix-turn-helix domain	CDS	393	15		PF12728	Helix-turn-helix domain
4264216	4264785	+	189	-	BVU3421	-	putative ribose phosphate pyrophosphokinase	CDS	570	16			
4264788	4265285	+	165	-	BVU3422	COG4474S	hypothetical protein	CDS	498	16		PF06908	Protein of unknown function (DUF1273)
4265569	4266246	+	225	-	BVU3423	COG1974KT	putative SOS-response transcriptional repressor	CDS	678	17		PF00717	Peptidase S24-like
4267699	4268196	+	165	-	BVU3426	-	Putative Helix-turn-helix domain	CDS	498	17		PF01381, PF13560, PF12844, PF13443	Helix-turn-helix , Helix-turn-helix domain , Helix-turn-helix domain , Cro/C1-type HTH DNA-binding domain
4268206	4269261	+	351	-	BVU3427	-	HNH endonuclease	CDS	1056	17		PF13392	HNH endonuclease

4269504	4269788	+	94	-	BVU3428	COG0776L	DNA-binding protein HU	CDS	285	17		PF00216	Bacterial DNA-binding protein
4269839	4269879	-	0	-	BVU_289tt	-	RNIE terminator	TransTerm	41				
4269849	4269889	+	0	-	BVU_290tt	-	RNIE terminator	TransTerm	41				
4269875	4271446	-	523	-	BVU3429	COG2865K	Putative Helix-turn-helix domain	CDS	1572	18		PF04326, PF13749, PF08279, PF08220, PF13412	Divergent AAA domain , ATP-dependent DNA helicase recG C-terminal , HTH domain , DeoR-like helix-turn-helix domain , Winged helix-turn-helix DNA-binding
4271584	4274874	-	1096	-	BVU3430	COG0358L	DNA primase	CDS	3291	18		PF01807, PF08275, PF13662	CHC2 zinc finger , DNA primase catalytic core; N-terminal domain , Toprim domain
4275074	4274871	-			BVU_RS21735		Putative Helix-turn-helix domain	CDS		18		PF12728	HTH_17 Helix-turn-helix domain
4275265	4275582	-	105	-	BVU3432	-	Putative Helix-turn-helix domain	CDS	318	18		PF13411	MerR HTH family regulatory protein
4275790	4275599	-			BVU_RS21835		hypothetical protein	CDS		18			
4275934	4276383	+	149	-	BVU3433	-	Putative Helix-turn-helix domain	CDS	450	19		PF01381	Helix-turn-helix

Table 2.3. Genomes Screened for the Presence of *Pv*CTn-like MGEs

Organism	RefSeq Accession	BVU3433 homologs	Metagenome Marker Gene Mapping
Alistipes finegoldii DSM 17242	NC_018011		X
Alistipes indistinctus YIT 12060	NZ_ADLD00000000		X
Alistipes putredinis DSM 17216	NZ_ABFK00000000		X
Alistipes shahii WAL 8301	NC_021030		X
Alistipes sp. HGB5	NZ_AENZ00000000		
Alistipes sp. JC136	NZ_CAEG00000000		X
Bacteroides caccae ATCC 43185	NZ_AAVM00000000		X
Bacteroides caccae CL03T12C61	NZ_AGXF00000000	HMPREF106101608	
Bacteroides cellulosilyticus CL02T12C19	NZ_AGXG00000000		
Bacteroides cellulosilyticus DSM 14838	NZ_ACCH00000000		
Bacteroides cellulosilyticus WH2	NZ_ATFI00000000		X
Bacteroides clarus YIT 12056	NZ_AFBM00000000		X
Bacteroides coprocola DSM 17136	NZ_ABIY00000000		X
Bacteroides coprophilus DSM 18228	NZ_ACBW00000000	BACCOPRO00036	X
Bacteroides dorei 5_1_36/D4	NZ_ACDI00000000	BSEG02450	
Bacteroides dorei 9_1_42FAA	NZ_ACAA00000000		
Bacteroides dorei CL02T00C15	NZ_AGXH00000000		X
Bacteroides dorei CL02T12C06	NZ_AGXJ00000000		
Bacteroides dorei CL03T12C01	NZ_AGXI00000000		
Bacteroides dorei DSM 17855	NZ_ABWZ00000000	BACDOR02565	
Bacteroides eggerthii 1_2_48FAA	NZ_ACWG00000000		X
Bacteroides eggerthii DSM 20697	NZ_ABVO00000000		
Bacteroides faecis MAJ27	NZ_AGDG00000000		X
Bacteroides finegoldii CL09T03C10	NZ_AGXW00000000		X
Bacteroides finegoldii DSM 17565	NZ_ABXI00000000	BACFIN06730	
Bacteroides fluxus YIT 12057	NZ_AFBN00000000		X
Bacteroides fragilis 2_1_16	NZ_ACPP00000000		

Bacteroides fragilis 2_1_56FAA	NZ_ACWI00000000		
Bacteroides fragilis 3_1_12	NZ_ABZX00000000		
Bacteroides fragilis 3_2_5	NZ_ACIB00000000		
Bacteroides fragilis 638R	NC_016776		
Bacteroides fragilis CL03T00C08	NZ_AGXK00000000		
Bacteroides fragilis CL03T12C07	NZ_AGXL00000000		
Bacteroides fragilis CL05T00C42	NZ_AGXO00000000		
Bacteroides fragilis CL05T12C13	NZ_AGXP00000000		
Bacteroides fragilis CL07T00C01	NZ_AGXM00000000		
Bacteroides fragilis CL07T12C05	NZ_AGXN00000000		
Bacteroides fragilis HMW 610	NZ_AGXQ00000000		
Bacteroides fragilis HMW 615	NZ_AGXR00000000	HMPREF120402563	
Bacteroides fragilis HMW 616	NZ_ALOB00000000		X
Bacteroides fragilis NCTC 9343	NC_003228		X
Bacteroides fragilis YCH46	NC_006347		
Bacteroides helcogenes P 36-108	NC_014933		X
Bacteroides intestinalis DSM 17393	NZ_ABJL00000000		X
Bacteroides nordii CL02T12C05	NZ_AGXS00000000		X
Bacteroides oleiciplenus YIT 12058	NZ_ADLF00000000		X
Bacteroides ovatus 3_1_23	NZ_ACRS00000000		
Bacteroides ovatus 3_8_47FAA	NZ_ACWH00000000	HMPREF101701100	
Bacteroides ovatus ATCC 8483	NZ_AAXF00000000	BACOVA02352	
Bacteroides ovatus CL02T12C04	NZ_AGXT00000000		X
Bacteroides ovatus CL03T12C18	NZ_AGXU00000000		
Bacteroides ovatus SD CC 2a	NZ_ADMP00000000		
Bacteroides ovatus SD CMC 3f	NZ_ADMO00000000		
Bacteroides plebeius DSM 17135	NZ_ABQC00000000		X
Bacteroides salanitronis DSM 18170	NC_015164		X
Bacteroides salyersiae CL02T12C01	NZ_AGXV00000000		X

Bacteroides sp. 2_1_33B	NZ_ACPR00000000		
Bacteroides sp. 2_1_7	NZ_ABZY00000000		
Bacteroides sp. 2_2_4	NZ_ABZZ00000000		
Bacteroides sp. 20_3	NZ_ACRQ00000000		
*Bacteroides s*p. 3_1_19	NZ_ADCJ00000000		
Bacteroides sp. 3_1_33FAA	NZ_ACPS00000000		
Bacteroides sp. CF50	NC_022526		X
Bacteroides ovatus D2	NZ_ACGA00000000	BacD2010100022192	
Bacteroides sp. D20	NZ_ACPT00000000		
Bacteroides sp. D22	NZ_ADCK00000000		
Bacteroides stercoris ATCC 43183	NZ_ABFZ00000000		x
Bacteroides thetaiotaomicron 1_1_6	NZ_ACIC00000000		
Bacteroides thetaiotaomicron 1_1_14	NZ_ACRP00000000		
Bacteroides thetaiotaomicron 3731	NC_Bthetaiotaomicron3731		
Bacteroides thetaiotaomicron 7330	NC_Bthetaiotaomicron7330	Bthe7334223	
Bacteroides thetaiotaomicron VPI-C11-15	SAMN18350527		
Bacteroides thetaiotaomicron VPI-0633-1	SAMN18350524		
Bacteroides thetaiotaomicron VPI-0940-1	SAMN18350525		
Bacteroides thetaiotaomicron VPI-5482	NC_004703		X
Bacteroides thetaiotaomicron VPI-J19-343	SAMN18350526		
Bacteroides uniformis VPI-61	SAMN18350528		
Bacteroides uniformis VPI-909	SAMN18350529	BUC092479	
Bacteroides uniformis VPI-C20-25	SAMN18350530		
Bacteroides uniformis VPI-C7-17	SAMN18350531	BUE011375	
Bacteroides dorei VPI-2277^{o}	SAMN18350535		
Bacteroides dorei VPI-3776A^{o}	SAMN18350536		
Bacteroides nordii WAL7936*	SAMN18350534		
Bacteroides uniformis 4_1_36	NZ_ACTC00000000		
Bacteroides uniformis ATCC 8492	NZ_AAYH00000000		X
Bacteroides uniformis CL03T00C23	NZ_AGXX00000000	HMPREF107203335	

Bacteroides uniformis CL03T12C37	NZ_AGXY00000000	HMPREF107301754	
Bacteroides uniformis str. VPI-R5-33	SAMN18350532		
Bacteroides vulgatus 3_1_40A	NZ_ACRT00000000	HMPREF901101412	
Bacteroides vulgatus 4_3_47FAA	NZ_ACDR00000000	BSFG01134	
Bacteroides vulgatus ATCC 8482	NC_009614	BVU3433	
Bacteroides vulgatus CL09T03C04	NZ_AGXZ00000000		X
Bacteroides vulgatus PC510	NZ_ADKO00000000		
Bacteroides vulgatus str. VPI-2365	SAMN14525940	BVL053895	
Bacteroides vulgatus str. VPI-4025	SAMN14525944	BVL013309	
Bacteroides vulgatus str. VPI-4496.2	SAMN14525945		
Bacteroides vulgatus str. VPI-5710	SAMN14525942		
Bacteroides dorei str. VPI-6598B	SAMN18350537		
Bacteroides vulgatus VPI-4245	SAMN14525939	BVG051495	
Bacteroides vulgatus VPI-4506	SAMN14525946		
Bacteroides vulgatus VPI-6186	SAMN14525941		
Bacteroides WH2	NC_BWH2		
Bacteroides xylanisolvens 1_1_30	NZ_ADCL00000000		
Bacteroides xylanisolvens 2_1_22	NZ_ACPQ00000000		
Bacteroides xylanisolvens CL03T12C04	NZ_AGXE00000000		
Bacteroides xylanisolvens D1	NZ_ACAB00000000		
Bacteroides xylanisolvens SD CC 1b	NZ_ADKP00000000		X
Bacteroides xylanisolvens XB1A	NC_BxylanisolvensXB1A		
Barnesiella intestinihominis YIT 11860	NZ_ADLE00000000		
Dysgonomonas gadei ATCC BAA-286	NZ_ADLV00000000		X
*Dysgonomonas mossii D*SM 22836	NZ_ADLW00000000		X
Odoribacter laneus YIT 12061	NZ_ADMC00000000		X
Odoribacter splanchnicus DSM 20712	NC_015160	Odosp3613	X
Parabacteroides distasonis ATCC 8503	NC_009615		X

Parabacteroides distasonis CL03T12C09	NZ_AGZM00000000		
Parabacteroides distasonis CL09T03C24	NZ_AGZN00000000		
Parabacteroides goldsteinii CL02T12C30	NZ_AGZO00000000		X
Parabacteroides johnsonii CL02T12C29	NZ_AGZP00000000	HMPREF107702926	
Parabacteroides johnsonii DSM 18315	NZ_ABYH00000000		X
Parabacteroides merdae ATCC 43184	NZ_AAXE00000000		X
Parabacteroides merdae CL03T12C32	NZ_AGZQ00000000		
Parabacteroides merdae CL09T00C40	NZ_AJPU00000000	HMPREF107800645	
Parabacteroides sp. D13	NZ_ACPW00000000		
Parabacteroides sp. D25	NZ_ACUA00000000		
Paraprevotella clara YIT 11840	NZ_AFFY00000000		X
Paraprevotella xylaniphila YIT 11841	NZ_AFBR00000000		X
Porphyromonas gingivalis ATCC 33277	NC_010729		X
Porphyromonas gingivalis TDC60	NC_015571		
Porphyromonas gingivalis W83	NC_002950		
Prevotella copri DSM 18205	NZ_ACBX00000000		X
Prevotella salivae DSM 15606	NZ_AEQO00000000		X
Prevotella stercorea DSM 18206	NZ_AFZZ00000000		X
Tannerella forsythia ATCC 43037	NC_016610		X
Tannerella sp. 6_1_58FAA_CT1	NZ_ACWX00000000		X
* Originally deposited in the VPI strain collection as *B. uniformis*			
º Originally deposited in VPI strain collection as *B. vulgatus*			

Table 2.4A. RNAseq data generated for Bacteroidota species, a transconjugant, and conjugation mass used in this study

Table S4 RNAseq data generated for Bacteroidota species, a transconjugant, and conjugation mass used in this study

Sample	Sequencer	Chemistry	Read Length (bp)	Raw Reads	Quality Filtered Read Pairs	Mapped to *P. vulgatus* ATCC 8482	% reads mapped	Mapped to *B. thetaiotaomicron* VPI-5482	% reads mapped	Reads mapped to *Pv*CTn	% reads mapped
P. vulgatus ATCC 8482 Δtdk - I	NextSeq500	NextSeq 500/550 Mid Output Kit v2.5	2x75	10,097,723	8886439	8045065	90.53%	n/a	n/a	4707	0.05%
P. vulgatus ATCC 8482 Δtdk - II	NextSeq500	NextSeq 500/550 Mid Output Kit v2.5	2x75	30,161,318	27600884	25489270	92.35%	n/a	n/a	9493	0.03%
P. vulgatus ATCC 8482 Δtdk - III	NextSeq500	NextSeq 500/550 Mid Output Kit v2.5	2x75	17,500,257	15910528	15224174	95.69%	n/a	n/a	6452	0.04%
P. vulgatus ATCC 8482 Δtdk ΔBVU3433 - I	NextSeq500	NextSeq 500/550 Mid Output Kit v2.5	2x75	6,119,937	5177945	4561126	88.09%	n/a	n/a	266350	5.14%
P. vulgatus ATCC 8482 Δtdk ΔBVU3433 - II	NextSeq500	NextSeq 500/550 Mid Output Kit v2.5	2x75	27,471,788	25254452	23845294	94.42%	n/a	n/a	1447882	5.73%
P. vulgatus ATCC 8482 Δtdk ΔBVU3433 - III	NextSeq500	NextSeq 500/550 Mid Output Kit v2.5	2x75	15,293,186	13912474	12576060	90.39%	n/a	n/a	825308	5.93%
B. thetaiotaomicron VPI-5482 *Pv*CTn:*tetQ_ΔBVU3433* - I	NextSeq500	NextSeq 500/550 Mid Output Kit v2.5	2x75	5,430,184	4712565	201288	4.27%	4192502	88.96%	3667	0.08%
B. thetaiotaomicron VPI-5482 *Pv*CTn:*tetQ_ΔBVU3433* - II	NextSeq500	NextSeq 500/550 Mid Output Kit v2.5	2x75	10,389,722	9572865	559727	5.85%	8848453	92.43%	6518	0.07%
B. thetaiotaomicron VPI-5482 *Pv*CTn:*tetQ_ΔBVU3433* - III	NextSeq500	NextSeq 500/550 Mid Output Kit v2.5	2x75	15,075,277	13751137	1957476	14.24%	12947228	94.15%	4015	0.03%
P. vulgatus pNBU2_*tetQ* + *B. thetaiotaomicron* pNBU2_*ermG* - I	NextSeq500	NextSeq 500/550 Mid Output Kit v2.5	2x75	5,708,649	4856309	4023585	82.85%	1128385	23.24%	1334	0.03%
P. vulgatus pNBU2_*tetQ* + *B. thetaiotaomicron* pNBU2_*ermG* - II	NextSeq500	NextSeq 500/550 Mid Output Kit v2.5	2x75	7,796,486	7185141	5869262	81.69%	1460588	20.33%	2805	0.04%
P. vulgatus pNBU2_*tetQ* + *B. thetaiotaomicron* pNBU2_*ermG* - III	NextSeq500	NextSeq 500/550 Mid Output Kit v2.5	2x75	11,529,792	10671344	8934369	83.72%	2360197	22.12%	3692	0.03%

Table 2.4B. RNASeq Differential Expression

Log2 fold change compared to *P. vulgatus* WT

Synonym	Product	*P. vulgatus* Δ*BVU3433*	*B. thetaiotaomicron* Δtdk pNBU2_erm IS4_*Pv*CTn:*tetQ*_Δ*BVU3433*	*P. vulgatus* & *B. thetaiotaomicron* conjugation
BVU3359	transposase	**-1**	**-1.169925001**	**-1.807354922**
BVU3360	hypothetical protein	**undefined**	0	-2
BVU3361	hypothetical protein	**-3.502500341**	0.333214679	**-1.075288127**
BVU3362	ADP-ribosylglycohydrolase	**-3.560714954**	-0.2175914351	**-1.468553009**
BVU3363	hypothetical protein	**undefined**	undefined	undefined
BVU3364	hypothetical protein	**-3.459431619**	0.5440680444	*B. vulgatus* Δ*tdk* expression only
BVU3365	putative mobilization protein	**-3.459431619**	0.4771212547	-0.4150374993
BVU3366	putative ParA-related protein	**-5.793115555**	**2.164352856**	-0.7349879447
BVU3367	hypothetical protein	**-6.95419631**	***B. vulgatus* Δ*tdk* expression only**	**-0.8365012677**
BVU3368	hypothetical protein	**-7.286722764**	*B. vulgatus* Δ*tdk* expression only	**-0.7884958948**
BVU3369	conserved protein found in conjugative transposon TraE	**-8.796039609**	*B. vulgatus* Δ*tdk* expression only	**-0.8395353278**
BVU3370	hypothetical protein	**-9.909893084**	*B. vulgatus* Δ*tdk* expression only	**-1.906890596**
BVU3371	conserved hypothetical protein found in conjugative transposon TraG	**-9.05166212**	***B. vulgatus* Δ*tdk* expression only**	**-0.7983661388**
BVU3372	putative reverse transcriptase	**-10.54259223**	**1.174830607**	**0.3379013556**
BVU3373	conserved hypothetical protein found in conjugative transposon TraG	**-9.847840356**	*B. vulgatus* Δ*tdk* expression only	**-0.3785116233**
BVU3374	hypothetical protein	**-8.131856961**	***B. vulgatus* Δ*tdk* expression only**	-0.07400058144
BVU3376	hypothetical protein	***P. vulgatus* Δ*BVU3433* expression only**	*B. vulgatus* Δ*tdk* expression only	-0.7369655942
BVU3377	conserved hypothetical protein found in conjugative transposon TraJ	***P. vulgatus* Δ*BVU3433* expression only**	*B. vulgatus* Δ*tdk* expression only	-2
BVU3378	conserved protein found in conjugative transposon TraK	***P. vulgatus* Δ*BVU3433* expression only**	*B. vulgatus* Δ*tdk* expression only	-3.169925001
BVU3379	hypothetical protein	**-9.603626345**	*B. vulgatus* Δ*tdk* expression only	0.3521825181
BVU3380	conserved protein found in conjugative transposon TraM	***P. vulgatus* Δ*BVU3433* expression only**	*B. vulgatus* Δ*tdk* expression only	*B. vulgatus* Δ*tdk* expression only
BVU3381	conserved protein found in conjugative transposon TraN	***P. vulgatus* Δ*BVU3433* expression only**	*B. vulgatus* Δ*tdk* expression only	0.1760912591
BVU3382	hypothetical protein	***P. vulgatus* Δ*BVU3433* expression only**	*B. vulgatus* Δ*tdk* expression only	0
BVU3383	putative mobilization protein	***P. vulgatus* Δ*BVU3433* expression only**	0.7781512504	0.07918124605
BVU3384	putative conjugative transposon gene TraX	***P. vulgatus* Δ*BVU3433* expression only**	-1	-1
BVU3385	hypothetical protein	-0.8365012677	**0.4206398988**	0.03342375549
BVU3386	putative ATPase involved in DNA repair	-1	**0.2725414579**	-0.1834461411

BVU3387	hypothetical protein	-0.6374299206	0.1513763712	0.312929219
BVU3388	hypothetical protein	**-6.598560104**	*B. vulgatus* Δ*tdk* expression only	-0.1269121125
BVU3389	putative DNA mismatch repair protein mutS	**-6.64647689**	*B. vulgatus* Δ*tdk* expression only	0.1337482574
BVU3390	putative DNA primase	**-7.553629294**	***B. vulgatus* Δ*tdk* expression only**	**-0.5849625007**
BVU3391	hypothetical protein	**-8.173926932**	***B. vulgatus* Δ*tdk* expression only**	**-0.896164189**
BVU3392	hypothetical protein	**-7.857980995**	*B. vulgatus* Δ*tdk* expression only	-0.6040713237
BVU3393	putative DNA topoisomerase I	**-8.294620749**	***B. vulgatus* Δ*tdk* expression only**	**-1.04580369**
BVU3394	hypothetical protein	**-8.282625134**	**1.602059991**	**-0.8278190246**
BVU3395	hypothetical protein	**-8.228016971**	***B. vulgatus* Δ*tdk* expression only**	**-1.173648087**
BVU3396	hypothetical protein	**-8.083213368**	***B. vulgatus* Δ*tdk* expression only**	-0.251538767
BVU3397	hypothetical protein	*P. vulgatus* Δ*BVU3433* expression only	undefined	*P. vulgatus* & *B. thetaiotaomicron* conjugation expression only
BVU3398	hypothetical protein	*P. vulgatus* Δ*BVU3433* expression only	undefined	*P. vulgatus* & *B. thetaiotaomicron* conjugation expression only
BVU3399	hypothetical protein	*P. vulgatus* Δ*BVU3433* expression only	*B. vulgatus* Δ*tdk* expression only	-2
BVU3400	hypothetical protein	*P. vulgatus* Δ*BVU3433* expression only	undefined	*P. vulgatus* & *B. thetaiotaomicron* conjugation expression only
BVU3401	ATPase involved in chromosome partitioning	**-6.607330314**	**-1.309855263**	-0.1963972128
BVU3402	hypothetical protein	**-5.707819249**	-0.1255308821	0.1492980825
BVU3403	conserved hypothetical protein, putative ribose phosphate pyrophosphokinase	**0.09691001301**	0.1283184773	0.07716595481
BVU3404	hypothetical protein	**0.234083206**	**0.4002812245**	0.1623029746
BVU3405	putative phosphatase	0.06694678963	**0.2956349638**	0.086848687
BVU3406	putative DNA processing enzyme DprA	0.02996322338	**0.4658402443**	0.06567877564
BVU3407	hypothetical protein	**-3.459431619**	*B. vulgatus* Δ*tdk* expression only	-2.459431619
BVU3408	hypothetical protein	**-1.321928095**	-0.2775339755	0.2174839442
BVU3409	ATP-dependent exonuclease V, alpha subunit - helicase superfamily I member	-1.050626073	-0.1059334447	0.313619123
BVU3410	type I restriction enzyme HsdR, putative	-1.080919995	0.05642012303	**0.4432078774**
BVU3411	hypothetical protein	-0.9349049718	0.1570800924	**0.432209518**
BVU3412	type I restriction-modification system S subunit	-0.8566358248	0.1386934265	**0.4187389307**
BVU3413	type I restriction-modification system M subunit	-1.201633861	0.1080525908	0.235451926
BVU3414	hypothetical protein	-1.724892762	**0.299406463**	0.343725891
BVU3415	Putative Helix-turn-helix domain	-1.787270676	0.2425641867	0.2199563959

BVU3416	hypothetical protein	**undefined**	undefined	*P. vulgatus* & *B. thetaiotaomicron* conjugation expression only
BVU3417	hypothetical protein	**undefined**	undefined	undefined
BVU3418	hypothetical protein	**undefined**	*B. thetaiotaomicron* Δtdk pNBU2_erm IS4_*Pv*CTn:*tetQ*_Δ*BVU3433* expression only	*P. vulgatus* & *B. thetaiotaomicron* conjugation expression only
BVU3419	hypothetical protein	**undefined**	*B. vulgatus* Δ*tdk* expression only	undefined
BVU3420	Putative Helix-turn-helix domain	**-5.254241287**	0.2839966564	-0.05658352837
BVU3421	putative ribose phosphate pyrophosphokinase	**-2**	**-2.530514717**	0.3521825181
BVU3422	hypothetical protein	**-2.222392421**	**-0.2995602819**	**-0.3870231231**
BVU3423	putative SOS-response transcriptional repressor	-2.415037499	**0.2656519561**	-0.1076755175
BVU3426	Putative Helix-turn-helix domain	-2.228268988	-0.2954558835	**-2.22940148**
BVU3427	HNH endonuclease	-2.307054818	0.05022711302	**-2.269788896**
BVU3428	DNA-binding protein HU	-1.398549376	-0.03355255971	**-3.175571565**
BVU3429	Putative Helix-turn-helix domain	-2.923378718	-0.05153030064	**-0.4901033144**
BVU3430	DNA primase	-4.329920886	**-2.133583154**	0.1271047984
BVU_RS21735	Putative Helix-turn-helix domain	No expression	**-1.431339312**	0.07918124605
BVU3432	Putative Helix-turn-helix domain	**-4.981852653**	**-1.924972955**	0.2635676041
BVU_RS21835	hypothetical protein	**-4.970616193**	**-1.933181893**	0.3611666111
BVU3433	Putative Helix-turn-helix domain	No expression (Δ*BVU3433*)	No expression (Δ*BVU3433*)	**-1.246793765**

Significant values are ($q < 0.01$) in **bold**

\

Table 2.5. Annotated regions of *BVU3433* encoding *Pv*CTn-like elements

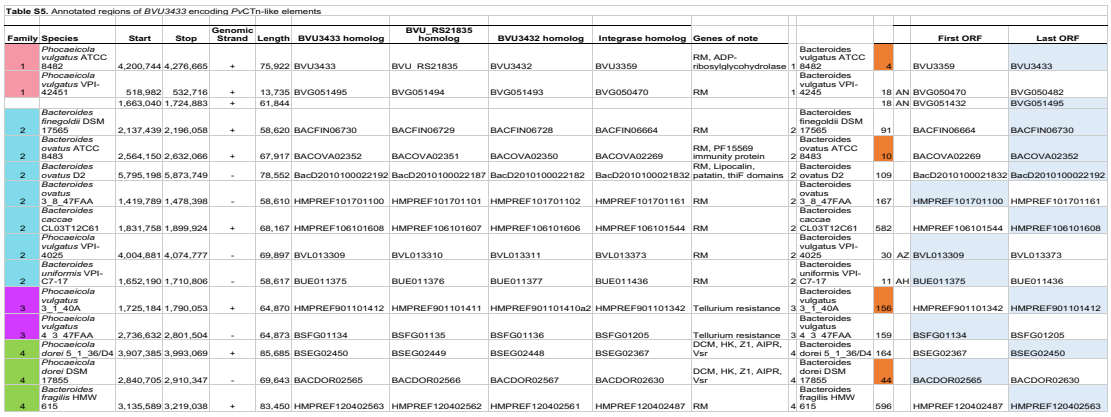


Table S5. Annotated regions of *BVU3433* encoding *Pv*CTn-like elements

Family	Species	Start	Stop	Genomic Strand	Length	BVU3433 homolog	BVU_RS21835 homolog	BVU3432 homolog	Integrase homolog	Genes of note				First ORF	Last ORF
1	*Phocaeicola vulgatus* ATCC 8482	4,200,744	4,276,665	+	75,922	BVU3433	BVU_RS21835	BVU3432	BVU3359	RM, ADP-ribosylglycohydrolase	1	Bacteroides vulgatus ATCC 8482	4	BVU3359	BVU3433
1	*Phocaeicola vulgatus* VPI-42451	518,982	532,716	+	13,735	BVG051495	BVG051494	BVG051493	BVG050470	RM	1	Bacteroides vulgatus VPI-4245	18	AN BVG050470	BVG050482
		1,663,040	1,724,883	+	61,844								18	AN BVG051432	BVG051495
2	*Bacteroides finegoldii* DSM 17565	2,137,439	2,196,058	+	58,620	BACFIN06730	BACFIN06729	BACFIN06728	BACFIN06664	RM	2	Bacteroides finegoldii DSM 17565	91	BACFIN06664	BACFIN06730
2	*Bacteroides ovatus* ATCC 8483	2,564,150	2,632,066	+	67,917	BACOVA02352	BACOVA02351	BACOVA02350	BACOVA02269	RM, PF15569 immunity protein	2	Bacteroides ovatus ATCC 8483	10	BACOVA02269	BACOVA02352
2	*Bacteroides ovatus* D2	5,795,198	5,873,749	-	78,552	BacD2010100022192	BacD2010100022187	BacD2010100022182	BacD2010100021832	RM, Lipocalin, patatin, thiF domains	2	Bacteroides ovatus D2	109	BacD2010100021832	BacD2010100022192
2	*Bacteroides ovatus* 3_8_47FAA	1,419,789	1,478,398	-	58,610	HMPREF101701100	HMPREF101701101	HMPREF101701102	HMPREF101701161	RM	2	Bacteroides ovatus 3_8_47FAA	167	HMPREF101701100	HMPREF101701161
2	*Bacteroides caccae* CL03T12C61	1,831,758	1,899,924	+	68,167	HMPREF106101608	HMPREF106101607	HMPREF106101606	HMPREF106101544	RM	2	Bacteroides caccae CL03T12C61	582	HMPREF106101544	HMPREF106101608
2	*Phocaeicola vulgatus* VPI-4025	4,004,881	4,074,777	-	69,897	BVL013309	BVL013310	BVL013311	BVL013373	RM	2	Bacteroides vulgatus VPI-4025	30	AZ BVL013309	BVL013373
2	*Bacteroides uniformis* VPI-C7-17	1,652,190	1,710,806	-	58,617	BUE011375	BUE011376	BUE011377	BUE011436	RM	2	Bacteroides uniformis VPI-C7-17	11	AH BUE011375	BUE011436
3	*Phocaeicola vulgatus* 3_1_40A	1,725,184	1,790,053	+	64,870	HMPREF901101412	HMPREF901101411	HMPREF901101410a2	HMPREF901101342	Tellurium resistance	3	Bacteroides vulgatus 3_1_40A	156	HMPREF901101342	HMPREF901101412
3	*Phocaeicola vulgatus* 4_3_47FAA	2,736,632	2,801,504	-	64,873	BSFG01134	BSFG01135	BSFG01136	BSFG01205	Tellurium resistance	3	Bacteroides vulgatus 4_3_47FAA	159	BSFG01134	BSFG01205
4	*Phocaeicola dorei* 5_1_36/D4	3,907,385	3,993,069	+	85,685	BSEG02450	BSEG02449	BSEG02448	BSEG02367	DCM, HK, Z1, AIPR, Vsr	4	Bacteroides dorei 5_1_36/D4	164	BSEG02367	BSEG02450
4	*Phocaeicola dorei* DSM 17855	2,840,705	2,910,347	-	69,643	BACDOR02565	BACDOR02566	BACDOR02567	BACDOR02630	DCM, HK, Z1, AIPR, Vsr	4	Bacteroides dorei DSM 17855	44	BACDOR02565	BACDOR02630
4	*Bacteroides fragilis* HMW 615	3,135,589	3,219,038	+	83,450	HMPREF120402563	HMPREF120402562	HMPREF120402561	HMPREF120402487	RM	4	Bacteroides fragilis HMW 615	596	HMPREF120402487	HMPREF120402563

Table 2.6. Table of metagenome samples screened for *Pv*CTn

			Quality Filtered Results						Quality Filtered Results		
Sample Name	**Run**	**Raw Read Pairs**	**1P/2P**	**1U**	**2U**	**Sample Name**	**Run**	**Raw Read Pairs**	**1P/2P**	**1U**	**2U**
PRJEB7774 – Salzburg, Austria											
Healthy						**CRC**					
SAMEA3136648	ERR688506	32,474,886	31,500,976	781,417	130,949	SAMEA3136748	ERR688505	20,056,458	17,120,791	1,924,569	456,599
SAMEA3136638	ERR688507	29,839,288	27,629,753	1,536,448	377,573	SAMEA3136769	ERR688513	26,382,550	23,148,453	1,886,754	673,822
SAMEA3136670	ERR688509	25,513,317	22,956,180	1,729,273	402,693	SAMEA3136751	ERR688515	25,837,477	22,154,313	2,307,252	650,980
SAMEA3136657	ERR688510	24,762,672	21,899,615	1,909,649	441,081	SAMEA3136738	ERR688517	25,516,040	22,348,148	1,699,160	864,153
SAMEA3136660	ERR688511	19,835,598	18,146,290	1,168,779	269,448	SAMEA3136747	ERR688521	26,617,685	23,363,876	2,109,163	540,233
SAMEA3136630	ERR688514	24,050,022	21,957,535	1,303,313	414,249	SAMEA3136743	ERR688525	23,424,383	21,506,970	1,291,862	341,336
SAMEA3136634	ERR688516	21,177,564	19,365,815	1,072,283	414,820	SAMEA3136754	ERR688527	29,927,254	25,538,097	2,816,303	755,114
SAMEA3136674	ERR688518	24,213,175	21,783,321	1,590,836	424,402	SAMEA3136765	ERR688534	31,308,192	27,553,544	2,098,063	943,387
SAMEA3136663	ERR688520	22,844,950	20,851,435	1,324,262	365,873	SAMEA3136724	ERR688541	31,580,158	30,080,886	1,158,447	232,528
SAMEA3136676	ERR688523	25,384,083	22,771,994	1,778,256	405,107	SAMEA3136755	ERR688547	23,261,680	20,252,830	1,949,666	500,972
SAMEA3136644	ERR688528	28,603,351	25,652,386	1,903,062	563,599	SAMEA3136766	ERR688548	29,363,386	26,464,203	1,824,104	559,229
SAMEA3136654	ERR688529	28,854,500	25,278,136	2,255,504	656,125	SAMEA3136728	ERR688553	27,886,772	24,684,593	1,815,256	777,589
SAMEA3136673	ERR688530	27,980,088	25,205,102	1,770,597	524,755	SAMEA3136756	ERR688558	24,446,743	21,264,618	2,008,163	577,340
SAMEA3136665	ERR688532	22,703,007	20,313,287	1,609,182	382,995	SAMEA3136753	ERR688567	23,197,613	20,269,436	1,976,574	467,848
SAMEA3136664	ERR688533	26,484,125	24,467,102	1,356,007	370,229	SAMEA3136737	ERR688568	19,520,755	17,711,436	958,219	529,747
SAMEA3136656	ERR688535	31,034,101	27,839,909	2,091,850	546,949	SAMEA3136726	ERR688569	31,222,562	30,399,151	665,767	122,672
SAMEA3136650	ERR688537	29,365,450	26,639,334	1,592,798	592,340	SAMEA3136734	ERR688570	22,124,801	19,330,652	1,699,641	584,116
SAMEA3136626	ERR688539	29,307,692	28,063,907	888,827	284,106	SAMEA3136758	ERR688571	29,360,386	26,041,846	2,156,881	549,735
SAMEA3136668	ERR688543	19,512,735	17,215,004	1,596,325	344,368	SAMEA3136750	ERR688572	32,431,127	28,142,253	2,744,436	726,532
SAMEA3136645	ERR688544	25,741,116	22,514,410	2,102,255	544,464	SAMEA3136742	ERR688573	27,671,795	23,877,181	2,392,507	682,740
SAMEA3136662	ERR688550	23,763,998	21,764,681	1,390,537	328,620	SAMEA3136760	ERR688574	27,911,137	23,544,063	2,495,254	877,899
SAMEA3136667	ERR688551	23,118,858	20,542,237	1,724,685	426,942	SAMEA3136727	ERR688575	30,227,502	26,307,887	2,176,736	966,233
SAMEA3136624	ERR688552	24,414,011	20,697,687	1,660,047	1,453,198	SAMEA3136752	ERR688576	29,991,991	25,620,107	2,740,581	762,503
SAMEA3136625	ERR688556	21,976,389	17,845,927	2,268,584	892,454	SAMEA3136740	ERR688577	28,686,450	24,907,264	2,179,193	870,821
SAMEA3136679	ERR688557	21,517,664	19,004,987	1,687,479	381,070	SAMEA3136731	ERR688578	27,061,552	23,346,907	1,999,026	959,928
SAMEA3136632	ERR688559	32,818,713	30,279,788	1,436,562	659,788	SAMEA3136729	ERR688579	25,170,549	18,984,930	2,423,842	2,393,583
SAMEA3136677	ERR688561	25,549,375	23,216,784	1,612,759	366,829	SAMEA3136767	ERR688580	27,416,746	23,252,209	2,647,479	693,101
SAMEA3136633	ERR688562	29,099,906	26,163,139	1,886,410	494,954	SAMEA3136761	ERR688581	30,717,938	28,355,657	1,600,857	405,836
SAMEA3136646	ERR688563	24,787,138	22,631,499	1,438,110	370,614	SAMEA3136725	ERR688582	33,201,178	29,795,875	2,121,171	783,191
SAMEA3136652	ERR688564	29,786,700	27,864,235	1,340,236	369,976	SAMEA3136763	ERR688583	36,202,656	31,701,516	2,881,887	789,167
SAMEA3136658	ERR688565	27,165,858	23,932,485	2,161,241	513,292	SAMEA3136732	ERR688618	25,369,054	21,947,347	1,890,208	844,156
SAMEA3136623	ERR688566	19,453,783	17,895,660	957,878	325,598	SAMEA3136746	ERR688634	29,017,379	25,071,811	2,609,666	598,369
SAMEA3136629	ERR688590	20,201,402	18,516,884	1,040,597	344,990	SAMEA3136759	ERR688635	22,537,125	18,982,007	2,235,431	604,087
SAMEA3136675	ERR688593	21,987,677	19,762,400	1,491,886	379,176	SAMEA3136733	ERR688636	28,120,088	24,910,152	1,818,183	786,134

SAMEA3136639	ERR688594	21,835,542	19,410,511	1,558,486	420,499	SAMEA3136757	ERR688637	30,100,955	26,304,393	2,453,814	622,849
SAMEA3136678	ERR688595	22,482,098	19,852,628	1,752,157	407,579	SAMEA3136768	ERR688638	36,858,391	30,823,685	3,953,816	891,544
SAMEA3136651	ERR688596	28,090,317	24,699,479	2,164,374	588,374	SAMEA3136749	ERR688639	25,919,573	22,224,635	2,433,846	574,300
SAMEA3136636	ERR688598	28,370,594	25,294,774	2,048,095	495,454	SAMEA3136730	ERR688640	28,651,553	24,795,885	2,232,362	853,420
SAMEA3136655	ERR688599	26,625,167	23,926,021	1,804,683	430,935	SAMEA3136744	ERR688641	28,467,172	25,223,946	2,106,670	555,329
SAMEA3136627	ERR688600	23,383,277	21,396,836	1,287,874	545,453	SAMEA3136764	ERR688642	31,804,185	27,775,516	2,492,354	768,297
SAMEA3136631	ERR688603	24,780,764	22,812,464	1,257,699	361,687	SAMEA3136735	ERR688643	24,539,947	22,137,592	1,352,441	630,814
SAMEA3136666	ERR688605	20,578,474	18,625,276	1,330,451	320,960	SAMEA3136745	ERR688644	23,237,641	20,268,916	1,899,280	500,927
SAMEA3136640	ERR688606	27,786,037	25,327,434	1,635,315	433,266	SAMEA3136741	ERR688647	26,712,989	23,686,757	1,895,307	565,892
SAMEA3136649	ERR688607	19,278,860	17,600,913	1,154,371	266,110	SAMEA3136739	ERR688649	24,786,392	22,485,881	1,235,640	681,228
SAMEA3136672	ERR688608	23,631,423	21,130,414	1,669,313	401,227	SAMEA3136762	ERR688650	33,015,193	28,746,655	2,728,882	730,228
SAMEA3136642	ERR688610	28,554,120	25,584,471	1,960,312	517,860	SAMEA3136736	ERR688651	21,302,038	18,909,782	1,343,479	612,675
SAMEA3136628	ERR688612	20,163,192	18,279,124	1,154,805	373,272						
SAMEA3136671	ERR688614	22,585,513	20,337,400	1,513,854	388,132						
SAMEA3136653	ERR688616	21,617,085	19,122,657	1,568,862	472,074						
SAMEA3136643	ERR688617	20,561,125	18,390,990	1,384,565	397,214						
SAMEA3136641	ERR688619	32,723,661	29,252,759	2,104,514	693,827						
SAMEA3136637	ERR688622	24,583,040	19,525,739	2,539,731	1,133,355						
SAMEA3136659	ERR688624	30,665,000	27,249,450	2,178,947	648,864						
SAMEA3136669	ERR688646	26,879,103	24,215,309	1,758,560	467,435						
SAMEA3136647	ERR688648	23,309,322	20,744,586	1,433,690	587,253						
SAMEA3178936	ERR710426	23,997,592	22,450,542	947,315	305,832						
SAMEA3178937	ERR710427	27,127,751	24,987,201	1,328,101	392,099						
SAMEA3178938	ERR710428	25,133,862	22,997,197	1,326,587	388,372						
SAMEA3178939	ERR710429	20,376,523	17,979,686	1,673,051	274,384						
SAMEA3178940	ERR710430	17,301,869	15,481,528	1,331,566	180,052						
SAMEA3178943	ERR710432	25,550,387	23,612,068	1,183,650	374,892						
PRJEB12449 – Hong Kong, China											
Healthy						**CRC**					
MMRS11932626ST-27-0-0	ERR1293879	4,165,813	4,165,723	89	1	MMRS12272136ST-27-0-0	ERR1293875	7,958,522	7,371,992	281,977	265,780
MMRS17603756ST-27-0-0	ERR1293814	8,700,361	7,894,696	484,792	265,598	MMRS14379078ST-27-0-0	ERR1293545	7,669,488	7,211,573	255,191	177,883
MMRS20257302ST-27-0-0	ERR1293758	6,487,580	5,896,949	309,853	232,747	MMRS14602194ST-27-0-0	ERR1293810	9,360,213	8,676,523	389,789	255,755
MMRS25211151ST-27-0-0	ERR1293702	6,445,549	5,993,043	289,161	144,162	MMRS15137911ST-27-0-0	ERR1293915	7,654,891	6,975,973	368,728	268,348
MMRS29805707ST-27-0-0	ERR1293710	8,569,687	8,569,603	60	24	MMRS16644320ST-27-0-0	ERR1293738	11,336,913	10,470,880	487,292	320,279
MMRS32573774ST-27-0-0	ERR1293670	8,109,956	8,109,893	52	11	MMRS18295921ST-27-0-0	ERR1293565	8,719,308	8,204,698	291,066	197,228
MMRS33294861ST-27-0-0	ERR1293666	7,372,625	6,701,284	373,654	247,558	MMRS24340226ST-27-0-0	ERR1293581	5,982,791	4,960,047	644,203	281,985
MMRS34569532ST-27-0-0	ERR1293862	7,391,296	6,746,362	368,812	227,674	MMRS25641480ST-27-0-0	ERR1293750	9,465,476	8,623,873	490,748	290,118
MMRS35808664ST-27-0-0	ERR1293818	9,546,412	8,679,346	537,972	272,809	MMRS31306426ST-27-0-0	ERR1293674	4,881,834	4,052,301	521,541	228,637
MMRS36405092ST-27-0-0	ERR1293509	6,481,635	6,481,608	9	18	MMRS34525504ST-27-0-0	ERR1293792	5,294,840	4,562,516	431,125	239,223
MMRS38861560ST-27-0-0	ERR1293847	8,803,195	8,066,802	400,699	276,927	MMRS41266697ST-27-0-0	ERR1293798	7,550,634	6,922,418	306,425	274,202

MMRS38954404ST-27-0-0	ERR1293794	9,070,717	8,238,122	490,333	282,007	MMRS41711735ST-27-0-0	ERR1293678	4,361,080	3,472,509	574,590	225,675
MMRS39415781ST-27-0-0	ERR1293890	9,078,459	9,078,310	138	11	MMRS42780924ST-27-0-0	ERR1293894	16,927,053	15,352,981	930,221	530,297
MMRS39582183ST-27-0-0	ERR1293662	7,283,007	6,678,628	334,839	233,266	MMRS43563715ST-27-0-0	ERR1293906	7,592,250	6,928,265	359,929	255,025
MMRS40913371ST-27-0-0	ERR1293871	8,953,361	8,194,943	422,473	281,575	MMRS43939546ST-27-0-0	ERR1293726	9,987,060	9,192,380	436,915	300,857
MMRS51307502ST-27-0-0	ERR1293926	6,324,255	5,463,722	478,472	305,826	MMRS44197125ST-27-0-0	ERR1293730	5,044,382	4,527,593	264,701	203,983
MMRS51737257ST-27-0-0	ERR1293609	10,775,417	9,820,550	474,399	406,638	MMRS44203591ST-27-0-0	ERR1293622	4,964,689	4,207,199	480,879	209,419
MMRS54246132ST-27-0-0	ERR1293694	8,112,395	6,987,506	636,651	392,830	MMRS44587603ST-27-0-0	ERR1293827	6,001,523	5,527,876	274,737	173,278
MMRS56137463ST-27-0-0	ERR1293754	8,835,619	8,564,488	155,572	108,716	MMRS47662059ST-27-0-0	ERR1293577	4,957,021	4,040,592	582,020	246,074
MMRS57276462ST-27-0-0	ERR1293593	8,145,402	7,515,092	343,589	245,594	MMRS47685461ST-27-0-0	ERR1293746	8,354,177	7,625,735	372,328	296,883
MMRS57569834ST-27-0-0	ERR1293597	7,864,862	7,864,799	35	28	MMRS51410669ST-27-0-0	ERR1293613	6,662,166	6,132,419	296,647	201,640
MMRS61683006ST-27-0-0	ERR1293930	12,765,378	11,857,109	521,304	337,380	MMRS51728985ST-27-0-0	ERR1293557	473,328	436,751	20,680	13,174
MMRS62192467ST-27-0-0	ERR1293854	7,718,788	7,230,216	266,472	196,306	MMRS54316675ST-27-0-0	ERR1293634	7,712,194	6,896,394	464,095	290,842
MMRS62666492ST-27-0-0	ERR1293549	6,440,038	6,439,965	48	25	MMRS55294627ST-27-0-0	ERR1293722	8,669,359	8,001,592	373,719	252,707
MMRS64281551ST-27-0-0	ERR1293525	7,573,520	7,052,014	285,669	204,484	MMRS59699117ST-27-0-0	ERR1293573	5,558,353	5,092,169	234,455	195,148
MMRS65742083ST-27-0-0	ERR1293698	8,781,106	7,896,657	540,077	284,544	MMRS62376181ST-27-0-0	ERR1293537	8,782,706	8,209,737	317,444	224,856
MMRS65862658ST-27-0-0	ERR1293911	8,484,754	7,834,599	343,980	269,866	MMRS64281211ST-27-0-0	ERR1293837	9,824,906	9,029,274	446,211	289,045
MMRS66868025ST-27-0-0	ERR1293505	7,181,410	6,204,249	537,823	353,507	MMRS65205033ST-27-0-0	ERR1293529	5,945,244	5,448,397	267,831	189,003
MMRS67096717ST-27-0-0	ERR1293553	4,702,076	3,988,457	415,240	236,359	MMRS65945127ST-27-0-0	ERR1293690	8,536,642	8,192,225	214,461	120,063
MMRS68630931ST-27-0-0	ERR1293626	7,845,816	7,218,205	320,503	258,127	MMRS67247116ST-27-0-0	ERR1293654	11,744,915	11,744,842	45	28
MMRS73787048ST-27-0-0	ERR1293770	9,028,424	8,342,094	351,603	293,779	MMRS68910755ST-27-0-0	ERR1293630	8,364,630	7,624,826	374,371	301,325
MMRS79893829ST-27-0-0	ERR1293714	4,788,801	3,877,457	552,439	265,956	MMRS71762412ST-27-0-0	ERR1293605	8,200,562	7,406,233	451,931	279,918
MMRS81130264ST-27-0-0	ERR1293923	8,437,556	8,437,399	153	4	MMRS72802364ST-27-0-0	ERR1293569	8,532,435	7,887,609	341,923	267,853
MMRS81135225ST-27-0-0	ERR1293513	9,189,906	8,562,391	336,173	263,997	MMRS72980899ST-27-0-0	ERR1293766	8,690,800	7,992,027	392,779	253,373
MMRS84085284ST-27-0-0	ERR1293533	7,797,273	7,797,242	25	6	MMRS77018926ST-27-0-0	ERR1293658	9,053,685	7,935,940	634,556	383,674
MMRS84159866ST-27-0-0	ERR1293686	8,552,994	7,850,932	376,328	284,239	MMRS77321055ST-27-0-0	ERR1293517	7,793,635	7,411,710	214,712	150,796
MMRS85438660ST-27-0-0	ERR1293778	8,405,926	7,811,136	308,455	255,805	MMRS77487595ST-27-0-0	ERR1293786	7,912,666	7,158,548	422,633	285,752
MMRS85548821ST-27-0-0	ERR1293706	7,719,859	7,719,838	10	11	MMRS78791111ST-27-0-0	ERR1293830	8,059,625	7,441,309	327,003	253,515
MMRS86168210ST-27-0-0	ERR1293682	6,266,439	5,370,910	487,363	317,723	MMRS80093644ST-27-0-0	ERR1293646	6,656,716	6,064,492	312,949	243,149
MMRS87723283ST-27-0-0	ERR1293867	9,069,691	8,343,567	398,657	276,355	MMRS81194191ST-27-0-0	ERR1293883	6,549,745	6,036,572	245,947	225,338
MMRS88581325ST-27-0-0	ERR1293782	7,563,080	7,051,992	295,574	186,984	MMRS81671217ST-27-0-0	ERR1293886	8,792,592	8,792,138	439	15
MMRS91297590ST-27-0-0	ERR1293585	5,213,787	4,283,766	543,900	290,430	MMRS84954772ST-27-0-0	ERR1293903	9,324,920	8,592,759	412,645	269,206
MMRS92546335ST-27-0-0	ERR1293589	4,487,796	3,852,513	379,424	204,109	MMRS85382779ST-27-0-0	ERR1293734	7,441,983	6,921,603	288,598	204,261
MMRS93621581ST-27-0-0	ERR1293806	10,039,947	9,181,750	497,869	299,466	MMRS86178473ST-27-0-0	ERR1293521	8,002,007	7,252,300	433,874	260,150
MMRS94549160ST-27-0-0	ERR1293742	8,433,256	8,012,343	220,616	181,265	MMRS86958796ST-27-0-0	ERR1293720	4,862,033	4,015,269	541,690	225,648
MMRS95277876ST-27-0-0	ERR1293898	4,668,761	4,668,705	50	6	MMRS90191266ST-27-0-0	ERR1293774	7,719,851	6,857,229	493,978	302,814
MMRS95479054ST-27-0-0	ERR1293802	8,469,848	8,355,692	67,335	45,591	MMRS92544445ST-27-0-0	ERR1293842	9,219,995	8,503,720	358,655	316,618
MMRS95674036ST-27-0-0	ERR1293601	5,399,558	4,960,577	253,281	152,690	MMRS93508324ST-27-0-0	ERR1293638	7,895,629	7,342,799	301,901	224,476
MMRS97307143ST-27-0-0	ERR1293850	7,363,809	6,579,348	450,299	274,355	MMRS96422329ST-27-0-0	ERR1293822	5,721,665	5,226,743	247,197	207,597
						MMRS97052561ST-27-0-0	ERR1293561	10,360,805	9,388,500	570,062	332,922
						MMRS97327636ST-27-0-0	ERR1293934	7,314,317	6,850,510	265,455	176,296
						MMRS99045180ST-27-0-0	ERR1293501	5,179,510	4,803,045	198,999	150,638

PRJEB10878 – Washington DC, USA											
Healthy						**CRC**					
SAMEA3541477	ERR1018193	30,614,519	24,387,823	3,217,681	1,245,254	SAMEA3541469	ERR1018185	30,374,603	24,816,830	2,490,258	1,478,727
SAMEA3541479	ERR1018195	32,082,108	26,398,574	2,954,176	1,256,651	SAMEA3541470	ERR1018186	26,427,396	20,230,611	2,889,137	1,653,758
SAMEA3541489	ERR1018205	28,517,594	23,087,399	2,750,310	1,226,523	SAMEA3541471	ERR1018187	29,936,691	24,186,351	2,589,639	1,451,782
SAMEA3541493	ERR1018209	28,067,170	23,260,103	2,496,838	1,239,547	SAMEA3541472	ERR1018188	27,880,851	22,812,194	2,543,300	1,162,750
SAMEA3541501	ERR1018217	31,665,069	29,704,168	1,314,513	476,643	SAMEA3541473	ERR1018189	32,217,415	24,266,308	4,093,531	1,819,887
SAMEA3541516	ERR1018232	33,032,051	29,840,803	1,852,002	725,755	SAMEA3541474	ERR1018190	26,051,114	20,648,261	2,743,538	1,155,414
SAMEA3541517	ERR1018233	35,214,496	32,410,402	1,653,101	598,108	SAMEA3541475	ERR1018191	31,228,455	25,077,746	3,094,275	1,358,360
SAMEA3541518	ERR1018234	30,808,659	26,379,380	2,236,902	1,058,611	SAMEA3541476	ERR1018192	30,036,687	27,040,632	2,110,547	566,308
SAMEA3541519	ERR1018235	34,236,068	31,054,194	1,896,828	643,663	SAMEA3541478	ERR1018194	30,736,143	24,436,806	3,203,675	1,394,480
SAMEA3541520	ERR1018236	34,767,133	31,907,904	1,649,156	634,381	SAMEA3541480	ERR1018196	33,075,319	27,211,759	3,175,181	1,247,463
SAMEA3541521	ERR1018237	34,002,045	30,793,356	1,708,169	775,875	SAMEA3541481	ERR1018197	30,037,451	25,088,656	2,256,506	1,358,820
SAMEA3541522	ERR1018238	36,038,719	32,758,509	1,897,918	728,056	SAMEA3541482	ERR1018198	28,675,806	21,853,107	3,156,627	1,779,338
SAMEA3541523	ERR1018239	29,579,886	26,099,199	1,799,705	872,837	SAMEA3541483	ERR1018199	14,794,931	11,935,241	1,447,011	734,738
SAMEA3541524	ERR1018240	33,739,277	31,105,234	1,442,445	641,972	SAMEA3541484	ERR1018200	29,273,446	24,524,977	2,201,835	1,242,269
SAMEA3541525	ERR1018241	33,659,451	30,969,939	1,516,656	605,596	SAMEA3541485	ERR1018201	28,292,499	21,330,735	3,740,310	1,509,969
SAMEA3541526	ERR1018242	31,921,893	27,882,019	2,182,128	1,249,348	SAMEA3541486	ERR1018202	25,646,078	21,047,454	2,226,962	1,115,020
SAMEA3541527	ERR1018243	31,904,075	29,053,816	1,546,074	657,274	SAMEA3541487	ERR1018203	27,058,316	22,645,540	2,392,934	907,983
SAMEA3541528	ERR1018244	30,260,317	27,842,015	1,249,573	662,701	SAMEA3541488	ERR1018204	28,324,241	20,159,828	4,433,475	1,551,343
SAMEA3541529	ERR1018245	26,349,074	21,251,220	2,144,391	2,029,421	SAMEA3541490	ERR1018206	31,129,040	24,660,414	3,651,579	1,324,130
SAMEA3541530	ERR1018246	26,394,471	24,099,613	1,340,463	493,332	SAMEA3541491	ERR1018207	22,818,639	18,121,977	2,241,650	1,216,626
SAMEA3541531	ERR1018247	28,557,198	25,523,016	1,686,292	678,544	SAMEA3541492	ERR1018208	29,411,059	22,074,465	3,499,200	1,832,374
SAMEA3541532	ERR1018248	29,295,754	26,790,722	1,467,544	550,008	SAMEA3541494	ERR1018210	30,410,236	25,645,951	3,580,149	582,295
SAMEA3541533	ERR1018249	24,846,934	21,091,358	1,730,635	1,414,147	SAMEA3541495	ERR1018211	30,968,869	24,359,904	3,206,336	1,514,417
SAMEA3541534	ERR1018250	23,762,242	21,126,641	1,565,173	533,541	SAMEA3541496	ERR1018212	29,832,977	24,020,524	2,830,471	1,343,282
SAMEA3541535	ERR1018251	27,749,839	25,701,522	1,281,615	558,917	SAMEA3541497	ERR1018213	28,738,139	22,199,125	3,715,741	1,253,332
SAMEA3541536	ERR1018252	23,156,930	21,629,672	1,065,516	339,347	SAMEA3541498	ERR1018214	31,174,741	29,154,633	1,375,035	471,450
SAMEA3541537	ERR1018253	24,245,250	18,162,331	2,494,824	2,200,560	SAMEA3541499	ERR1018215	35,808,593	29,781,354	2,436,299	2,494,302
SAMEA3541538	ERR1018254	28,353,974	25,811,711	1,347,776	647,862	SAMEA3541500	ERR1018216	23,514,642	17,896,665	2,199,404	2,173,654
SAMEA3541539	ERR1018255	22,009,545	17,196,663	2,031,889	1,812,561	SAMEA3541503	ERR1018219	36,185,222	33,142,591	1,811,068	660,978
SAMEA3541540	ERR1018256	22,452,181	20,552,328	1,134,319	407,334	SAMEA3541504	ERR1018220	34,079,841	31,596,467	1,489,121	550,530
SAMEA3541541	ERR1018257	30,797,686	28,125,625	1,391,576	740,162	SAMEA3541505	ERR1018221	30,593,520	28,196,079	1,603,452	585,715
SAMEA3541542	ERR1018258	28,981,180	27,124,859	1,262,800	434,843	SAMEA3541506	ERR1018222	28,517,493	24,801,937	2,135,858	730,665
SAMEA3541543	ERR1018259	28,280,953	26,578,521	1,189,354	387,925	SAMEA3541507	ERR1018223	30,667,313	27,769,253	1,562,998	770,078
SAMEA3541544	ERR1018260	30,977,906	26,403,645	2,076,711	1,719,330	SAMEA3541508	ERR1018224	25,154,452	23,686,054	983,027	362,974
SAMEA3541545	ERR1018261	27,160,683	22,350,445	2,192,080	1,738,038	SAMEA3541509	ERR1018225	21,602,596	19,731,913	1,046,480	433,372
SAMEA3541546	ERR1018262	38,631,754	35,034,619	1,841,223	931,772	SAMEA3541510	ERR1018226	34,219,487	30,802,487	1,933,594	729,119
SAMEA3541547	ERR1018263	27,789,674	25,516,354	1,282,361	544,805	SAMEA3541511	ERR1018227	36,130,113	33,211,467	1,628,049	685,745
SAMEA3541548	ERR1018264	29,519,443	22,804,527	3,034,077	2,259,891	SAMEA3541512	ERR1018228	31,284,880	28,257,677	1,675,262	673,884

SAMEA3541549	ERR1018265	31,392,044	28,919,425	1,415,514	578,525		SAMEA3541513	ERR1018229	29,269,564	26,762,401	1,613,461	645,947
SAMEA3541550	ERR1018266	30,987,813	28,328,767	1,421,582	656,155		SAMEA3541514	ERR1018230	29,992,194	26,731,872	1,849,939	697,752
SAMEA3541551	ERR1018267	25,735,887	22,122,383	1,881,978	1,160,489		SAMEA3541515	ERR1018231	26,813,809	24,586,299	1,250,570	551,199
SAMEA3541552	ERR1018268	29,182,339	26,713,033	1,315,906	628,449		SAMEA3541563	ERR1018279	31,890,697	25,159,838	3,143,993	2,068,840
SAMEA3541553	ERR1018269	30,338,245	28,187,869	1,220,558	522,724		SAMEA3541564	ERR1018280	32,566,793	26,538,027	3,040,322	1,854,786
SAMEA3541554	ERR1018270	31,670,543	29,452,872	1,266,028	525,001		SAMEA3541565	ERR1018281	26,093,875	18,285,621	4,917,065	1,106,664
SAMEA3541555	ERR1018271	27,595,011	22,142,021	2,443,846	1,989,149		SAMEA3541566	ERR1018282	28,451,973	20,971,996	4,820,812	1,049,654
SAMEA3541556	ERR1018272	30,126,113	26,766,382	1,997,547	671,678		SAMEA3541567	ERR1018283	25,622,872	18,944,859	4,251,322	966,771
SAMEA3541557	ERR1018273	31,464,688	28,774,725	1,708,166	722,711		SAMEA3541568	ERR1018284	25,906,119	18,955,161	4,393,128	1,125,739
SAMEA3541558	ERR1018274	31,640,182	30,041,933	1,160,716	330,676		SAMEA3541569	ERR1018285	31,074,892	22,909,304	5,149,003	1,215,448
SAMEA3541559	ERR1018275	21,608,942	19,422,429	1,292,214	492,968		SAMEA3541570	ERR1018286	27,860,482	22,540,404	3,272,015	989,906
SAMEA3541560	ERR1018276	22,094,138	20,357,644	993,142	406,869		SAMEA3541571	ERR1018287	18,445,604	17,753,299	660,318	26,879
SAMEA3541561	ERR1018277	30,335,506	28,132,994	1,152,001	604,681		SAMEA3541572	ERR1018288	19,465,986	14,850,451	2,335,105	1,297,291
SAMEA3541562	ERR1018278	31,661,345	26,970,171	2,238,000	1,694,979		SAMEA3541573	ERR1018289	17,512,836	13,929,978	1,877,108	1,033,315
SAMEA3541592	ERR1018308	29,040,137	26,542,894	1,598,412	479,960		SAMEA3541574	ERR1018290	19,143,105	15,267,408	1,922,746	1,170,634
SAMEA3541593	ERR1018309	28,609,717	25,838,673	1,794,450	477,649		SAMEA3541575	ERR1018291	19,210,534	14,986,972	2,155,558	1,223,306
							SAMEA3541576	ERR1018292	18,172,884	14,745,924	1,730,178	1,041,571
							SAMEA3541577	ERR1018293	18,512,218	14,940,036	1,833,387	1,049,881
							SAMEA3541578	ERR1018294	20,203,586	15,166,711	2,343,330	1,518,766
							SAMEA3541579	ERR1018295	21,211,853	17,125,486	2,054,185	1,230,862
							SAMEA3541580	ERR1018296	5,480,782	5,074,627	270,138	67,884
							SAMEA3541581	ERR1018297	24,143,910	19,133,953	2,372,980	1,585,561
							SAMEA3541582	ERR1018298	18,813,807	17,094,188	1,152,125	438,353
							SAMEA3541583	ERR1018299	25,269,573	23,571,955	1,132,747	285,968
							SAMEA3541584	ERR1018300	20,120,144	15,980,123	2,044,408	1,222,152
							SAMEA3541585	ERR1018301	22,170,740	20,236,861	1,273,100	316,094
							SAMEA3541586	ERR1018302	32,725,364	30,754,526	1,299,655	368,070
							SAMEA3541587	ERR1018303	28,657,938	25,787,201	1,910,726	451,870
							SAMEA3541588	ERR1018304	29,521,220	27,119,416	1,587,359	409,277
							SAMEA3541589	ERR1018305	29,259,097	26,891,840	1,561,941	389,517
							SAMEA3541590	ERR1018306	24,678,251	22,751,365	1,271,764	366,077
							SAMEA3541591	ERR1018307	25,535,223	23,306,394	1,479,880	355,618
							SAMEA3541594	ERR1018310	33,093,058	30,020,405	1,939,704	582,353
							SAMEA3541595	ERR1018311	29,394,943	27,524,800	1,104,984	487,683
							SAMEA3541596	ERR1018312	31,669,978	29,454,219	1,292,438	571,578

CHAPTER 3: F**amily of CTnDOT-like conjugative transposons represent a pervasive global reservoir for antibiotic resistance**

Abstract

Antibiotic resistant bacteria pose an escalating challenge to human health, particularly in clinical settings. The pace at which antibiotic resistance can emerge and spread within bacterial populations exceeds the rate at which researchers can develop and deploy novel antibiotics. A pivotal driver of antibiotic resistance spread is horizontal gene transfer (HGT) via mechanisms such as conjugation. The conjugative transposon CTnDOT has been an important model for studying the regulation and function of conjugative elements and antibiotic resistance among gut microbes. However, the sequence, structural and functional diversity of CTnDOT-like elements has not been well characterized. Here, we detected the *in vitro* mobilization of a novel CTnDOT-like element from *Parabacteroides merdae* (PmDOT) into *Bacteroides thetaiotaomicron*. While both CTnDOT and PmDOT encode tetracycline resistance (*tetQ*) only CTnDOT encodes the 12.8kb *ermF* integron containing erythromycin and streptomycin resistance genes. Subsequent genome screens identified intact, CTnDOT-like elements in ~50% of the 134 human-gut associated Bacteroidota genomes analyzed. Most of these CTn-DOT-like elements confer tetracycline resistance (70/77) while only a minority provided dual

tetracycline and erythromycin resistance (11/77) to their bacterial hosts. Furthermore, analysis human fecal metagenomes confirm the widespread occurrence of CTnDOT-like elements among globally distributed human populations. Together, these data indicate that CTnDOT-like elements are a diverse and important group of MGEs in the human gut microbiome and likely continue to contribute to the transfer of antibiotic resistance in the gut.

Introduction

Horizontal gene transfer (HGT) plays a pivotal role in shaping the diversity of and adaptations in bacterial strains (Rankin et al. 2011). The human gut microbiome not only exhibits elevated HGT activity, but it is also an important reservoir for antibiotic resistance genes (Smillie et al. 2011, van Shaik 2015). Antibiotic resistance and its transfer among commensal and pathogenic bacterial strains by mobile genetic elements is of significant clinical concern (Turner et al. 2019). Conjugation, facilitated by Type IV Secretion System (T4SS), is a key mechanism of HGT in human gut-associated bacteria (Johnson & Grossman 2015). Therefore, understanding the diversity and functions of conjugative transposons (CTns) is critical for unraveling the dynamics of antibiotic resistance spread in gut microbial communities.

CTnDOT is a model integrative and conjugative element (ICE) among gut Bacteroidota as it plays a pivotal role in the dissemination of tetracycline and erythromycin resistance (Tet^R and Erm^R) (Cheng et al. 2000). Early studies using

Southern hybridizations and PCR of marker genes determined that >80% of natural *Bacteroides* clinical and community isolates in 1990 were Tet^R, compared to 20-30% in the 1970s (Shoemaker et al. 2001). This massive increase in Tet^R was largely attributed to the increased mobilization of CTnDOT among these gut microbes likely due to the prescription of tetracycline during this time (Shoemaker et al. 2001). Despite these insights, a comprehensive genomic analysis of CTnDOT-like elements, particularly in more recent metagenomic samples, remains unexplored.

Horizontal transfer of CTnDOT requires a series of key steps including excision, mobilization, transfer, replication, and integration that are induced upon exposure to tetracycline (Waters & Salyers 2013). Initially, the *tetQ-rteA-rteB* operon is subject to translational attenuation. However, upon tetracycline exposure, translation resumes, enabling RteB to initiate RteC transcription (Jeters et al. 2009). Subsequently, RteC activates the excision operon, facilitating the removal of CTnDOT from the chromosome (Keeton et al. 2013). Transcriptional regulation also occurs, with Xis2c and Xis2d governing the *tra* region, while Xis2d and Exc enhance *mob* transcription (Hopp et al. 2015). While this mechanism is well-established for CTnDOT, close relatives exhibit different behaviors, such as CTnERL, which has a transfer origin (*oriT*) located near the center of the element, which allows the transfer of a portion of itself with a mutagenized integrase gene (Whittle et al. 2006). Although the regulatory genes of CTnDOT are well-characterized it is unclear how broadly among CTnDOT-like elements these regulatory genes are conserved.

In our pursuit of novel mobile genetic elements (MGEs) through a transposon mutagenesis mobilization method (TMMM) (Ortañez & Degnan 2024), we discovered a functional CTnDOT-like element in *Parabacteroides merdae* (PmDOT) capable of integrating into *B. thetaiotaomicron*. Comparative analysis of CTnDOT, PmDOT and a group of other CTnDOT-like elements enabled us to identify the patterns of regulatory conservation among these elements. Further, by identifying the core genes of these CTnDOT-like elements, we were able to identify an expanded network of CTnDOT family elements among a panel of gut Bacteroidota. Additionally, a global metagenome analysis revealed the widespread prevalence of CTnDOT-like elements not only in industrial societies, but also in hunter-gatherer communities as well. Our characterization of previously overlooked CTnDOT-like elements promises valuable insights into the diverse mechanisms governing the transfer of antibiotic resistance genes.

Materials and Methods

Growth of bacterial strains

For the subsequent experiments *E. coli* S-17 λ *pir* donor strain with the pSAM_BT vector was grown aerobically at 37ºC in LB agar and liquid medium with 100 μg/mL of ampicillin as needed. Anaerobic Bacteroidota strains (Table 2.1) were grown in tryptone yeast extract glucose (TYG) liquid medium (Holdeman 1977) and BHI+HB agar comprised of Brain-Heart Infusion agar supplemented 10% v/v with defibrinated horse blood (Quad Five, Ryegate, MT) and amended with appropriate antibiotics (tetracycline 2

μg/mL, erythromycin 25 μg/mL, gentamicin 200 μg/mL). All anaerobic cultures and experiments were incubated at 37ºC in an atmosphere of 20% CO_2, 7% H_2 and the balance of N_2 gas in a flexible vinyl anaerobic chamber (Coy Laboratory Products, Grass Lake, MI) or stoppered tubes.

Mobilization of putative CTns

Transposon mediated screening of MGEs was employed to identify functional elements in *B. thetaiotaomicron* str. VPI-5482 Δ*tdk* as described previously (Ortañez & Degnan 2024). Briefly, the pSAM_BT vector was used to generate a library of *B. thetaiotaomicron* str. VPI-5482 Δ*tdk* Erm^R mutants. Resulting transconjugants were colony purified and stored at -80ºC. This mutant pool was then used as a conjugation donor with *P. merdae* ATCC 43184 and other strains with natively Tet^R strains to identify novel doubly resistant transconjugants. Another set of conjugation experiments were independently performed *in vitro*, using a pNBU2 generated Erm^R *B. thetaiotaomicron* str. VPI-5482 Δ*tdk* using as a recipient. The donors utilized in these experiments are detailed in Table 2.5.

All conjugations were carried out as described in (Ortañez & Degnan 2024). In brief, the recipient and donor strains were grown overnight in 5 mL of appropriate medium +/- antibiotic in anaerobic (TYG) or aerobic conditions (LB). Overnight cultures were then subcultured into the same medium at 1:25, 1:50, and 1:100 dilutions. Anaerobic cultures were stoppered, removed from the chamber, and incubated statically

at 37ºC and aerobic cultures were grown while shaking at 37ºC. Growth was monitored and pairs of donor and recipient cells were pelleted (4,000 x g for 5 min) when both reached an OD_{600} of ~0.4. The supernatants were removed, and cells were resuspended with 1mL TYG medium and pelleted again. After removing the supernatant, 1mL TYG medium was used to combine donors and recipients. The combined cells were then spread onto BHI+HB agar and incubated aerobically (*E. coli* donors) or anaerobically (Bacteroidota donors) at 37ºC for 24 hrs. The following day, conjugation masses were scraped up and resuspended in 5 mL TYG medium. The resuspended conjugation masses were then serially diluted and plated onto BHI+HB agar with antibiotic supplements and grown anaerobically at 37ºC for 48hrs.

Genome sequencing of a double transconjugant

A double transconjugant strain was subjected to whole genome sequencing to identify the MGE that had been acquired. The culture was revived and grown in TYG liquid and subjected to a standard DNA extraction protocol (Campbell et al. 2020). Qubit quality-checked DNA was subjected to library construction with the NEBNext Ultra II FS kit for Illumina (New England Biolabs, Ipswich, MA) at the UC Riverside IIGB Genomics Core. The sample library was quantified, pooled, and sequenced on an Illumina MiSeq with v3 chemistry generating 2 x 76 nt paired end reads.

Raw sequencing data was processed with Trimmomatic v0.39 (Bolger et al. 2014) to remove adaptor sequences and quality trim the reads (SLIDINGWINDOW:4:15 LEADING:2 TRAILING:2 MINLEN:55). High quality paired-end reads were then assembled with SPADES v3.15.4 (Prjibelski et al. 2020) using the default parameters. Resulting contigs greater than 500 bp were then subjected to gene prediction and functional annotation using Anvi'o v6.2 (Eren et al. 2020). Functional annotations were inferred using default thresholds for the anvi-run-hmms script with the PFAM (v33.1) and TIGRFAMs (15.0) databases and for the anvi-run-ncbi-cogs script with NCBI's Cluster of Orthologous Genes (COGs) database

Core and Pan genome determination of CTnDOT

We selected a panel of CTnDOT-like elements to identify the extent of gene conservation. The predicted CTnDOT-like elements were identified in the genome assemblies of bacterial strains in our collection (Table 3.2). The selection criteria required contiguous BLASTn matches with ≥85% nucleotide identity and spanning 35 ± 8 kbp of the assembly. Note that all CTnDOT-like elements encoded a *tetQ* gene (≥97% amino acid identity). First, the CTnDOT-like regions were aligned with the prototype CTnDOT element from *B. thetaiotaomicron* VPI-DOT to identify conserved and variable features (e.g., regulatory genes) using Mauve v2015-02-25 (Darling et al. 2004). In addition, select intergenic regions were aligned when required using Muscle v3.8.1551 (Edgar 2004). Second, protein coding sequences were grouped based on pairwise BLASTp (E value ≤ 0.0001) and clustered using MCL v14.137 (Inflation = 20) (Enright et al. 2002).

Custom Perl code was then used to identify the core/pan genes based on the MCL clusters for all combinations of the CTnDOT-like elements.

Comparisons of CTnDOT family among gut Bacteroidota

We used the set of core genes identified to detect putative CTnDOT-like sequences in a local database of 124 additional *Bacteroidota* genomes and 562 non-Bacteroidota gut microbial genomes (Table 3.3). BLASTp was used to find putative homologs (-evalue 1e-40). Putative CTnDOT-like genomic regions containing core gene homologs were cross-referenced with previous MGE predictions (Frye et al. 2021 and Ortañez & Degnan 2024) and individual boundaries of each CTnDOT-like were manually evaluated based on functional gene annotations surrounding genes and alignments of the genomic regions with related Bacteroidota strains using Mauve (Table 3.3). Inclusion criterion for predicted intact CTnDOT-like element were (*i*) ≥6 CORE genes and (*ii*) despite incomplete genome assemblies the CTn region appears continuous. We defined continuity as regions that completely encompassed in a single contig/scaffold or the region spans the end of two contigs/scaffolds that are conceivably disrupted only due to an assembly artifact. Discontinuous elements that were defined as "fragmentary" have >2 CTn aligning regions within >2 contigs/scaffolds. Identified genomic regions corresponding to putative CTns were then compared using pairwise BLASTn (E value ≤ 0.0001), filtered for ≥20% percent length aligned (PLA) (Leigh 2021) and then clustered using MCL v14-137 (Inflation = 20) (Enright 2002). Clusters were visualized in Cytoscape v3.10.1 (Shannon 2003).

Additional alignments of CTn elements were performed as needed using Mauve. Furthermore, 13 bacterial conserved core genes for all the examined strains were identified with BLASTp as described previously (*fusA, rpsH, aspS, ffh, gltX, infB, leuS, rplB, rpsE, rpsK, topA, tufA, rpoB*)(Degnan et al. 2014). Alignments of the amino acid sequences for the individual genes were performed using Muscle v3.8.1551 (Edgar 2004), then combined into a single matrix and maximum likelihood phylogeny was reconstructed using RAxML v8.2.12 (Stamatakis 2014) on the CIPRES server (Miller et al. 2010). The JTT protein substitution matrix was used, and bootstrapping was allowed to be automatically stopped.

Detection of CTnDOT-like elements in human gut metagenomes

To identify the prevalence of CTnDOT-like elements in human gut metagenome samples we employed a marker gene approach (Ortañez & Degnan 2024). First, nucleotide sequences a species-specific marker were retrieved from 46 representative genomes in our panel (Table 3.3) corresponding to universally conserved 30S ribosomal protein S5 (*rpsE*). Second, sequences for the core CTnDOT-like elements were retrieved. Then a panel of 10 short read metagenomic datasets were retrieved from the NCBI SRA database representing patients from diverse industrialized populations (PRJNA664754, PRJNA588313, PRJEB8094, PRJNA834885, PRJEB7774, PRJEB12449, PRJEB10878) and hunter-gatherer populations (PRJNA268964, PRJNA485056, PRJEB49206)(Table 3.4). These data were quality filtered with Trimmomatic as described above and mapped to the CTnDOT-like regions and the *rpsE* marker gene regions using Bowtie2 with the

default settings, which only allow reads to be aligned to a single reference sequence location (Langmead et al 2012). Raw read counts per *rpsE* or CTnDOT-like element were then normalized to Reads Per Kilobase per Million mapped reads (RPKM) values for each patient sample. Statistical comparisons were carried out in JMP Pro v16.

RESULTS

Transfer of PmDOT detected during screen for mobilizable elements

A conjugation experiment was carried out using *P. merdae* ATCC 43184 (Tet^R) with a transposon mutant library of *B. thetaiotaomicron* VPI-5482 (Erm^R). Successful transconjugants were isolated that were dually resistant, encoding both the Tet^R and Erm^R markers. Subsequent PCR analysis determined that all transconjugants (14/14) were a *B. thetaiotaomicron* genetic background that had acquired a CTnDOT-like element from *P. merdae* that we have named PmDOT. We carried out whole genome sequencing of a *B. thetaiotaomicron*-PmDOT transconjugant which confirmed the transfer of the entire ~48 kb conjugative element (Figure 3.1A).

Comparative analysis of the integrated PmDOT and CTnDOT (~64 kb) along with their flanking regions identified broad conservation of the core conjugative transposon structure. This includes key genes responsible for tetracycline resistance (*tetQ*) and Type IV Secretion (*tra*), which were highly conserved at the nucleotide level, ranging from 88% – 96% (Figure 3.1A). Notably, all genes associated with the regulation of CTnDOT excision and conjugative transfer (*rteB, rteC, xis2c, xis2d, exc, intDOT*) are

conserved in PmDOT (Figure 3.1A). However, we did identify several gene blocks present in one CTn and absent from the other. For example, PmDOT encodes a two gene operon immediately downstream of the *tetQ-rteA-rteB* operon that is absent in CTnDOT. The specific function of these genes is undetermined; PARMER00792 is a hypothetical protein, and PARMER00793 encodes a KAP family P-loop domain. Notably, CTnDOT encodes a ~12.8 kb integron adjacent to *intDOT* that encodes the genes for streptomycin (*aadS*) and erythromycin (*ermF*) resistance which is absent in PmDOT (Figure 3.1A) (Whittle et al. 2001).

Integration of CTnDOT is mediated via the integrase enzyme IntDOT which recognizes a short *attB* sequence motif GTAnnTTTGC and results in formation of two direct repeats (*attL, attR*) (Cheng et al. 2000, Wood and Gardner 2015). It has been previously shown that CTnDOT can integrate into at least 6 possible *attB* sites in the *Bacteroides thetaiotaomicron* genome (Cheng et al. 2000). The PmDOT element encodes an *intDOT* homolog that is 98% amino acid identical to the copy in CTnDOT. Further, most of the differences (8/9) are conservative amino acid changes, and none alter the predicted structure (Phyre2; Kelley et al. 2015). As such the alignment of the attachment sites for PmDOT and *B. thetaiotaomicron* PmDOT with CTnDOT reveals a high degree of conservation with the published CTnDOT *attB* motif (Figure 3.1B). While PmDOT in *P. merdae* matches the motif exactly, we did detect a single 1bp mismatch (G → T) at the 5' end of *B. thetaiotaomicron attB* site, which is preserved in the *attL* site once PmDOT is integrated (Figure 3.1B). This minor deviation may be stochastic or it may associated with one or more of the three amino acid differences detected in the core binding domain

of the PmDOT *intDOT* gene slightly altering the site specificity of the enzyme. Critically however, no differences were detected in the active site of PmDOT *intDOT* or among any of the key catalytic residues (Kim et al. 2010).

Regulatory mechanism of representative CTnDOT-like elements

CTnDOT has been a model for studying the mechanisms governing the regulation, excision, and mobilization of CTn elements among gut microbes (Whittle et al. 2001, Cheng et al. 2000, Whittle et al. 2002, Keeton et al. 2013, Hopp et al. 2015, Waters et al. 2013, Wood & Gardner 2015). In addition to PmDOT, reanalysis of previous MGE predictions (Frye et al. 2022) identified a set of putative CTnDOT-like elements among the human gut associated microbes in our lab strain collection (Table 3.5). These strains included: *Ph. dorei* DSM 17855, *Ph. coprocola* DSM 17136, *Ph. stercoris* ATCC 43183, *Ph. intestinalis* DSM 17393, *Ph. plebeius* DSM 17135, *B. eggerthii* DSM 20697, *P. johnsonii* DSM 18315, and *B. thetaiotaomicron* 3731. All these MGEs encoded *tetQ*, appear largely intact having been identified on 1 contig or found on the ends to two adjacent contigs, and showed substantial sequence similarity when aligned to CTnDOT (Figure 3.5).

Although these strains exhibit an average 91–95% nucleotide identity to CTnDOT in aligned regions (Figure 3.5), only CTnDOT contains the ~12.8kb *ermF* integron. However, the majority of the additional CTns (7/9) encode all the primary genes responsible for regulating the excision and mobilization processes within CTnDOT (Table 3.1). The two remaining CTnDOT-like elements from *Ph. dorei* and *P. johnsonii*,

diverge in this regard. In the case of the *Ph. dorei* CTn, all the genes upstream of *tetQ-rteA-rteB* operon in CTnDOT are missing and replaced with proteins of unknown function (Figure 3.5). Although the *Ph.* dorei CTn shares ≥50% amino acid identity with 3/7 CTnDOT regulatory proteins, it lacks the entire excision operon (*xis2c-xis2d-orf3-exc*) and *intDOT*. The absence of *exc* is expected to impede the ability of the *Ph.* dorei CTn to undergo excision. Furthermore, Xis2c, Xis2d and Exc collectively play a crucial role in activating the transcription of the *tra* and *mob* operons, which are responsible for forming the T4SS structure and facilitating CTn mobilization, respectively. Despite retaining intact *tra* genes, we anticipate that *Ph. dorei* CTn lacks the necessary genes for mobilization. Also, *Ph. dorei* differs in size compared to CTnDOT, ~47kb as opposed to CTnDOT's ~64kb.

Similarly, *P. johnsonii* CTn is even smaller, sized at ~31kb. Nonetheless, the *P. johnsonii* CTn retains the essential genes required for T4SS formation, excision, and mobilization of the CTn. However, the *P. johnsonii* CTn is missing *intDOT*, *rteA*, and *rteB*. The primary role of RteB is activating the downstream gene *rteC*, a pivotal step in the regulatory cascade governing the transcription of the excision T4SS, and mobilization operons. Nevertheless, *in vitro* conjugation experiments involving *P. johnsonii* (Tet^R) and *B. thetaiotaomicron* VPI-5482 (Erm^R) yielded dual-resistant colonies, indicating that the *P. johnsonii* CTn is capable of mobilizing into *B. thetaiotaomicron* (Table 3.6). This would suggest that the regulatory mechanisms governing the mobilization of the *P. johnsonii* CTn significantly differ from those of CTnDOT.

Tetracycline-dependent regulation of the excision and transfer of CTnDOT by the *tetQ-rteA-rteB* operon is mediated by translational attenuation of the *tetQ* 5' untranslated region (UTR) (Wang et al. 2004). We aligned the *tetQ* 5' UTR for all CTnDOT-like elements and found there were virtually identical. Only two nucleotide substitutions were detected in the *tetQ* 5' UTR of our CTnDOT sequence, the *P. johnsonii* CTn, and the *B. thetaiotaomicron* 3731 CTn respectively when compared to the published CTnDOT sequence (Wang et al. 2005). However, based on their positions these mutations are unlikely to impact the RNA secondary structure crucial for the translational attenuation as they are situated within loops and not in hairpin stems, implying that the structure is preserved (Wang et al. 2005). Furthermore, CTnDOT-like elements exhibit complete conservation of the -7 site, the -33 site, transcription initiation site, and the 3 amino acid leader peptide region. Together this suggests that all these CTnDOT-like elements are poised to exhibit identical or similar tetracycline-dependent regulation.

Core genes of CTnDOT-like elements

Conjugative transposons, like the genomes they integrate into are subject to rearrangements, recombination, gene losses and gains. Therefore, to define the key genes associated with these 10 CTnDOT-like elements and to facilitate detection of additional CTnDOT-like elements, we conducted a core and pan genome analysis. We identified 10 universally conserved core genes that define this family of CTns: *tetQ*, *traA*, *traC, traD, traG, orf1, orf2, rteC*, and two uncharacterized genes (*BTH032404* and *BTH033992*) (Figure 3.2A). As noted above, the missing regulatory and integrase genes from the *Ph.*

dorei and *P. johnsonii* CTns (Table 3.1) results in their absence from the core list. It follows that the list of accessory genes among these CTns is substantially larger with 137 across the 10 CTns. This is indicative of the large role rearrangements, recombination, gene losses and gains play in CTn evolution (Figure 3.2B, Table 3.7). Further it turns out that the core genes are in fact centrally located within the CTnDOT element, all localized within the central 25 kb region of the 64 kb element (Figure 3.2C).

Diversity and ubiquity of CTnDOT-like elements

Using the CTnDOT core gene panel we identified an additional 67 predicted CTnDOT-like elements from a panel of 124 Bacteroidota genome assemblies (Figure 3.3, Table 3.5). The inclusion criteria required the presence of ≥6 core genes (81% have all 10 core genes) and that the CTn appeared intact (see methods). Final CTn boundaries were determined using genome alignments and a network analysis was performed clustering all 77 CTns based on their percentage length aligned (PLA) (as in Frye et al. 2021, Ortañez & Degnan 2024). Using this method, we found that all the CTnDOT-like elements clustered as a single family while outgroup CTns, *Pv*CTn and B12-CTns clustered separately (Figure 3.3A). We subsequently filtered the PLA values to ≥90% and re-clustered the CTns identifying 10 subfamilies with 2 or more members and 15 singletons. Like CTnDOT the majority of the CTns (70/77) confer tetracycline resistance to their host strains, encoding *tetQ* alleles ($\bar{\mathbf{x}}$ ~ 97% amino acid identity). This includes all members of the four largest subfamilies of sequence similar CTns (Nos. 1–4). We cross referenced the presence of the CTns in genomes with when the host strains were

collected (when available) and found that three of these four clusters have been circulating among gut microbe strains for over 5 decades (Figure 3.3B).

Unlike tetracycline resistance, the other resistance alleles encoded by CTnDOT are not as widely conserved (*aadS, tetX1, tetX2,* and *ermF*). In fact, the specific integron encoding these alleles is only found in four other members of the CTnDOT subfamily cluster (No. 2). Despite lacking this particular erythromycin integron, we found that six CTnDOT-like elements also confer dual tetracycline and erythromycin resistance. While the *tetQ* allele remains in a syntenic location, these CTns have distinct integrated regions encoding *ermF* alleles with 97–98% amino acid identity to *ermF* from CTnDOT (Figure 3.3C).

Further characterization of gene content differences among the CTnDOT-like elements primarily identified novel integrated insertion sequences and group II introns. We also detected CTns encoding a phage lysozyme (PF00959), a two gene operon encoding a SnoaL-like domain gene (PF12680) and an acetyltransferase (GNAT) family gene (PF00583) immediately upstream of some *tetQ* alleles, and homologs of the thoeris operon, which is a bacterial antiphage defense system (PF08937, PF18185) (Ka et al. 2020).

As expected, the distribution of CTnDOT-like elements among species and strains is consistent with frequent horizontal transfer (Figure 3.3D). We note that six strains encode 2 separate CTnDOT-like elements, and in four of these cases both encode a *tetQ* allele. Moreover, in several of the densely sampled species such as *B. fragilis, B. xylanisolvens* and *Ph. dorei*, CTnDOT-like elements are very common and we note that

they are often from distinct subfamilies (Figure 3.3D). Although we successfully identified 77 intact CTns, 19 additional partial/inactivated CTnDOT-like elements were detected. These elements had fewer than 6 core genes and/or appeared highly fragmented due to apparent genome-level recombination or incomplete genome assembly (Figure 3.3D). We note that additional strains were identified that encoded *tetQ* or *ermF* alleles that were not associated with CTnDOT-like elements.

CTnDOT-like elements are globally distributed

The genome sequenced strains analyzed for the presence of CTnDOT-like elements represent a collection of individual representatives of much larger microbial populations. As such we investigated the prevalence of CTnDOT-like elements among densely sampled human gut metagenomic datasets (Table 3.4). These samples were chosen to represent diverse cultural and geographic populations including both industrialized and hunter-gatherer populations. As expected, evidence of gut Bacteroidota based on the marker gene *rpsE* was confirmed in every sample, with levels often varying among samples in a dataset by 2 to 3 orders of magnitude (Figure 3.4A). Unexpectedly, short read evidence of CTnDOT-like elements were detected in every dataset and sample examined regardless of their lack of industrialization (Figure 3.4A). Further, after RPKM normalization we found that the average ratio of CTnDOT-like read coverage was ≥1 in all populations indicating the prevalence of these types of elements in gut microbiomes.

It is possible that the prevalence of the CTnDOT-like elements in these samples could be a result of a single highly repetitive gene region (e.g., integrase, or insertion

sequence). To address this, we investigated the read coverage mapping of the metagenomic data across the length of the CTnDOT-like elements. Comparisons of read coverage mapping of samples from an industrialized and hunter-gatherer population, clearly show conservation of the CTnDOT backbone (Figure 3.4B). This pattern of conservation is similar to the pattern observed in the alignment of *ermF* encoding CTnDOT-like elements detected above (Figure 3.3C). Furthermore, these coverage heatmaps highlight evidence for the *ermF* integron, as well as all of the CTnDOT antibiotic resistance alleles in many of the individual patient samples (Figure 3.4B).

In addition to the healthy patient gut metagenomes we analyzed, the industrialized datasets also included paired samples that were treated with antibiotics (USA1, DNK, CAN, CHN1) or that were colorectal cancer patients (AUT, CHN2, USA2) (Table 3.4). Given that CTnDOT conjugation activity is mediated by tetracycline exposure (Wang et al. 2005), we speculated that perhaps other CTnDOT-like elements may respond to different antibiotics or disease induced dysbiosis leading to changes in the ratio of CTnDOT-like RPKM to *rpsE* RPKM. However, no such general trends were detected within patients before and after antibiotic exposure or between healthy and CRC patients (e.g., Figure 2.8). Therefore, it remains unclear if other environmental signals might be responsible for changes in conjugation activity of CTnDOT-like elements.

DISCUSSION

The spread of antibiotic resistance genes is widely recognized to be a global concern (CDC 2023, WHO 2023). For example, previous reports identified increases in CTnDOT-like elements that spread tetracycline and erythromycin resistance, which corresponded to prescription of the same antibiotics (Shoemaker et al. 2001). Although these two antibiotics are generally no longer frontline drugs for human infections in many industrialized countries, these CTnDOT-like elements persist in gut microbiomes (Figure 3.3, Figure 3.4). Here, we have characterized the extent of conservation and variation among these CTnDOT-like elements among individual strains and human gut metagenomes. We observed the persistence of these MGEs and their associated antibiotic resistance genes among diverse gut microbes from globally disparate human populations.

Redefining the CTnDOT-like family

Mobile genetic elements are inherently difficult to organize and classify due to the frequency of their recombination and reassortment. For instance, the mixing of tyrosine recombinase (integrase) has been observed in phylogenies within and between ICE and prophage groups, suggesting a pattern of integrase exchange between ICEs and prophages throughout evolutionary history (Jiang et al. 2019, Ortañez & Degnan 2024). Here we have used an approach that focuses on patterns of overall proportion of DNA conservation among MGEs (Frye et al. 2021, Ortañez & Degnan 2024). As a result, we identified 77 intact and 19 additional partial/inactivated CTnDOT-like elements among a

panel of 134 gut bacterial strains. We note that previous authors have characterized a smaller number of MGEs identified at the time that shared homology with CTnDOT (Whittle et al. 2002, Liu et al. 2019). For example, in the ICEberg 2.0 database CTnDOT is grouped with 6 other integrative and conjugative elements (ICEs) based on the presence of ≥60% protein homology to *intDOT* (Liu et al. 2019). While our analysis recapitulated the grouping of CTnDOT with the CTns from *B. fragilis* YCH46, *B. salanitronis* DSM 18170, and *B. xylanisolvens* XB1A, the *Poryphyromonas gingivalis* CTnPg1-a was not clustered with CTnDOT. Although this CTn encodes a homolog of *intDOT,* we found that it had no more than 2 – 14% percent length aligned (PLA) with any of the 77 CTnDOT-like family members in our study (Figure 3.3A). Thus, our analysis that relies on the overall proportion of DNA conservation indicates that CTnPg1-a should be classified as a member of a distinct CTn family.

Furthermore, several CTns recently detected among clinical isolates of *B. fragilis* meet our definition of CTnDOT-like elements (Boiten et al. 2023, Vineis et al. 2024). All of these CTns were noted to have extensive sequence homology with CTnDOT including the presence of *tetQ*. One CTn, CTn214 was recovered from a *B. fragilis* metagenome-assembled genome from a patient that developed intestinal inflammation (pouchitis)(Vineis et al. 2024). This CTn was found to exist both in an integrative form as well as a plasmid-like circular form and exhibited tetracycline-mediated gene regulation consistent with CTnDOT (Vineis et al. 2024). The second set of these CTns were found in the genomes of multidrug resistant (MDR) strains of *B. fragilis* isolated between 2014 and 2020 (Boiten et al. 2023). While these strains were isolated due to their resistance to

frontline antibiotics meropenem and metronidazole, they encoded multiple other resistance genes including *tetQ*. Together these data appear to confirm that CTnDOT-like elements persist in gut associated Bacteroidota acting as a reservoir for tetracycline resistance.

Supporting this notion, a study utilizing Split Read Insertion Detection (SRID) to identify MGEs revealed that the majority of CTnDOT-like elements harbor resistance to tetracycline via the *tetQ* gene (Jiang et al. 2019). Consistent with these findings, our investigation found that 70/77 CTnDOT-like elements retained an intact *tetQ* gene (Figure 3.6). The SRID study also revealed diverse insertion sites for antibiotic resistance in CTnDOT-like elements, demonstrating varied insertions and substitutions across five sites (Jiang et al. 2019). Although our analysis did not capture the full spectrum of antibiotic resistance gene diversity reported by Jiang et al. 2019, it did reveal the presence of the streptomycin resistance gene (*aadS*) in 5/77 CTnDOT-like elements, previously unconfirmed in CTnDOT. These findings highlight the broad antibiotic resistance gene diversity in CTnDOT-like elements shaped by recombination and reassortment.

Global evidence of CTnDOT-like elements in human fecal metagenomes

While genome sequences can provide snapshots of individual strains, fecal metagenomes can provide semi quantitative measures of gene frequencies from entire communities. Using this approach, we detected widespread evidence of CTnDOT-like elements among a diverse panel of human microbiomes (Figure 3.4A, Table 3.4). While CTnDOT is known to be induced by tetracycline exposure (Wang et al. 2005), one might

assume its heightened prevalence in antibiotic exposed industrial societies compared to hunter-gatherer communities. Surprisingly, our data revealed a comparable or even greater abundance of CTnDOT-like elements in hunter-gatherer communities than in industrialized populations (Figure 3.4A). Further, patient samples from the hunter-gatherer communities showed clear evidence of antibiotic resistance genes *aadS, tetX, ermF* and *tetQ* (Figure 3.4B).

Together, this suggests that although tetracycline is a primary inducer of the CTnDOT excision and mobilization regulatory cascade in *in vitro* tests, there are possibly other natural factors playing roles in its spread and persistence in human gut-associated bacteria (Wang et al. 2005). This may include the acquisition and dissemination of other genes that can improve the fitness of their bacterial host (e.g., antiphage defense systems). Interestingly, CTnDOT was previously observed to be transferrable to *E. coli* (Li et al. 1995). However, we searched 562 non-Bacteroidota gut genomes with the CTnDOT core genes and did not identify any evidence of intact or partial CTnDOT-like elements. This suggests that CTnDOT-like elements in nature predominantly mobilize among gut Bacteroidota, which is consistent with previous observations of HGT in gut microbes (Smillie et al. 2011).

Activity and Regulation of CTnDOT-like elements

Computational studies can rapidly identify putative MGEs, however, without further experiments it is generally unclear which CTns are functional. Further, some CTns identified as being fragmented may be part of chimeric MGEs or with improved

assembly prove to be intact. Using the Tn mutagenesis mobilization method (TMMM) (Ortañez & Degnan 2024) we detected the mobilization of PmDOT that we determined was a novel CTnDOT-like element present in *P. merdae*.

PmDOT encodes all of the regulatory genes demonstrated to be responsible for the excision and conjugative transfer of CTnDOT (*rteB, rteC, xis2c, xis2d, exc, intDOT*). A notable difference detected in PmDOT is a two-gene operon, found in 29/30 PmDOT family variants (Subfamily 1), immediately downstream of the *tetQ-rteA-rteB* operon, which includes a KAP family P-loop domain protein (PARMER00793). KAP nucleoside-triphosphatases (NTPases) are commonly found in plasmids and are suggested to play a role in in protecting plasmid-carrying bacterial cells from bacteriophages (Aravind et al. 2004). While expression of these genes in *P. merdae* has not been confirmed, the possible protective function involving regulation of the NTP-dependent assembly or disassembly of membrane associated protein complexes (Aravind et al. 2004) may be beneficial for host bacteria that harbor Subfamily 1 CTns. This potential increase in fitness might contribute to the prevalence of Subfamily 1 compared to CTnDOT and Subfamily 2 among the panel of human gut-associated bacterial genomes examined (Figure 3.3B).

CTnDOT mobilization is intricately controlled through multiple layers of gene regulation (Waters and Salyers 2013). Here we have highlighted the distinctive regulatory elements among a subset of CTnDOT-like elements which suggests potential regulatory diversity among the broader CTnDOT-like family. For example, in *Ph. dorei* the lack of *xis2c, xis2d, orf3, exc*, and *intDOT* would indicate that this CTn lacks the essential genes for complete excision and mobilization. Initial *in vitro* tests of *Ph. dorei* CTn

mobilization into *B. thetaiotaomicron* supports this hypothesis, as no transconjugants were detected (Table 3.6). However, it is plausible that *B. thetaiotaomicron* and *Ph. dorei* are conjugatively incompatible, or that *Ph. dorei* CTnDOT-like element requires distinct environmental signals to induce mobilization (e.g., not tetracycline). Conversely, *P. johnsonii* CTnDOT, despite lacking orthologs of *intDOT*, *rteA*, and *rteB* genes, demonstrated *in vitro* mobilization when mated with *B. thetaiotaomicron* (Table 3.6). This strongly suggests that the regulatory mechanisms of *P. johnsonii* CTnDOT differ in part from those of CTnDOT. Further investigation is needed to fully detail these specific regulatory mechanisms.

Conclusion

Given the importance of horizontal gene transfer (HGT) for spreading antibiotic resistance in human associated microbes, characterizing the conservation and variation of the MGEs responsible for HGT is important. We find that the CTnDOT-like family of MGEs is widespread and diverse, occurring in ~50% of Bacteroidota isolates examined and the vast majority encoding one or more antibiotic resistance genes. Further, these CTnDOT-like elements are found globally in gut microbes of both industrialized as well as indigenous populations. Therefore, continued studies of these persistent reservoirs of antibiotic resistance, the risk of their acquisition of novel antibiotic resistance genes and potential strategies to interfere with the transmission of these elements are warranted.

FIGURES

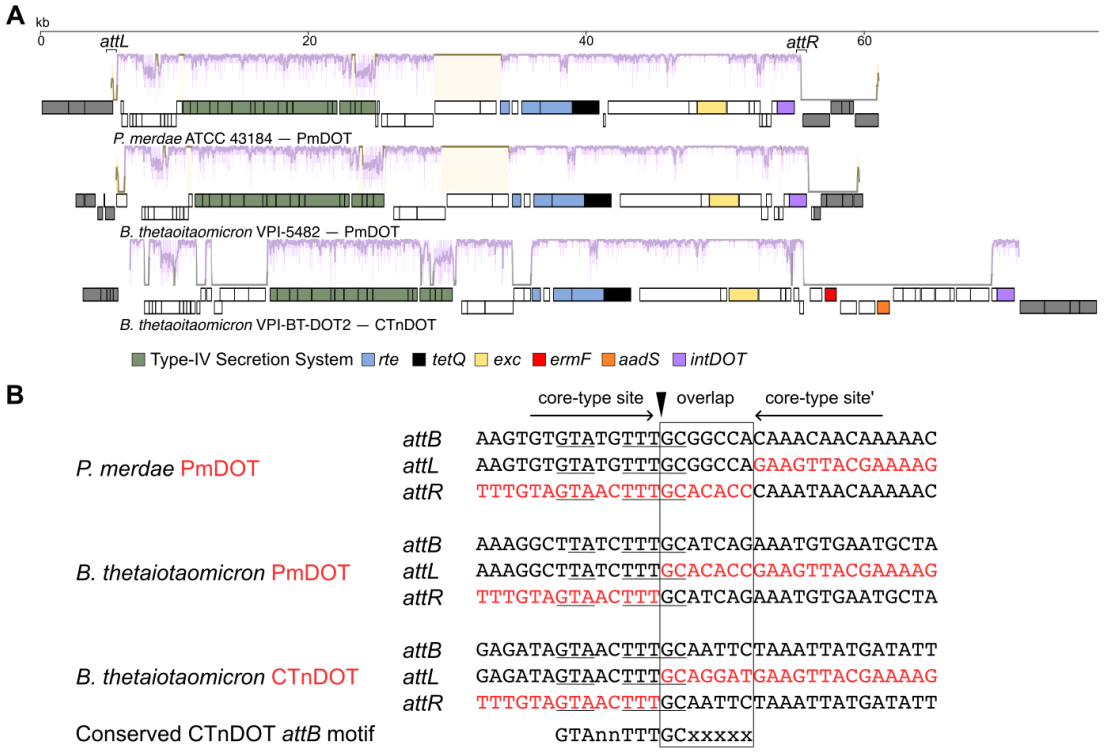


Figure 3.1. CTnDOT and PmDOT have variable and conserved regions
(A) Nucleotide alignment of PmDOT, *B. thetaiotaomicron* PmDOT, CTnDOT, and their flanking regions, aligned by Mauve. Plots indicate nucleotide sequence identity among the CTns ranging from 50 (baseline) to 100% (top). Purple regions represent locally colinear blocks present in all three CTns, while the tan regions are only encoded by PmDOT. Notable genes are highlighted by the color-coded key. Gray genes represent flanking chromosomal regions. (B) Alignments of PmDOT and CTnDOT MGE attachment sequences exhibit conservation of the IntDOT binding motif. Black letters indicate the bacterial target site, and red letters indicate the integrated ends of the CTn sequences. The conserved IntDOT binding motif (Cheng et al. 2000) found in the 'core-type site' is shown with the underlined letters and correspond to the imperfect direct repeats flanking the integrated element. The black triangle indicates the location of the first IntDOT mediated cut in the top strand and the result of the second staggered cut is shown by the 'overlap' box that extends through the 5bp coupling site (Xs) (Wood and Gardner 2015).

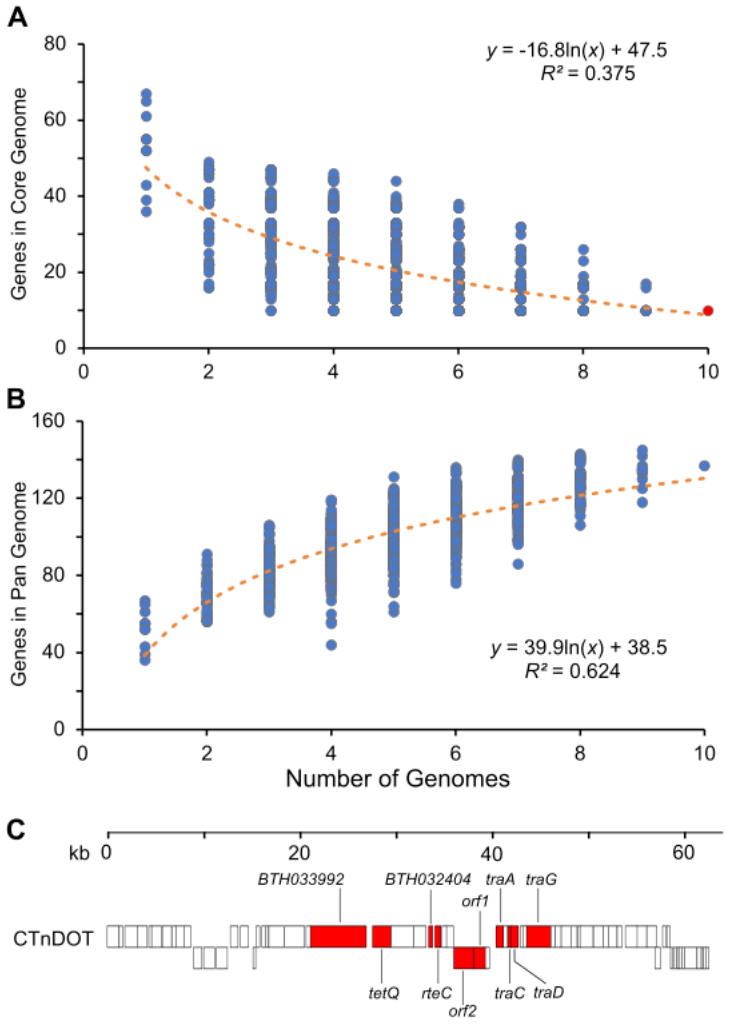


Figure 3.2. Core and pan genome of CTnDOT-like elements. Evaluation of (A) Core and (B) Pan genome of CTnDOT-like elements using 10 analyzed CTnDOT-like elements. (C) Schematic of CTnDOT with the 10 core genes indicated in red.

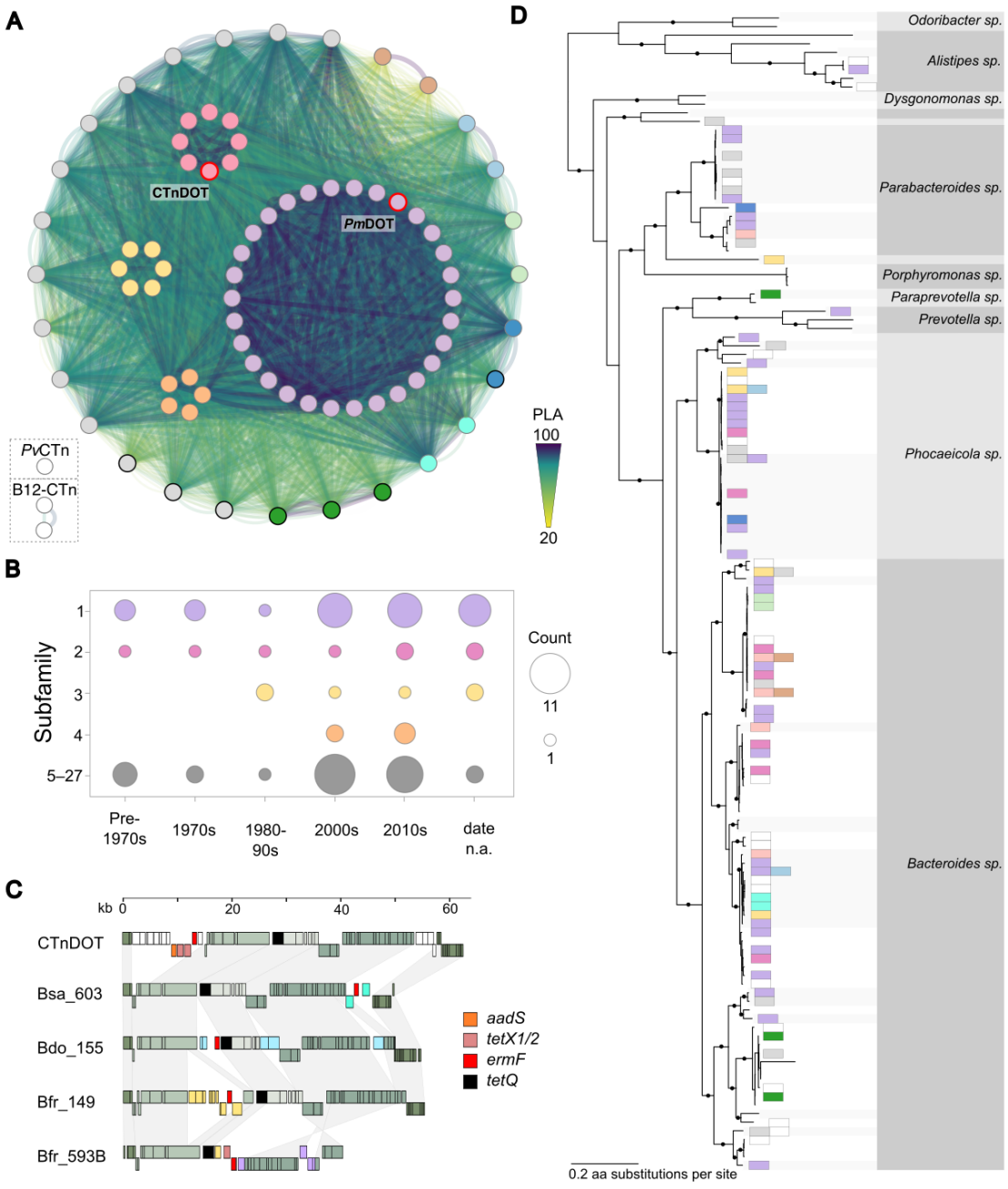

A
CTnDOT
PmDOT
PvCTn
B12-CTn
PLA
100
20
B
Subfamily
1
2
3
4
5–27
Pre-1970s
1970s
1980-90s
2000s
2010s
date n.a.
Count
11
1
C
kb 0
20
40
60
CTnDOT
Bsa_603
Bdo_155
Bfr_149
Bfr_593B
aadS
tetX1/2
ermF
tetQ
D
Odoribacter sp.
Alistipes sp.
Dysgonomonas sp.
Parabacteroides sp.
Porphyromonas sp.
Paraprevotella sp.
Prevotella sp.
Phocaeicola sp.
Bacteroides sp.
0.2 aa substitutions per site

Figure 3.3. Core gene detected CTnDOT-like elements are prevalent across human gut-associated bacteria genomes. (**A**) Network schematic of identified CTnDOT-like family of elements identified screening genomes for core CTnDOT genes and clustered based on percentage length aligned (PLA). Nodes represent individual predicted CTns and edges represent PLA values ≥20%. Subfamilies among CTns at 90% PLA are shown as distinct colored nodes and grey nodes represent singletons. CTnDOT and PmDOT are labeled and circled in red, while CTns with black borders represent CTns lacking a *tetQ* gene. Edge widths, colors and transparency indicate PLA (20% =yellow, narrow, near transparent; 100% blue, thick, solid). Labeled network can be found in **Figure 3.6**. (**B**) Bubble plot indicating the number of strains collected during the indicated time intervals that encode CTns forming 90% PLA subfamilies. (**C**) An alignment of representative *ermF-tetQ* encoding CTns illustrates the variable insertion locations of the 97–98% amino acid identical *ermF* alleles (red). CTnDOT backbone genes are shown in shades of green, antibiotic genes are colored according to the key and unique genes are shown as distinct colors. (**D**) Species and strain phylogeny of gut associated Bacteroidota screened for CTnDOT-like elements. Bootstrap values ≥75 are shown as black dots (•) on the relevant nodes of the maximum likelihood phylogeny. Colored rectangles next to the branch tips indicate the presence of one or more CTnDOT-like element and correspond to the 90% PLA cluster colors in **A**. White rectangles indicate fragmented CTnDOT-like elements not included in the network analysis. Labeled phylogeny can be found in **Figure 3.7**.

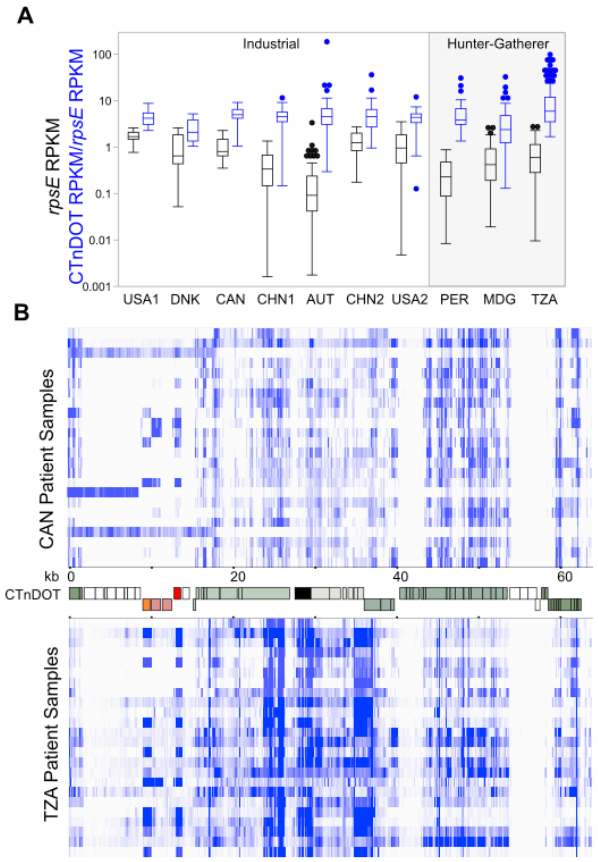


Figure 3.4. Detection of CTnDOT-like elements in diverse metagenomes.
(**A**) Bacteroidota and CTnDOT-like elements were universally detected in metagenome reads from 8 geographically distinct human populations. Box and whisker plots are shown for Bacteroidota (*rpsE* RPKM) detected and the ratio of CTnDOT-like elements to Bacteroidota detected per sample (blue). (**B**) Representative heatmaps show CTnDOT read coverage from 24 patient samples from an industrialized population (CAN) and a hunter-gatherer population (TZA). Read coverage for each sample was normalized per 1 million reads mapped, and regions with ≥ 5,000X read coverage are solid blue. Coloring of CTnDOT schematic is the same as in **Figure 3.3**.

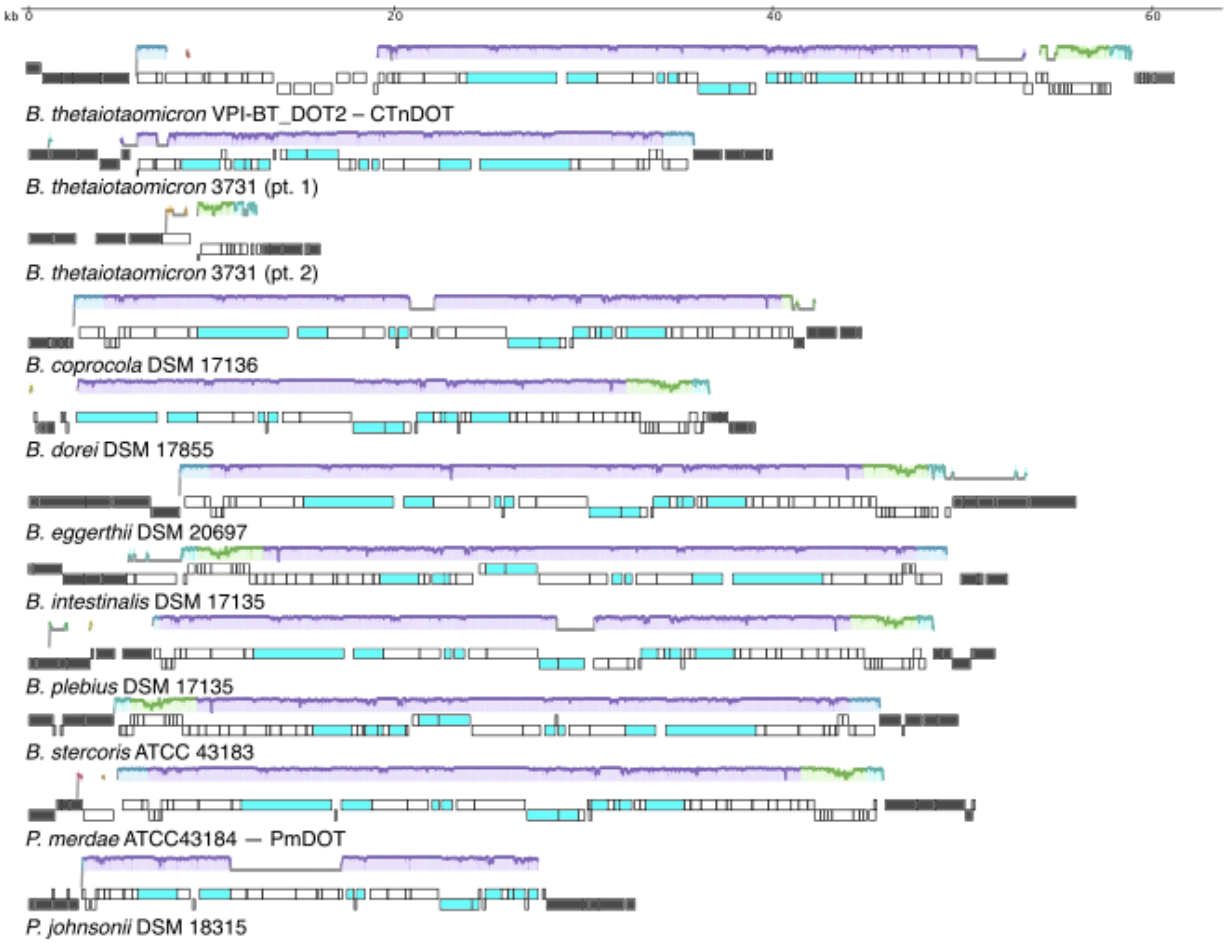


Figure 3.5. Alignment of a representative set of CTnDOT-like elements. Mauve genome alignment of CTnDOT with other CTnDOT-like elements. Teal genes represent core genes and genes in flanking regions are shown in gray. Plots indicate nucleotide sequence identity among the CTns ranging from 50 (baseline) to 100% (top).

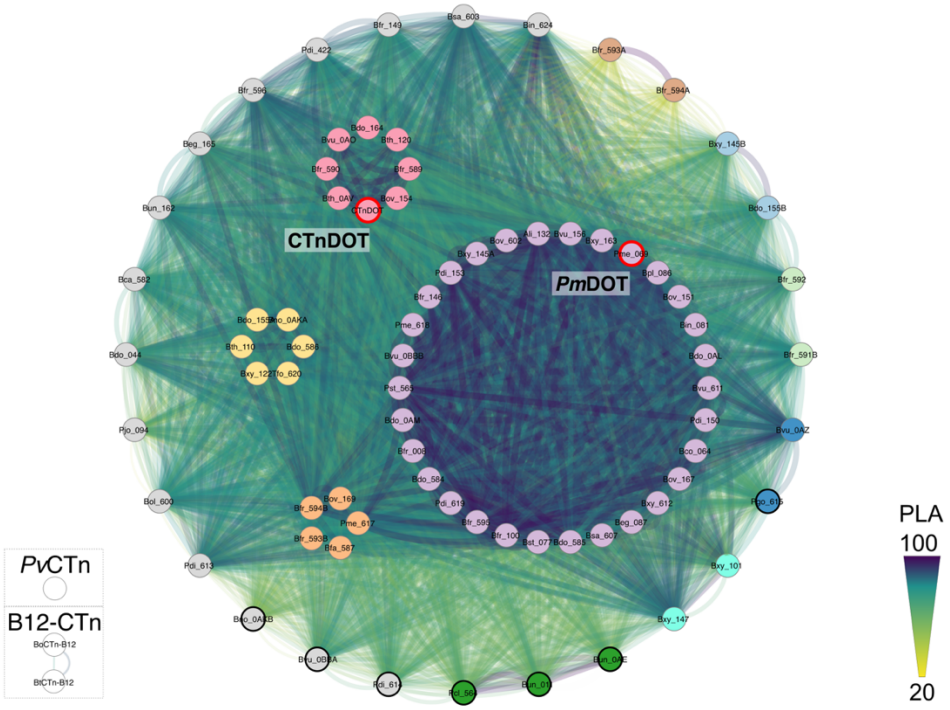


Figure 3.6. Labeled network of schematic of identified CTnDOT-like family of elements. Identified CTnDOT-like elements and indicated outgroups were clustered based on percentage length aligned (PLA). Nodes represent individual predicted CTns and edges represent PLA values ≥20%. Subfamilies among CTns at 90% PLA are shown as distinct colored nodes and grey nodes represent singletons. CTnDOT and PmDOT are labeled and circled in red, while CTns with black borders represent CTns lacking a *tetQ* gene. Edge widths, colors and transparency indicate PLA (20% =yellow, narrow, near transparent; 100% blue, thick, solid). Labels correspond to those found in Table 3.5.

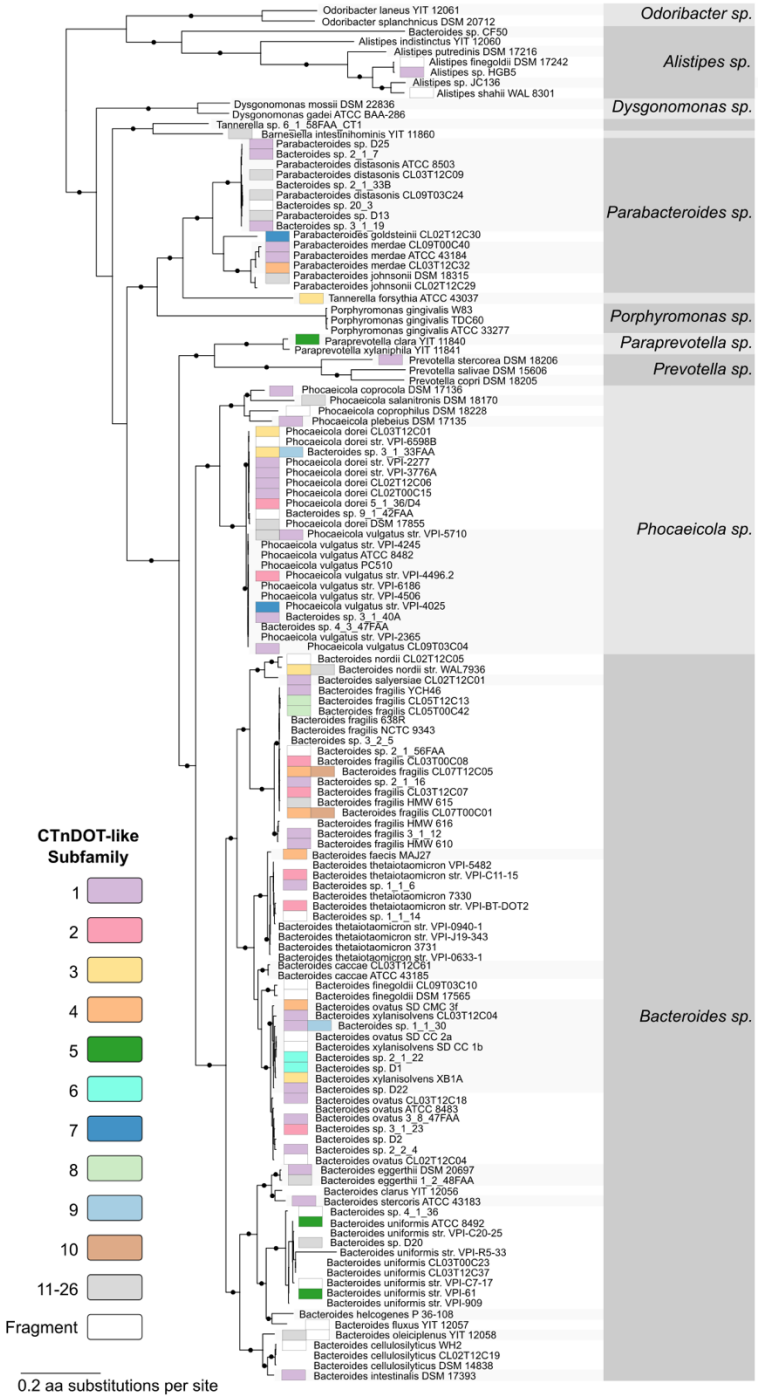

Odoribacter laneus YIT 12061
Odoribacter splanchnicus DSM 20712
Odoribacter sp.
Bacteroides sp. CF50
Alistipes indistinctus YIT 12060
Alistipes putredinis DSM 17216
Alistipes finegoldii DSM 17242
Alistipes sp. HGB5
Alistipes sp. JC136
Alistipes shahii WAL 8301
Alistipes sp.
Dysgonomonas mossii DSM 22836
Dysgonomonas gadei ATCC BAA-286
Dysgonomonas sp.
Tannerella sp. 6_1_58FAA_CT1
Barnesiella intestinihominis YIT 11860
Parabacteroides sp. D25
Bacteroides sp. 2_1_7
Parabacteroides distasonis ATCC 8503
Parabacteroides distasonis CL03T12C09
Bacteroides sp. 2_1_33B
Parabacteroides distasonis CL09T03C24
Bacteroides sp. 20_3
Parabacteroides sp. D13
Bacteroides sp. 3_1_19
Parabacteroides goldsteinii CL02T12C30
Parabacteroides merdae CL09T00C40
Parabacteroides merdae ATCC 43184
Parabacteroides merdae CL03T12C32
Parabacteroides johnsonii DSM 18315
Parabacteroides johnsonii CL02T12C29
Parabacteroides sp.
Tannerella forsythia ATCC 43037
Porphyromonas gingivalis W83
Porphyromonas gingivalis TDC60
Porphyromonas gingivalis ATCC 33277
Porphyromonas sp.
Paraprevotella clara YIT 11840
Paraprevotella xylaniphila YIT 11841
Paraprevotella sp.
Prevotella stercorea DSM 18206
Prevotella salivae DSM 15606
Prevotella copri DSM 18205
Prevotella sp.
Phocaeicola coprocola DSM 17136
Phocaeicola salanitronis DSM 18170
Phocaeicola coprophilus DSM 18228
Phocaeicola plebeius DSM 17135
Phocaeicola dorei CL03T12C01
Phocaeicola dorei str. VPI-6598B
Bacteroides sp. 3_1_33FAA
Phocaeicola dorei str. VPI-2277
Phocaeicola dorei str. VPI-3776A
Phocaeicola dorei CL02T12C06
Phocaeicola dorei CL02T00C15
Phocaeicola dorei 5_1_36/D4
Bacteroides sp. 9_1_42FAA
Phocaeicola dorei DSM 17855
Phocaeicola vulgatus str. VPI-5710
Phocaeicola vulgatus str. VPI-4245
Phocaeicola vulgatus ATCC 8482
Phocaeicola vulgatus PC510
Phocaeicola vulgatus str. VPI-4496.2
Phocaeicola vulgatus str. VPI-6186
Phocaeicola vulgatus str. VPI-4506
Phocaeicola vulgatus str. VPI-4025
Bacteroides sp. 3_1_40A
Bacteroides sp. 4_3_47FAA
Phocaeicola vulgatus str. VPI-2365
Phocaeicola vulgatus CL09T03C04
Phocaeicola sp.
Bacteroides nordii CL02T12C05
Bacteroides nordii str. WAL7936
Bacteroides salyersiae CL02T12C01
Bacteroides fragilis YCH46
Bacteroides fragilis CL05T12C13
Bacteroides fragilis CL05T00C42
Bacteroides fragilis 638R
Bacteroides fragilis NCTC 9343
Bacteroides sp. 3_2_5
Bacteroides sp. 2_1_56FAA
Bacteroides fragilis CL03T00C08
Bacteroides fragilis CL07T12C05
Bacteroides sp. 2_1_16
Bacteroides fragilis CL03T12C07
Bacteroides fragilis HMW 615
Bacteroides fragilis CL07T00C01
Bacteroides fragilis HMW 616
Bacteroides fragilis 3_1_12
Bacteroides fragilis HMW 610
Bacteroides faecis MAJ27
Bacteroides thetaiotaomicron VPI-5482
Bacteroides thetaiotaomicron str. VPI-C11-15
Bacteroides sp. 1_1_6
Bacteroides thetaiotaomicron 7330
Bacteroides thetaiotaomicron str. VPI-BT-DOT2
Bacteroides sp. 1_1_14
Bacteroides thetaiotaomicron str. VPI-0940-1
Bacteroides thetaiotaomicron str. VPI-J19-343
Bacteroides thetaiotaomicron 3731
Bacteroides thetaiotaomicron str. VPI-0633-1
Bacteroides caccae CL03T12C61
Bacteroides caccae ATCC 43185
Bacteroides finegoldii CL09T03C10
Bacteroides finegoldii DSM 17565
Bacteroides ovatus SD CMC 3f
Bacteroides xylanisolvens CL03T12C04
Bacteroides sp. 1_1_30
Bacteroides ovatus SD CC 2a
Bacteroides xylanisolvens SD CC 1b
Bacteroides sp. 2_1_22
Bacteroides sp. D1
Bacteroides xylanisolvens XB1A
Bacteroides sp. D22
Bacteroides ovatus CL03T12C18
Bacteroides ovatus ATCC 8483
Bacteroides ovatus 3_8_47FAA
Bacteroides sp. 3_1_23
Bacteroides sp. D2
Bacteroides sp. 2_2_4
Bacteroides ovatus CL02T12C04
Bacteroides eggerthii DSM 20697
Bacteroides eggerthii 1_2_48FAA
Bacteroides clarus YIT 12056
Bacteroides stercoris ATCC 43183
Bacteroides sp. 4_1_36
Bacteroides uniformis ATCC 8492
Bacteroides uniformis str. VPI-C20-25
Bacteroides sp. D20
Bacteroides uniformis str. VPI-R5-33
Bacteroides uniformis CL03T00C23
Bacteroides uniformis CL03T12C37
Bacteroides uniformis str. VPI-C7-17
Bacteroides uniformis str. VPI-61
Bacteroides uniformis str. VPI-909
Bacteroides helcogenes P 36-108
Bacteroides fluxus YIT 12057
Bacteroides oleiciplenus YIT 12058
Bacteroides cellulosilyticus WH2
Bacteroides cellulosilyticus CL02T12C19
Bacteroides cellulosilyticus DSM 14838
Bacteroides intestinalis DSM 17393
Bacteroides sp.
CTnDOT-like Subfamily
1
2
3
4
5
6
7
8
9
10
11-26
Fragment
0.2 aa substitutions per site

Figure 3.7. Detailed species phylogeny of gut Bacteroidota with CTnDOT-like elements. Maximum likelihood phylogeny of strains of gut associated Bacteroidota screened for CTnDOT-like elements. Bootstrap values ≥75 are shown as black dots (•) on the relevant nodes. Colored rectangles next to the branch tips indicate the presence of one or more CTnDOT-like element and correspond to the subfamily (90% PLA cluster) colors in noted in the key. White rectangles indicate partial or highly fragmented CTnDOT-like elements not included in the network analysis.

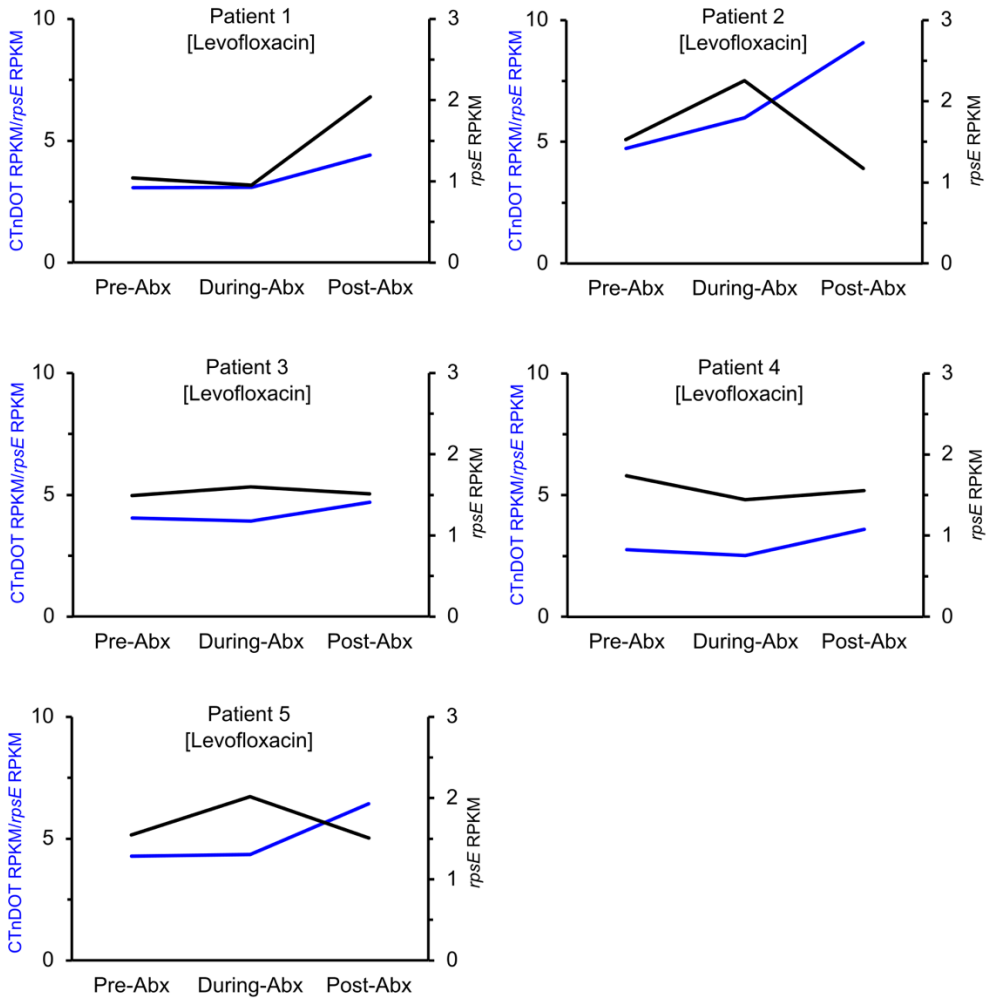


Figure 3.8. CTnDOT-like counts in antibiotic treated metagenomes. Example plots of *rpsE* and *rpsE* normalized CTnDOT RPKM values before, during, and after Levofloxacin antibiotic perturbance for five patients (Anthony et al. 2022). Although some trends were detected, none were significant in this or the other studies examined.

TABLES

Table 3.1 Conservation of regulatory proteins in CTnDOT-like elements

Strain	***exc***	***intDOT***	***rteA***	***rteB***	***rteC***	***xis2c***	***xis2d***	***aadS***	***ermF***	***tetQ***
CTnDOT	+	+	+	+	+	+	+	+	+	+
PmDOT	+	+	+	+	+	+	+	–	–	+
Ph. dorei DSM 17855	–	–	+	+	+	–	–	–	–	+
Ph. coprocola DSM 17136	+	+	+	+	+	+	+	–	–	+
B. stercoris ATCC 43183	+	+	+	+	+	+	+	–	–	+
B. intestinalis DSM 17393	+	+	+	+	+	+	+	–	–	+
Ph. plebeius DSM 17135	+	+	+	+	+	+	+	–	–	+
B. eggerthii DSM 20697	+	+	+	+	+	+	+	–	–	+
P. johnsonii DSM 18315	+	–	–	–	+	+	+	–	–	+
B. thetaiotaomicron 3731	+	+	+	+	+	+	+	–	–	+

Table 3.2 Bacterial strains used in this study

Strain Name	Short name in text	Notes	Source and/or reference
Escherichia coli S17-1λpir pSAM_BC2_ermG		Barcoded donor used to generate *B. thetaiotaomicron* Tn library; AmpR	Ortañez & Degnan 2024
Bacteroides thetaiotaomicron VPI-5482 *Δtdk*	*B. thetaiotaomicron* str. VPI-5482 *Δtdk*	"wild-type" deletion background; Used to generate *B. thetaiotaomicron* Tn mutant library	Eggerth and Gagnon, 1933
Bacteroides thetaiotaomicron VPI-5482 *Δtdk* pSAM_BC2_ermG *(*Tn mutant library*)*		Recipient strain that that acquired PmDOT, ErmR	Ortañez & Degnan 2024
Bacteroides thetaiotaomicron VPI-5482 *Δtdk* pSAM_BC2_ermG PmDOT		*B. thetaiotaomicron* transconjugant with PmDOT insertion in *BT1582,* ErmR + TetR	This study
Bacteroides thetaiotaomicron VPI-5482 *pNBU2_tetQ*		Recipient strain to screen Tn libraries for MGEs, TetR	Ortañez & Degnan 2024
Parabacteroides merdae ATCC 43184	*P. merdae*	Donor strain that mobilized PmDOT into *B. thetaiotaomicron* VPI-5482 Tn mutant, Naturally TetR	Johnson et al. 1986
Phocaeicola dorei DSM 17855	*Ph. dorei*	Donor strain used to screen CTnDOT-like mobiliziation, Naturally TetR	Bakir et al. 200
Phocaeicola coprocola DSM 17136	*Ph. coprocola*	Donor strain used to screen CTnDOT-like mobiliziation, Naturally TetR	Kitahara et al. 2005
Phocaeicola stercoris ATCC 43183	*Ph. stercoris*	Donor strain used to screen CTnDOT-like mobiliziation, Naturally TetR	Johnson et al. 1986
Phocaeicola intestinalis DSM 17393	*Ph. intestinalis*	Donor strain used to screen CTnDOT-like mobiliziation, Naturally TetR	Bakir et al. 2006
Phocaeicola plebius DSM 17135	*Ph. plebius*	Donor strain used to screen CTnDOT-like mobiliziation, Naturally TetR	Kitahara et al. 2005
Bacteroides eggerthii DSM 20697	*B. eggerthii*	Donor strain used to screen CTnDOT-like mobiliziation, Naturally TetR	Holdeman and Moore 1974
Parabacteroides johnsonii DSM 18315	*P. johnsonii*	Donor strain used to screen CTnDOT-like mobiliziation, Naturally TetR	Sakamoto et al. 2007
Bacteroides thetaiotaomicron 3731	*B. thetaiotaomicron 3731*	Donor strain used to screen CTnDOT-like mobiliziation, Naturally TetR	
Bacteroides cellulosilyticus WH2		Donor strain used to screen CTnDOT-like mobiliziation, Naturally TetR	
Bacteroides thetaiotaomicron VPI-5482 *Δtdk* pNBU2_ermG		Recipient strain to test mobilzation of CTnDOT-like elements, ErmR	Frye et al. 2022
Tn, transposon; AmpR ErmR, erythromycin resistant			

Table 3.3 All gut microbe genomes screened for CTnDOT.

Organism	RefSeq Accession	CTnDOT Detection	Metagenome *rpsE* Gene Mapping
Alistipes finegoldii DSM 17242	NC_018011		X
Alistipes indistinctus YIT 12060	NZ_ADLD00000000		X
Alistipes putredinis DSM 17216	NZ_ABFK00000000		X
Alistipes shahii WAL 8301	NC_021030		X
Alistipes sp. HGB5	NZ_AENZ00000000	INTACT	
Alistipes sp. JC136	NZ_CAEG00000000		X
Bacteroides caccae ATCC 43185	NZ_AAVM00000000		X
Bacteroides caccae CL03T12C61	NZ_AGXF00000000	INTACT	
Bacteroides cellulosilyticus CL02T12C19	NZ_AGXG00000000		
Bacteroides cellulosilyticus DSM 14838	NZ_ACCH00000000		
Bacteroides cellulosilyticus WH2	NZ_ATFI00000000		X
Bacteroides clarus YIT 12056	NZ_AFBM00000000		X
Phocaeicola coprocola DSM 17136	NZ_ABIY00000000	CORE	X
Phocaeicola coprophilus DSM 18228	NZ_ACBW00000000		X
Phocaeicola dorei 5_1_36/D4	NZ_ACDI00000000	INTACT	
Phocaeicola dorei 9_1_42FAA	NZ_ACAA00000000		
Phocaeicola dorei CL02T00C15	NZ_AGXH00000000	INTACT	X
Phocaeicola dorei CL02T12C06	NZ_AGXJ00000000	INTACT	
Phocaeicola dorei CL03T12C01	NZ_AGXI00000000	INTACT	
Phocaeicola dorei DSM 17855	NZ_ABWZ00000000	CORE	
Phocaeicola dorei VPI-6598B	SAMN18350537		
Phocaeicola dorei VPI-2277^{o}	SAMN18350535	INTACT	
Phocaeicola dorei VPI-3776A^{o}	SAMN18350536	INTACT	
Bacteroides eggerthii 1_2_48FAA	NZ_ACWG00000000	INTACT	X
Bacteroides eggerthii DSM 20697	NZ_ABVO00000000	CORE	
Bacteroides faecis MAJ27	NZ_AGDG00000000	INTACT	X
Bacteroides finegoldii CL09T03C10	NZ_AGXW00000000		X
Bacteroides finegoldii DSM 17565	NZ_ABXI00000000		
Bacteroides fluxus YIT 12057	NZ_AFBN00000000		X
Bacteroides fragilis 2_1_16	NZ_ACPP00000000	INTACT	

Bacteroides fragilis 2_1_56FAA	NZ_ACWI00000000	INTACT	
Bacteroides fragilis 3_1_12	NZ_ABZX00000000	INTACT	
Bacteroides fragilis 3_2_5	NZ_ACIB00000000		
Bacteroides fragilis 638R	NC_016776		
Bacteroides fragilis CL03T00C08	NZ_AGXK00000000	INTACT	
Bacteroides fragilis CL03T12C07	NZ_AGXL00000000	INTACT	
Bacteroides fragilis CL05T00C42	NZ_AGXO00000000	INTACT	
Bacteroides fragilis CL05T12C13	NZ_AGXP00000000	INTACT	
Bacteroides fragilis CL07T00C01	NZ_AGXM00000000	INTACT,INTACT	
Bacteroides fragilis CL07T12C05	NZ_AGXN00000000	INTACT,INTACT	
Bacteroides fragilis HMW 610	NZ_AGXQ00000000	INTACT	
Bacteroides fragilis HMW 615	NZ_AGXR00000000	INTACT	
Bacteroides fragilis HMW 616	NZ_ALOB00000000		X
Bacteroides fragilis NCTC 9343	NC_003228		X
Bacteroides fragilis YCH46	NC_006347	INTACT	
Bacteroides helcogenes P 36-108	NC_014933		X
Bacteroides intestinalis DSM 17393	NZ_ABJL00000000	CORE	X
Bacteroides nordii CL02T12C05	NZ_AGXS00000000		X
Bacteroides nordii WAL7936*	SAMN18350534	INTACT,INTACT	
Bacteroides oleiciplenus YIT 12058	NZ_ADLF00000000	INTACT	X
Bacteroides ovatus 3_1_23	NZ_ACRS00000000	INTACT	
Bacteroides ovatus 3_8_47FAA	NZ_ACWH00000000	INTACT	
Bacteroides ovatus ATCC 8483	NZ_AAXF00000000		
Bacteroides ovatus CL02T12C04	NZ_AGXT00000000		X
Bacteroides ovatus CL03T12C18	NZ_AGXU00000000	INTACT	
Bacteroides ovatus SD CC 2a	NZ_ADMP00000000		
Bacteroides ovatus SD CMC 3f	NZ_ADMO00000000	INTACT	
Phocaeicola plebeius DSM 17135	NZ_ABQC00000000	CORE	X
Phocaeicola salanitronis DSM 18170	NC_015164	INTACT	X
Bacteroides salyersiae CL02T12C01	NZ_AGXV00000000	INTACT	X
Bacteroides sp. 2_1_33B	NZ_ACPR00000000		
Bacteroides sp. 2_1_7	NZ_ABZY00000000	INTACT	
Bacteroides sp. 2_2_4	NZ_ABZZ00000000	INTACT	
Bacteroides sp. 20_3	NZ_ACRQ00000000		
*Bacteroides s*p. 3_1_19	NZ_ADCJ00000000	INTACT	

Bacteroides sp. 3_1_33FAA	NZ_ACPS00000000	INTACT,INTACT	
Bacteroides sp. CF50	NC_022526		X
Bacteroides sp. D2	NZ_ACGA00000000		
Bacteroides sp. D20	NZ_ACPT00000000	INTACT	
Bacteroides sp. D22	NZ_ADCK00000000	INTACT	
Bacteroides stercoris ATCC 43183	NZ_ABFZ00000000	CORE	x
Bacteroides thetaiotaomicron 1_1_14	NZ_ACRP00000000		
Bacteroides thetaiotaomicron 1_1_6	NZ_ACIC00000000	INTACT	
Bacteroides thetaiotaomicron 3731	NC_Bthetaiotaomicron3731	CORE	
Bacteroides thetaiotaomicron 7330	NC_Bthetaiotaomicron7330		
Bacteroides thetaiotaomicron VPI-BT-DOT2	BTH03	**CTnDOT**	
Bacteroides thetaiotaomicron VPI-0633-1	SAMN18350524		
Bacteroides thetaiotaomicron VPI-0940-1	SAMN18350525		
Bacteroides thetaiotaomicron VPI-5482	NC_004703		X
Bacteroides thetaiotaomicron VPI-C11-15	SAMN18350527	INTACT	
Bacteroides thetaiotaomicron VPI-J19-343	SAMN18350526		
Bacteroides uniformis 4_1_36	NZ_ACTC00000000		
Bacteroides uniformis ATCC 8492	NZ_AAYH00000000	INTACT	X
Bacteroides uniformis CL03T00C23	NZ_AGXX00000000		
Bacteroides uniformis CL03T12C37	NZ_AGXY00000000		
Bacteroides uniformis VPI-R5-33	SAMN18350532		
Bacteroides uniformis VPI-61	SAMN18350528	INTACT	
Bacteroides uniformis VPI-909	SAMN18350529		
Bacteroides uniformis VPI-C20-25	SAMN18350530		
Bacteroides uniformis VPI-C7-17	SAMN18350531		
Phocaeicola vulgatus 3_1_40A	NZ_ACRT00000000	INTACT	
Phocaeicola vulgatus 4_3_47FAA	NZ_ACDR00000000		
Phocaeicola vulgatus ATCC 8482	NC_009614		
Phocaeicola vulgatus CL09T03C04	NZ_AGXZ00000000	INTACT	X

Phocaeicola vulgatus PC510	NZ_ADKO00000000		
Phocaeicola *vulgatus* VPI-2365	SAMN14525940		
Phocaeicola *vulgatus* VPI-4025	SAMN14525944	INTACT	
Phocaeicola *vulgatus* VPI-4496.2	SAMN14525945	INTACT	
Phocaeicola *vulgatus* VPI-5710	SAMN14525942	INTACT	
Phocaeicola vulgatus VPI-4245	SAMN14525939		
Phocaeicola vulgatus VPI-4506	SAMN14525946		
Phocaeicola vulgatus VPI-6186	SAMN14525941		
Bacteroides xylanisolvens 1_1_30	NZ_ADCL00000000	INTACT,INTACT	
Bacteroides xylanisolvens 2_1_22	NZ_ACPQ00000000	INTACT	
Bacteroides xylanisolvens CL03T12C04	NZ_AGXE00000000	INTACT	
Bacteroides xylanisolvens D1	NZ_ACAB00000000	INTACT	
Bacteroides xylanisolvens SD CC 1b	NZ_ADKP00000000		X
Bacteroides xylanisolvens XB1A	NC_BxylanisolvensXB1A	INTACT	
Barnesiella intestinihominis YIT 11860	NZ_ADLE00000000	INTACT	
Dysgonomonas gadei ATCC BAA-286	NZ_ADLV00000000		X
*Dysgonomonas mossii D*SM 22836	NZ_ADLW00000000		X
Odoribacter laneus YIT 12061	NZ_ADMC00000000		X
Odoribacter splanchnicus DSM 20712	NC_015160		X
Parabacteroides distasonis ATCC 8503	NC_009615		X
Parabacteroides distasonis CL03T12C09	NZ_AGZM00000000	INTACT	
Parabacteroides distasonis CL09T03C24	NZ_AGZN00000000	INTACT	
Parabacteroides goldsteinii CL02T12C30	NZ_AGZO00000000	INTACT	X
Parabacteroides johnsonii CL02T12C29	NZ_AGZP00000000		
Parabacteroides johnsonii DSM 18315	NZ_ABYH00000000	CORE	X
Parabacteroides merdae ATCC 43184	NZ_AAXE00000000	CORE	X
Parabacteroides merdae CL03T12C32	NZ_AGZQ00000000	INTACT	
Parabacteroides merdae CL09T00C40	NZ_AJPU00000000	INTACT	
Parabacteroides sp. D13	NZ_ACPW00000000	INTACT	
Parabacteroides sp. D25	NZ_ACUA00000000	INTACT	

Paraprevotella clara YIT 11840	NZ_AFFY00000000	INTACT	X
Paraprevotella xylaniphila YIT 11841	NZ_AFBR00000000		X
Porphyromonas gingivalis ATCC 33277	NC_010729		X
Porphyromonas gingivalis TDC60	NC_015571		
Porphyromonas gingivalis W83	NC_002950		
Prevotella copri DSM 18205	NZ_ACBX00000000		X
Prevotella salivae DSM 15606	NZ_AEQO00000000		X
Prevotella stercorea DSM 18206	NZ_AFZZ00000000	INTACT	X
Tannerella forsythia ATCC 43037	NC_016610	INTACT	X
Tannerella sp. 6_1_58FAA_CT1	NZ_ACWX00000000		X
* Originally deposited in the VPI strain collection as *B. uniformis*			
º Originally deposited in VPI strain collection as *B. vulgatus*			

Table 3.4 Calculated RPKM of patient metagenome samples corresponding to CTnDOT-like elements and Bacteroidota *rpsE* alleles

Code	Authors	BioProject	Collection Year	Location	Population	Note	n	Mean No. of Raw Reads Pairs per sample (Millions ± SE)	CTnDOT RPKM1	*rpsE* RPKM	CTnDOT RPKM/ *rpsE* RPKM
USA1	Anthony et al.	PRJNA664754	2017	Missouri, USA	Industrial	Pre-Abx	20*	19.9 ± 0.9	8.05 ± 0.90	1.76 ± 0.11	4.66 ± 0.46
						Post-Abx	114	5.4 ± 0.2	8.63 ± 0.39	1.89 ± 0.06	4.79 ± 0.21
DNK	Kang et al.	PRJNA588313	2015	Denmark	Industrial	Pre-Abx	10	46.3 ± 2.6	3.06 ± 1.25	1.02 ± 0.27	2.53 ± 0.43
						Post-Abx	20	45.4 ± 1.8	4.03 ± 0.90	1.15 ± 0.26	3.40 ± 0.76
CAN	Raymond et al.	PRJEB8094	2012	Quebec, Canada	Industrial	Pre-Abx/Control	24	58.1 ± 4.1	5.62 ± 0.78	1.06 ± 0.12	5.28 ± 0.42
						Control	12	78.4 ± 7.5	5.01 ± 0.98	1.1 ± 0.15	4.40 ±0.30
						Post-Abx	36	68.2 ± 4.4	5.51 ± 0.46	1.09 ± 0.09	5.49 ± 0.35
CHN1	Cheung et al.	PRJNA834885	n.a.	Hong Kong, China	Industrial	Pre-trip	67	21.5 ± 0.1	2.03 ± 0.23	0.44 ± 0.37	4.71 ± 0.24
						Post-trip	52	21.5 ± 0.2	2.46 ± 0.30	0.52 ± 0.06	4.97 ± 0.22
						Post-trip/Abx	15	21.6 ± 0.3	2.09 ± 0.42	0.38 ±0.09	5.79 ± 0.58
AUT	Feng et al.	PRJEB7774	2013	Salzburg, Austria	Industrial	Control	61	25.0 ± 0.5	1.42 ± 0.30	0.25 ± 0.06	8.40 ± 3.04
						CRC	46	27.6 ± 0.6	2.66 ± 0.45	0.65 ± 0.10	4.54 ± 0.29
USA2	Vogtmann et al.	PRJEB12449	1985/1987	Washington DC, USA	Industrial	Control	49	7.7 ± 0.2	6.21 ± 0.54	1.38 ± 0.10	5.70 ± 0.79
						CRC	52	7.7 ± 0.3	5.93 ±0.44	1.53 ± 0.07	4.07 ± 0.33
CHN2	Yu et al.	PRJEB10878	2012	Hong Kong, China	Industrial	Control	54	29.5 ± 0.5	4.70 ± 0.50	1.13 ± 0.11	4.43 ± 0.25
						CRC	73	27.1 ± 0.7	5.06 ± 0.43	1.19 ± 0.09	4.46 ± 0.20
PER	Obregon-Tito et al.	PRJNA268964	2015	Peru	Hunter-Gatherer	Control	20	30.3 ± 1.0	1.23 ± 0.17	0.31 ± 0.06	6.97 ± 1.69
MDG	Pasolli et al.	PRJNA485056	2019	Madagascar	Hunter-Gatherer	Control	68	34.1 ± 2.3	1.42 ± 0.14	0.66 ± 0.07	3.98 ± 0.60
TZA	Carter et al.	PRJEB49206	2023	Tanzania	Hunter-Gatherer	Control	135	105.9 ± 6.0	4.52 ± 0.23	0.79 ± 0.05	11.67 ± 1.29

* Reads from 3-4 pre antibiotic exposure samples were pooled for each of the 20 patients in this study.

1 Average Reads Per Kilobase per Million mapped reads values ± standard error

Table 3.5a Novel detection of CTnDOT-like elements.

CTn ID	Identification	Subfamily	Strain	Starting ORF	Ending ORF	Length	*tetQ*	*aadS*	*tetX2*	*tetX1*
CTnDOT	CORE	2	*Bacteroides thetaiotaomicron* VPI-BT-DOT2	BTH033970	BTH033995	64.0	X	X	X	X
				BTH032365	BTH032405					
Beg_087	CORE	1	*Bacteroides eggerthii* DSM 20697	BACEGG03049	BACEGG03116	73.3	X			
Bin_081	CORE	1	*Bacteroides intestinalis* DSM 17393	BACINT00772	BACINT00835	66.2	X			
Bst_077	CORE	1	*Bacteroides stercoris* ATCC 43183	BACSTE01062	BACSTE01116	48.5	X			
Bth_120	CORE	2	*Bacteroides thetaiotaomicron* 3731	Bthe_732569	Bthe_732600	39.1	X			
				Bthe_735212	Bthe_735219					
Pjo_094	CORE	21	*Parabacteroides johnsonii* DSM 18315	PRABACTJOHN01768	PRABACTJOHN01806	30.6	X			
Pme_069	CORE	1	*Parabacteroides merdae* ATCC 43184	PARMER00775	PARMER00826	48.5	X			
Bco_064	CORE	1	*Phocaeicola coprocola* DSM 17136	BACCOP01817	BACCOP01860	45.9	X			
Bdo_044	CORE	20	*Phocaeicola dorei* DSM 17855	BACDOR01615	BACDOR01670	46.6	X			
Bpl_086	CORE	1	*Phocaeicola plebeius* DSM 17135	BACPLE03530	BACPLE03583	53.3	X			
Bca_582	INTACT	18	*Bacteroides caccae* CL03T12C61	HMPREF1061011189	HMPREF1061011241	52.7	X			
Beg_165	INTACT	16	*Bacteroides eggerthii* 1_2_48FAA	HMPREF1016011625	HMPREF1016011683	62.0	X			
Bfa_587	INTACT	4	*Bacteroides faecis* MAJ27	BfaeM0101000016859	BfaeM0101000017134	53.8	X			
Bfr_100	INTACT	1	*Bacteroides fragilis* 3_1_12	Bfra3010100023342	Bfra3010100023602	49.7	X			
Bfr_589	INTACT	2	*Bacteroides fragilis* CL03T00C08	HMPREF1066036632	HMPREF1066036695	59.7	X	X	X	X

Bfr_590	INTACT	2	*Bacteroides fragilis* CL03T12C07	HMPREF1067025 05	HMPREF1067025 66	58.6	X	X	X	X
Bfr_591B	INTACT	8	*Bacteroides fragilis* CL05T00C42	HMPREF1079023 73	HMPREF1079024 28	53.5	X			
Bfr_592	INTACT	8	*Bacteroides fragilis* CL05T12C13	HMPREF1080029 91	HMPREF1080030 06	59.8	X			
				HMPREF1080038 73	HMPREF1080039 18					
Bfr_593A	INTACT	10	*Bacteroides fragilis* CL07T00C01	HMPREF1055026 90	HMPREF1055027 45	40.3	X			
Bfr_593B	INTACT	4	*Bacteroides fragilis* CL07T00C01	HMPREF1055005 08	HMPREF1055005 81	52.4	X	X		
Bfr_594A	INTACT	10	*Bacteroides fragilis* CL07T12C05	HMPREF1056001 80	HMPREF1056002 26	40.9	X			
Bfr_594B	INTACT	4	*Bacteroides fragilis* CL07T12C05	HMPREF1056023 39	HMPREF1056024 12	45.3	X			
Bfr_595	INTACT	1	*Bacteroides fragilis* HMW 610	HMPREF1203041 75	HMPREF1203042 22	48.6	X			
Bfr_596	INTACT	15	*Bacteroides fragilis* HMW 615	HMPREF1204033 58	HMPREF1204034 08	48.1	X			
Bfr_008	INTACT	1	*Bacteroides fragilis* YCH46	BFych460103	BFych460152	48.7	X			
Bno_0AKA	INTACT	3	*Bacteroides nordii* WAL7936	BUE090309	BUE090336	55.2	X			
				BUE090800	BUE090826					
Bno_0AKB	INTACT	24	*Bacteroides nordii* WAL7936	BUE091000	BUE091046	42.7				
Bol_600	INTACT	22	*Bacteroides oleiciplenus* YIT 12058	HMPREF9447054 25	HMPREF9447054 69	50.0	X			
Bov_167	INTACT	1	*Bacteroides ovatus* 3_8_47FAA	HMPREF1017019 69	HMPREF1017020 16	47.8	X			
Bov_602	INTACT	1	*Bacteroides ovatus* CL03T12C18	HMPREF1070003 82	HMPREF1070004 31	52.8	X			

Bov_169	INTACT	4	*Bacteroides ovatus* SD CMC 3f	CUY2152	CUY2198	74.2	X			
				CUY2769	CUY2801					
Bsa_607	INTACT	1	*Bacteroides salyersiae* CL02T12C01	HMPREF1071018898	HMPREF1071019954	60.0	X			
Bxy_145A	INTACT	1	*Bacteroides* sp. 1_1_30	HMPREF0127014423	HMPREF0127014480	67.2	X			
Bxy_145B	INTACT	9	*Bacteroides* sp. 1_1_30	HMPREF0127048334	HMPREF0127048366	40.1	X			
Bth_110	INTACT	3	*Bacteroides* sp. 1_1_6	BSIG02053	BSIG02068	57.6	X			
				BSIG05184	BSIG05228					
Bfr_146	INTACT	1	*Bacteroides* sp. 2_1_16	HMPREF0101034494	HMPREF0101035550	53.8	X			
Bxy_147	INTACT	6	*Bacteroides* sp. 2_1_22	HMPREF0102016677	HMPREF0102017728	74.0	X			
Bfr_149	INTACT	13	*Bacteroides* sp. 2_1_56FAA	HMPREF1018027769	HMPREF1018028827	55.4	X			
Pdi_150	INTACT	1	*Bacteroides* sp. 2_1_7	B2010100020951	B2010100021266	59.4	X			
Bov_151	INTACT	1	*Bacteroides* sp. 2_2_4	BSCG05214	BSCG05223	56.4	X			
				BSCG05498	BSCG05543					
Pdi_153	INTACT	1	*Bacteroides* sp. 3_1_19	HMPREF0104032208	HMPREF0104032256	50.2	X			
Bov_154	INTACT	2	*Bacteroides* sp. 3_1_23	HMPREF9010029986	HMPREF9010030053	48.0	X	X	X	X
Bdo_155A	INTACT	3	*Bacteroides* sp. 3_1_33FAA	HMPREF0105115555	HMPREF0105120808	46.8	X			
Bdo_155B	INTACT	9	*Bacteroides* sp. 3_1_33FAA	HMPREF0105365369	HMPREF0105372424	54.7	X			
Bvu_156	INTACT	1	*Bacteroides* sp. 3_1_40A	HMPREF9011044410	HMPREF9011044466	54.9	X			
Bxy_101	INTACT	6	*Bacteroides* sp. D1	BSAG04082	BSAG04136	73.5	X			
Bun_162	INTACT	17	*Bacteroides* sp. D20	HMPREF0	HMPREF0	55.4	X			

				96901759	96901818					
Bxy_163	INTACT	1	*Bacteroides* sp. D22	HMPREF0106040010	HMPREF0106040062	52.1	X			
Bth_0AV	INTACT	2	*Bacteroides thetaiotaomicron* VPI-C11-15	BTH052720	BTH052766	45.7	X			
Bun_011	INTACT	5	*Bacteroides uniformis* ATCC 8492	BACUNI01743	BACUNI01798	56.2				
				BACUNI03375	BACUNI03393					
Bun_0AE	INTACT	5	*Bacteroides uniformis* VPI-61	BUC071095	BUC071127	55.5				
				BUC073515	BUC073542					
Bxy_612	INTACT	1	*Bacteroides xylanisolvens* CL03T12C04	HMPREF1074030565	HMPREF1074030617	52.4	X			
Bxy_122	INTACT	3	*Bacteroides xylanisolvens* XB1A	Bxyl3097	Bxyl3157	72.2	X			
Bdo_164	INTACT	2	*Phocaeicola dorei* 5_1_36/D4	BSEG01291	BSEG01239	55.0	X			
Bdo_584	INTACT	1	*Phocaeicola dorei* CL02T00C15	HMPREF1063047870	HMPREF1063048420	58.9	X			
Bdo_585	INTACT	1	*Phocaeicola dorei* CL02T12C06	HMPREF1064045120	HMPREF1064045680	59.5	X			
Bdo_586	INTACT	3	*Phocaeicola dorei* CL03T12C01	HMPREF1065038450	HMPREF1065038900	44.7	X			
Bdo_0AL	INTACT	1	*Phocaeicola dorei VPI-2277*	BVE112618	BVE112664	47.8	X			
Bdo_0AM	INTACT	1	*Phocaeicola dorei VPI-3776A*	BVG033805	BVG033851	47.8	X			
Bsa_603	INTACT	12	*Phocaeicola salanitronis* DSM 18170	Bacsa2525	Bacsa2577	49.2	X			
Bvu_611	INTACT	1	*Phocaeicola vulgatus* CL09T03C04	HMPREF1058019530	HMPREF1058020000	49.6	X			
Bvu_0AZ	INTACT	7	*Phocaeicola vulgatus* VPI-4025	BVL012440	BVL012486	101.6	X			
				BVL012904	BVL012965					
Bvu_0AO	INTACT	2	*Phocaeicola vulgatus* VPI-4496.2	BVG070839	BVG070860	64.2	X	X	X	X
				BVG073019	BVG073063					

Bvu_0BBA	INTACT	25	*Phocaeicola vulgatus* VPI-5710	BVL020913	BVL020964	61.7				
				BVL022846	BVL022858					
Bvu_0BBB	INTACT	1	*Phocaeicola vulgatus* VPI-5710	BVL022506	BVL022552	47.8	X			
Bin_624	INTACT	11	*Barnesiella intestinihominis* YIT 11860	HMPREF944801844	HMPREF944801892	53.5	X			
Pdi_613	INTACT	23	*Parabacteroides distasonis* CL03T12C09	HMPREF107504019	HMPREF107504062	43.9	X			
Pdi_614	INTACT	27	*Parabacteroides distasonis* CL09T03C24	HMPREF105903585	HMPREF105903647	55.0				
Pgo_615	INTACT	7	*Parabacteroides goldsteinii* CL02T12C30	HMPREF107602145	HMPREF107602210	56.7				
Pme_617	INTACT	4	*Parabacteroides merdae* CL03T12C32	HMPREF106003703	HMPREF106003757	56.6	X			
Pme_618	INTACT	1	*Parabacteroides merdae* CL09T00C40	HMPREF107801847	HMPREF107801893	47.8	X			
Pdi_422	INTACT	14	*Parabacteroides* sp. D13	HMPREF061902717	HMPREF061902764	51.5	X			
Pdi_619	INTACT	1	*Parabacteroides* sp. D25	HMPREF099904211	HMPREF099904284	70.4	X			
Pcl_564	INTACT	5	*Paraprevotella clara* YIT 11840	HMPREF944100017	HMPREF944100081	44.1				
Pst_565	INTACT	1	*Prevotella stercorea* DSM 18206	HMPREF067300061	HMPREF067300112	48.7	X			
Tfo_620	INTACT	3	*Tannerella forsythia* ATCC 43037	BFO1222	BFO1282	49.2	X			
Ali_132	INTACT	1	*Alistipes* sp. HGB5	HMPREF972028551	HMPREF972028598	50.5	X			
–	FRAGMENTARY	–	*Bacteroides* sp. 20_3	HMPREF900803817	HMPREF900803501	–	X			
				HMPREF9	HMPREF9	–				

				00804800	00804805					
				HMPREF900804715	HMPREF900804747	–				
				HMPREF900804771	HMPREF900804792	–				
–	FRAGMENTARY	–	*Bacteroides oleiciplenus* YIT 12058	HMPREF944700001	HMPREF944700016	–				
				HMPREF944702764	HMPREF944702779	–	X			
				HMPREF944705001	HMPREF944705009	–				
				HMPREF944705489	HMPREF944705495	–				
				HMPREF944702764	HMPREF944702779	–				
–	FRAGMENTARY	–	*Bacteroides* sp. 9_1_42FAA	BSBG04789	BSBG04790	–	X			
				BSBG04794	BSBG04795	–	X			
				BSBG00680	BSBG00690	–				
				BSBG02890	BSBG02929	–				
				BSBG04655	BSBG04664	–				
–	FRAGMENTARY	–	*Bacteroides ovatus* SD CC 2a	CW14779	CW14780	–				
				CW14712	CW14724	–				
				CW14799	CW14833	–	X			
–	PARTIAL	–	*Bacteroides* sp. 1_1_14	HMPREF900700044	HMPREF900700152	–				
				HMPREF900700152	HMPREF900700234	–				
–	FRAGMENTARY	–	*Bacteroides nordii* CL02T12C05	HMPREF106800001	HMPREF106800082	–				
				HMPREF106800083	HMPREF106800093	–	X			

				HMP REF1 06804 292	HMP REF1 06804 325	–				
				HMP REF1 06804 292	HMP REF1 06804 346	–				
–	PARTI AL	–	*Bacteroides xylanisolvens* SD CC 1b	CW33 738	CW33 757	–	X			
–	PARTI AL	–	*Bacteroides cellulosilyticus* WH2	BWH 2RS01 04525	BWH 2RS01 04590	–	X			
–	PARTI AL	–	*Bacteroides fluxus* YIT 12057	HMP REF9 44600 072	HMP REF9 44600 089	–	X			
–	PARTI AL	–	*Phocaeicola vulgatus* PC510	CUU2 481	CUU2 501	–	X			
–	PARTI AL	–	*Alistipes shahii* WAL 8301	AL13 1210	AL13 1440	–	X			
–	PARTI AL	–	*Alistipes finegoldii* DSM 17242	Alfi05 44	Alfi05 66	–	X			

Table 3.5b Novel detection of CTnDOT-like elements

CTn ID	Identification	Subfamily	length	TetQ	aadS	ermF
CTnDOT	CORE	2		BTH033993	BTH033980	BTH033983
Beg_087	CORE	1	73272	BACEGG03060		
Bin_081	CORE	1	66214	BACINT00817		
Bst_077	CORE	1	48529	BACSTE01105		
Bth_120	CORE	2	35273	Bthe3732590		
			4013			
Pjo_094	CORE	21	30600	PRABACTJOHN01781		
Pme_069	CORE	1	48464	PARMER00787		
Bco_064	CORE	1	45890	BACCOP01828		~~BACCOP03001~~
Bdo_044	CORE	20	46596	BACDOR01628		
Bpl_086	CORE	1	53312	BACPLE03542		
Bca_582	INTACT	18	52661	HMPREF106101202		
Beg_165	INTACT	16	61968	HMPREF101601635		~~HMPREF101603622~~
Bfa_587	INTACT	4	53811	BfaeM010100016919		~~BfaeM010100023725~~
Bfr_100	INTACT	1	49745	Bfra3010100023552		~~Bfra3010100023687~~
Bfr_589	INTACT	2	59698	HMPREF106603670	HMPREF106603685	HMPREF106603682
Bfr_590	INTACT	2	58564	HMPREF106702543	HMPREF106702556	HMPREF106702553
Bfr_591B	INTACT	8	53454	HMPREF107902409		
Bfr_592	INTACT	8	16778	HMPREF108003909		
			43028	HMPREF108003909		
Bfr_593A	INTACT	10	52427	HMPREF105502727		
Bfr_593B	INTACT	4	73419	HMPREF105500556		HMPREF105500552
Bfr_594A	INTACT	10	45273	HMPREF105600189		
Bfr_594B	INTACT	4	74021	HMPREF105602364		
Bfr_595	INTACT	1	48591	HMPREF120304186		~~HMPREF120304084~~
Bfr_596	INTACT	15	48127	HMPREF120403369		~~HMPREF120402867~~
Bfr_008	INTACT	1	48722	BFych460141		
Bno_0AKA	INTACT	3	32669	BUE090321		
			13976			
Bno_0AKB	INTACT	24	29303	NO		
Bol_600	INTACT	22	50020	HMPREF944705460		
Bov_167	INTACT	1	47832	HMPREF101701979		

Bov_602	INTACT	1	52801	HMPREF107000394		
Bov_169	INTACT	4	52509	CUY2175		
			21671	CUY2175		
Bsa_607	INTACT	1	60012	HMPREF107101942		
Bxy_145A	INTACT	1	67198	HMPREF012701433		~~HMPREF012704848~~
Bxy_145B	INTACT	9	23712	HMPREF012704849		HMPREF012704848
Bth_110	INTACT	3	9218	BSIG05199		
			48430	BSIG05199		
Bfr_146	INTACT	1	53759	HMPREF010103535		
Bxy_147	INTACT	6	73970	HMPREF010201688		~~HMPREF010204411~~
Bfr_149	INTACT	13	55383	HMPREF101802805		HMPREF101802808
Pdi_150	INTACT	1	59366	B2010100021001		
Bov_151	INTACT	1	10756	BSCG05500		~~BSCG02203~~
Pdi_153	INTACT	1	50159	HMPREF010403218		
Bov_154	INTACT	2	48041	HMPREF901003026	HMPREF901003043	HMPREF901003040
Bdo_155A	INTACT	3	46774	HMPREF01051197		~~HMPREF01053680~~
Bdo_155B	INTACT	9	54677	HMPREF01053681		HMPREF01053680
Bvu_156	INTACT	1	54898	HMPREF901104456		~~HMPREF901102147~~
Bxy_101	INTACT	6	73529	BSAG04093		~~BSAG04317~~
Bun_162	INTACT	17	55376	HMPREF096901769		
Bxy_163	INTACT	1	52134	HMPREF010604021		
Bth_0AV	INTACT	2	40439	BTH052729		
Bun_011	INTACT	5	36588	–		
			19642	–		
Bun_0AE	INTACT	5	21944	NO		
			13599			
Bxy_612	INTACT	1	52426	HMPREF107403607		
Bxy_122	INTACT	3	72160	Bxyl3137		
Bdo_164	INTACT	2	55022	BSEG01279		
Bdo_584	INTACT	1	58875	HMPREF106304799		
Bdo_585	INTACT	1	59462	HMPREF106404524		
Bdo_586	INTACT	3	44684	HMPREF106503880		
Bdo_0AL	INTACT	1	49134	BVE112654		
Bdo_0AM	INTACT	1	49134	BVG033815		
Bsa_603	INTACT	12	49239	Bacsa2536		Bacsa2566
Bvu_611	INTACT	1	49595	HMPREF105801990		

Bvu_0AZ	INTACT	7	43981	BVL012450		
			55078			
Bvu_0AO	INTACT	2	26128	BVG073061	BVG070850	BVG070847
			37770			
Bvu_0BBA	INTACT	25	34679			
			14775			
Bvu_0BBB	INTACT	1	46218	BVL022542		
Bin_624	INTACT	11	53528	HMPREF944801882		HMPREF944801853
Pdi_613	INTACT	23	43940	HMPREF107504052		
Pdi_614	INTACT	27	**INTACT**			
Pgo_615	INTACT	7	56712	#N/A		
Pme_617	INTACT	4	56629	HMPREF106003713		
Pme_618	INTACT	1	47832	HMPREF107801857		
Pdi_422	INTACT	14	51536	HMPREF061902754		
Pdi_619	INTACT	1	70430	HMPREF099904245		
Pcl_564	INTACT	5	**INTACT**			
Pst_565	INTACT	1	48723	HMPREF067300072		
Tfo_620	INTACT	3	49194	BFO1235		
Ali_132	INTACT	1	50459	HMPREF97202861		
–	FRAGMENTARY	–		HMPREF900803817		
			11414	HMPREF900804804		
			27630	HMPREF900804804	HMPREF900804704	HMPREF900804707
			20342	HMPREF900804804		
–	FRAGMENTARY	–	21769	HMPREF944700007		HMPREF944705500
			21769	HMPREF944700007		HMPREF944705500
			3828	HMPREF944700007		HMPREF944705500
			3314	HMPREF944700007		HMPREF944705500
				HMPREF944702773		
–	FRAGMENTARY	–		BSBG04789		
				BSBG04794		
			11357	BSBG04789/BSBG04794		
			32674	BSBG04789/BSBG04794		
			11267	BSBG04789/BSBG04794		
–	FRAGMENTARY	–	4767	CW14800		
			5880	CW14800		
				CW14800		

–	PARTIAL	–	107152	HMPREF900700055		
			73861	HMPREF900700055		
–	FRAGMENTARY	–	77100	HMPREF106800085		
				HMPREF106800085		
			46331	HMPREF106800085		
–	PARTIAL	–	26942	CUU2491		~~CUU3501~~
–	PARTIAL	–	26903	CW33741		
–	PARTIAL	–	23768	HMPREF944600083		
–	PARTIAL	–		AL131330		
–	PARTIAL	–		Alfi0555		
–	PARTIAL	–		BWH2RS0104540		

Table 3.6 *In vitro* conjugation results for core CTnDOT-like elements.

Donor (TetR)	Transconjugants TetR + ErmR
Ph. dorei DSM 17855	-
Ph. coprocola DSM 17136	-
B. stercoris ATCC 43183	-
B. intestinalis DSM 17393	+
Ph. plebius DSM 17135	-
B. eggerthii DSM 20697	-
P. johnsonii DSM 18315	+
B. cellulosilyticus WH2	+
B. thetaiotaomicron 3731	-

Table 3.7 Functional summary of variable CTnDOT-like genes

Name	Synonym	PFAM	PFAM description
CTnDOT	BTH033970	PF00589, PF13102	Phage integrase family, Phage integrase SAM-like domain
CTnDOT	BTH033971	-	-
CTnDOT	BTH033972	PF00589	Phage integrase family
CTnDOT	BTH033973	-	-
CTnDOT	BTH033974	PF12728	Helix-turn-helix domain
CTnDOT	BTH033975	PF13481	AAA domain
CTnDOT	BTH033976	PF13155	Toprim-like
CTnDOT	BTH033977	PF05713	Bacterial mobilisation protein (MobC)
CTnDOT	BTH033978	PF03432	Relaxase/Mobilisation nuclease domain
CTnDOT	BTH033979	-	-
CTnDOT	BTH033980	PF04439	Streptomycin adenylyltransferase
CTnDOT	BTH033981	PF01494, PF13450	FAD binding domain, NAD(P)-binding Rossmann-like domain
CTnDOT	BTH033982	PF01494, PF13450, PF13241, PF00070	FAD binding domain, NAD(P)-binding Rossmann-like domain, Putative NAD(P)-binding, Pyridine nucleotide-disulphide oxidoreductase
CTnDOT	BTH033983	PF00398, PF13659, PF01135	Ribosomal RNA adenine dimethylase, Methyltransferase domain, Protein-L-isoaspartate(D-aspartate) O-methyltransferase (PCMT)
CTnDOT	BTH033984	PF00589, PF13495	Phage integrase family, Phage integrase, N-terminal SAM-like domain
CTnDOT	BTH033985	PF12728	Helix-turn-helix domain
CTnDOT	BTH033986	-	-
CTnDOT	BTH033987	PF12728	Helix-turn-helix domain
CTnDOT	BTH033988	PF12728, PF13411	Helix-turn-helix domain, MerR HTH family regulatory protein
CTnDOT	BTH033989	PF13351, PF13101	Protein of unknown function (DUF4099), Protein of unknown function (DUF3945)
CTnDOT	BTH033990	PF01131, PF13342, PF01751	DNA topoisomerase, C-terminal repeat of topoisomerase, Toprim domain
CTnDOT	BTH0321m	-	-
CTnDOT	BTH033991	PF08989	Domain of unknown function (DUF1896)
CTnDOT	BTH033994	PF02518, PF00072, PF00512	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase, Response regulator receiver domain, His Kinase A (phospho-acceptor) domain

CTnDOT	BTH032405	PF00158, PF02954, PF14532	Sigma-54 interaction domain, Bacterial regulatory protein, Fis family, Sigma-54 interaction domain
CTnDOT	BTH032402	PF13274	Protein of unknown function (DUF4065)
CTnDOT	BTH032401	-	-
CTnDOT	BTH032398	-	-
CTnDOT	BTH032396	PF11888	Protein of unknown function (DUF3408)
CTnDOT	BTH032393	PF13572	Domain of unknown function (DUF4134)
CTnDOT	BTH032392	PF13571	Domain of unknown function (DUF4133)
CTnDOT	BTH032390	PF12992	Domain of unknown function, B. Theta Gene description (DUF3876)
CTnDOT	BTH032389	PF13605	Domain of unknown function (DUF4141)
CTnDOT	BTH032388	PF07863	Homologues of TraJ from Bacteroides conjugative transposon
CTnDOT	BTH032387	-	-
CTnDOT	BTH032386	PF13150	Protein of unknown function (DUF3989)
CTnDOT	BTH032385	PF12508	Protein of unknown function (DUF3714)
CTnDOT	BTH032384	PF13595	Domain of unknown function (DUF4138)
CTnDOT	BTH032383	PF10626	Conjugative transposon protein TraO
CTnDOT	BTH032382	PF13155, PF01807	Toprim-like, CHC2 zinc finger
CTnDOT	BTH032381	PF12988	Domain of unknown function, B. Theta Gene description (DUF3872)
CTnDOT	BTH032380	PF00589	Phage integrase family
CTnDOT	BTH032379	PF00589, PF12835	Phage integrase family, Integrase
CTnDOT	BTH032378	PF00589, PF02899, PF13495	Phage integrase family, Phage integrase, N-terminal SAM-like domain, Phage integrase, N-terminal SAM-like domain
CTnDOT	BTH032377	PF01551	Peptidase family M23
CTnDOT	BTH032376	-	-
CTnDOT	BTH032375	-	-
CTnDOT	BTH032374	-	-
CTnDOT	BTH032373	-	-
CTnDOT	BTH032372	-	-
CTnDOT	BTH032371	PF12989	Domain of unknown function, B. Theta Gene description (DUF3873)
CTnDOT	BTH032370	-	-

CTnDOT	BTH032369	PF14284	PcfJ-like protein
CTnDOT	BTH032368	PF14058	PcfK-like protein
CTnDOT	BTH032367	-	-
CTnDOT	BTH032366	-	-
CTnDOT	BTH032365	-	-
Bacteroides dorei DSM 17855	BACDOR01615	-	-
Bacteroides dorei DSM 17855	BACDOR01616	-	-
Bacteroides dorei DSM 17855	BACDOR01617	-	-
Bacteroides dorei DSM 17855	BACDOR01618	-	-
Bacteroides dorei DSM 17855	BACDOR01619	-	-
Bacteroides dorei DSM 17855	BACDOR01620	-	-
Bacteroides dorei DSM 17855	BACDOR01621	-	-
Bacteroides dorei DSM 17855	BACDOR01622	-	-
Bacteroides dorei DSM 17855	BACDOR01623	-	-
Bacteroides dorei DSM 17855	BACDOR01624	-	-
Bacteroides dorei DSM 17855	BACDOR01625	-	-
Bacteroides dorei DSM 17855	BACDOR01626	-	-
Bacteroides dorei DSM 17855	BACDOR01629	PF02518, PF00512, PF00072	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase, His Kinase A (phosphoacceptor) domain, Response regulator receiver domain
Bacteroides dorei DSM 17855	BACDOR01630	PF00158, PF00072, PF02954, PF07728	Sigma-54 interaction domain, Response regulator receiver domain, Bacterial regulatory protein, Fis family, ATPase

			family associated with various cellular activities (AAA)
Bacteroides dorei DSM 17855	BACDOR01632	-	-
Bacteroides dorei DSM 17855	BACDOR01634	-	-
Bacteroides dorei DSM 17855	BACDOR01635	PF07693	KAP family P-loop domain
Bacteroides dorei DSM 17855	BACDOR01638	-	-
Bacteroides dorei DSM 17855	BACDOR01640	-	-
Bacteroides dorei DSM 17855	BACDOR01642	-	-
Bacteroides dorei DSM 17855	BACDOR01644	-	-
Bacteroides dorei DSM 17855	BACDOR01645	-	-
Bacteroides dorei DSM 17855	BACDOR01646	-	-
Bacteroides dorei DSM 17855	BACDOR01648	-	-
Bacteroides dorei DSM 17855	BACDOR01649	-	-
Bacteroides dorei DSM 17855	BACDOR01650	PF07863	Homologues of TraJ from Bacteroides conjugative transposon
Bacteroides dorei DSM 17855	BACDOR01651	-	-
Bacteroides dorei DSM 17855	BACDOR01652	-	-
Bacteroides dorei DSM 17855	BACDOR01653	-	-
Bacteroides dorei DSM 17855	BACDOR01654	-	-

Bacteroides dorei DSM 17855	BACDOR01655	PF10626	Conjugative transposon protein TraO
Bacteroides dorei DSM 17855	BACDOR01656	PF01807	CHC2 zinc finger
Bacteroides dorei DSM 17855	BACDOR01657	-	-
Bacteroides dorei DSM 17855	BACDOR01658	-	-
Bacteroides dorei DSM 17855	BACDOR01659	-	-
Bacteroides dorei DSM 17855	BACDOR01660	-	-
Bacteroides dorei DSM 17855	BACDOR01661	-	-
Bacteroides dorei DSM 17855	BACDOR01662	-	-
Bacteroides dorei DSM 17855	BACDOR01663	-	-
Bacteroides dorei DSM 17855	BACDOR01664	-	-
Bacteroides dorei DSM 17855	BACDOR01665	-	-
Bacteroides dorei DSM 17855	BACDOR01666	-	-
Bacteroides dorei DSM 17855	BACDOR01667	-	-
Bacteroides dorei DSM 17855	BACDOR01668	-	-
Bacteroides dorei DSM 17855	BACDOR01669	-	-
Bacteroides dorei DSM 17855	BACDOR01670	-	-
Bacteroides coprocola DSM 17136	BACCOP01817	PF00589	Phage integrase family

Bacteroides coprocola DSM 17136	BACCOP01818	-	-
Bacteroides coprocola DSM 17136	BACCOP01819	-	-
Bacteroides coprocola DSM 17136	BACCOP01820	-	-
Bacteroides coprocola DSM 17136	BACCOP01822	-	-
Bacteroides coprocola DSM 17136	BACCOP01821	-	-
Bacteroides coprocola DSM 17136	BACCOP01823	-	-
Bacteroides coprocola DSM 17136	BACCOP01824	-	-
Bacteroides coprocola DSM 17136	BACCOP01825	PF01131, PF01751	DNA topoisomerase, Toprim domain
Bacteroides coprocola DSM 17136	BACCOP01826	PF08989	Domain of unknown function (DUF1896)
Bacteroides coprocola DSM 17136	BACCOP01829	PF02518, PF00512, PF00072	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase, His Kinase A (phosphoacceptor) domain, Response regulator receiver domain
Bacteroides coprocola DSM 17136	BACCOP01830	PF00158, PF00072, PF02954, PF07728	Sigma-54 interaction domain, Response regulator receiver domain, Bacterial regulatory protein, Fis family, ATPase family associated with various cellular activities (AAA)
Bacteroides coprocola DSM 17136	BACCOP01832	-	-
Bacteroides coprocola DSM 17136	BACCOP01834	PF01609	Transposase DDE domain
Bacteroides coprocola DSM 17136	BACCOP01835	-	-
Bacteroides coprocola DSM 17136	BACCOP01836	-	-
Bacteroides coprocola DSM 17136	BACCOP01837	PF07693	KAP family P-loop domain

Bacteroides coprocola DSM 17136	BACCOP01840	-	-
Bacteroides coprocola DSM 17136	BACCOP01842	-	-
Bacteroides coprocola DSM 17136	BACCOP01844	-	-
Bacteroides coprocola DSM 17136	BACCOP01846	-	-
Bacteroides coprocola DSM 17136	BACCOP01847	-	-
Bacteroides coprocola DSM 17136	BACCOP01849	-	-
Bacteroides coprocola DSM 17136	BACCOP01850	-	-
Bacteroides coprocola DSM 17136	BACCOP01851	PF07863	Homologues of TraJ from Bacteroides conjugative transposon
Bacteroides coprocola DSM 17136	BACCOP01852	-	-
Bacteroides coprocola DSM 17136	BACCOP01853	-	-
Bacteroides coprocola DSM 17136	BACCOP01854	-	-
Bacteroides coprocola DSM 17136	BACCOP01855	-	-
Bacteroides coprocola DSM 17136	BACCOP01856	PF10626	Conjugative transposon protein TraO
Bacteroides coprocola DSM 17136	BACCOP01857	PF01807	CHC2 zinc finger
Bacteroides coprocola DSM 17136	BACCOP01858	-	-
Bacteroides coprocola DSM 17136	BACCOP01859	-	-
Bacteroides coprocola DSM 17136	BACCOP01860	-	-

Parabacteroides merdae ATCC 43184			
Parabacteroides merdae ATCC 43184			
Parabacteroides merdae ATCC 43184	PARMER00775	PF00589	Phage integrase family
Parabacteroides merdae ATCC 43184	PARMER00776	-	-
Parabacteroides merdae ATCC 43184	PARMER00777	-	-
Parabacteroides merdae ATCC 43184	PARMER00778	-	-
Parabacteroides merdae ATCC 43184	PARMER00780	-	-
Parabacteroides merdae ATCC 43184	PARMER00779	-	-
Parabacteroides merdae ATCC 43184	PARMER00781	-	-
Parabacteroides merdae ATCC 43184	PARMER00782	-	-
Parabacteroides merdae ATCC 43184	PARMER00783	PF01131, PF01751	DNA topoisomerase, Toprim domain
Parabacteroides merdae ATCC 43184	PARMER00784	PF08989	Domain of unknown function (DUF1896)
Parabacteroides merdae ATCC 43184	PARMER00786	-	-
Parabacteroides merdae ATCC 43184	PARMER00788	PF02518, PF00512, PF00072	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase, His Kinase A (phosphoacceptor) domain, Response regulator receiver domain
Parabacteroides merdae ATCC 43184	PARMER00789	PF00158, PF00072, PF02954, PF07728	Sigma-54 interaction domain, Response regulator receiver domain, Bacterial regulatory protein, Fis family, ATPase family associated with various cellular activities (AAA)
Parabacteroides merdae ATCC 43184	PARMER00792	-	-

Parabacteroides merdae ATCC 43184	PARMER00793	PF07693	KAP family P-loop domain
Parabacteroides merdae ATCC 43184	PARMER00796	-	-
Parabacteroides merdae ATCC 43184	PARMER00797	-	-
Parabacteroides merdae ATCC 43184	PARMER00799	-	-
Parabacteroides merdae ATCC 43184	PARMER00801	-	-
Parabacteroides merdae ATCC 43184	PARMER00803	-	-
Parabacteroides merdae ATCC 43184	PARMER00804	-	-
Parabacteroides merdae ATCC 43184	PARMER00806	-	-
Parabacteroides merdae ATCC 43184	PARMER00807	-	-
Parabacteroides merdae ATCC 43184	PARMER00808	PF07863	Homologues of TraJ from Bacteroides conjugative transposon
Parabacteroides merdae ATCC 43184	PARMER00809	-	-
Parabacteroides merdae ATCC 43184	PARMER00810	-	-
Parabacteroides merdae ATCC 43184	PARMER00811	-	-
Parabacteroides merdae ATCC 43184	PARMER00812	-	-
Parabacteroides merdae ATCC 43184	PARMER00813	PF10626	Conjugative transposon protein TraO
Parabacteroides merdae ATCC 43184	PARMER00814	PF01807	CHC2 zinc finger
Parabacteroides merdae ATCC 43184	PARMER00815	-	-

Parabacteroides merdae ATCC 43184	PARMER00816	-	-
Parabacteroides merdae ATCC 43184	PARMER00817	-	-
Parabacteroides merdae ATCC 43184	PARMER00818	-	-
Parabacteroides merdae ATCC 43184	PARMER00819	-	-
Parabacteroides merdae ATCC 43184	PARMER00820	-	-
Parabacteroides merdae ATCC 43184	PARMER00821	-	-
Parabacteroides merdae ATCC 43184	PARMER00822	-	-
Parabacteroides merdae ATCC 43184	PARMER00823	-	-
Parabacteroides merdae ATCC 43184	PARMER00824	-	-
Parabacteroides merdae ATCC 43184	PARMER00825	-	-
Parabacteroides merdae ATCC 43184	PARMER00827	-	-
Parabacteroides merdae ATCC 43184	PARMER00826	-	-
Bacteroides stercoris ATCC 43183	BACSTE01062	-	-
Bacteroides stercoris ATCC 43183	BACSTE01063	-	-
Bacteroides stercoris ATCC 43183	BACSTE01064	-	-
Bacteroides stercoris ATCC 43183	BACSTE01066	-	-
Bacteroides stercoris ATCC 43183	BACSTE01065	-	-

Bacteroides stercoris ATCC 43183	BACSTE01067	-	-
Bacteroides stercoris ATCC 43183	BACSTE01068	-	-
Bacteroides stercoris ATCC 43183	BACSTE01069	-	-
Bacteroides stercoris ATCC 43183	BACSTE01070	-	-
Bacteroides stercoris ATCC 43183	BACSTE01071	-	-
Bacteroides stercoris ATCC 43183	BACSTE01072	-	-
Bacteroides stercoris ATCC 43183	BACSTE01073	-	-
Bacteroides stercoris ATCC 43183	BACSTE01074	-	-
Bacteroides stercoris ATCC 43183	BACSTE01075	-	-
Bacteroides stercoris ATCC 43183	BACSTE01076	-	-
Bacteroides stercoris ATCC 43183	BACSTE01077	PF01807	CHC2 zinc finger
Bacteroides stercoris ATCC 43183	BACSTE01078	PF10626	Conjugative transposon protein TraO
Bacteroides stercoris ATCC 43183	BACSTE01079	-	-
Bacteroides stercoris ATCC 43183	BACSTE01080	-	-
Bacteroides stercoris ATCC 43183	BACSTE01081	-	-
Bacteroides stercoris ATCC 43183	BACSTE01082	-	-
Bacteroides stercoris ATCC 43183	BACSTE01083	PF07863	Homologues of TraJ from Bacteroides conjugative transposon

Bacteroides stercoris ATCC 43183	BACSTE01084	-	-
Bacteroides stercoris ATCC 43183	BACSTE01085	-	-
Bacteroides stercoris ATCC 43183	BACSTE01087	-	-
Bacteroides stercoris ATCC 43183	BACSTE01088	-	-
Bacteroides stercoris ATCC 43183	BACSTE01089	-	-
Bacteroides stercoris ATCC 43183	BACSTE01091	-	-
Bacteroides stercoris ATCC 43183	BACSTE01094	-	-
Bacteroides stercoris ATCC 43183	BACSTE01095	-	-
Bacteroides stercoris ATCC 43183	BACSTE01098	PF07693	KAP family P-loop domain
Bacteroides stercoris ATCC 43183	BACSTE01099	-	-
Bacteroides stercoris ATCC 43183	BACSTE01101	-	-
Bacteroides stercoris ATCC 43183	BACSTE01103	PF00158, PF00072, PF02954, PF07728	Sigma-54 interaction domain, Response regulator receiver domain, Bacterial regulatory protein, Fis family, ATPase family associated with various cellular activities (AAA)
Bacteroides stercoris ATCC 43183	BACSTE01104	PF02518, PF00512, PF00072	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase, His Kinase A (phosphoacceptor) domain, Response regulator receiver domain
Bacteroides stercoris ATCC 43183	BACSTE01107	PF08989	Domain of unknown function (DUF1896)
Bacteroides stercoris ATCC 43183	BACSTE01108	PF01131, PF01751	DNA topoisomerase, Toprim domain
Bacteroides stercoris ATCC 43183	BACSTE01109	-	-

Bacteroides stercoris ATCC 43183	BACSTE01110	-	-
Bacteroides stercoris ATCC 43183	BACSTE01111	-	-
Bacteroides stercoris ATCC 43183	BACSTE01113	-	-
Bacteroides stercoris ATCC 43183	BACSTE01112	-	-
Bacteroides stercoris ATCC 43183	BACSTE01114	-	-
Bacteroides stercoris ATCC 43183	BACSTE01115	-	-
Bacteroides stercoris ATCC 43183	BACSTE01116	PF00589	Phage integrase family
Bacteroides intestinalis DSM 17393	BACINT00772	-	-
Bacteroides intestinalis DSM 17393	BACINT00773	PF01011, PF00149	PQQ enzyme repeat, Calcineurin-like phosphoesterase
Bacteroides intestinalis DSM 17393	BACINT00774	-	-
Bacteroides intestinalis DSM 17393	BACINT00775	-	-
Bacteroides intestinalis DSM 17393	BACINT00776	-	-
Bacteroides intestinalis DSM 17393	BACINT00777	PF00578, PF08534	AhpC/TSA family, Redoxin
Bacteroides intestinalis DSM 17393	BACINT00778	-	-
Bacteroides intestinalis DSM 17393	BACINT00779	-	-
Bacteroides intestinalis DSM 17393	BACINT00780	-	-
Bacteroides intestinalis DSM 17393	BACINT00781	-	-

Bacteroides intestinalis DSM 17393	BACINT00782	-	-
Bacteroides intestinalis DSM 17393	BACINT00783	-	-
Bacteroides intestinalis DSM 17393	BACINT00784	-	-
Bacteroides intestinalis DSM 17393	BACINT00785	-	-
Bacteroides intestinalis DSM 17393	BACINT00786	-	-
Bacteroides intestinalis DSM 17393	BACINT00787	-	-
Bacteroides intestinalis DSM 17393	BACINT00788	-	-
Bacteroides intestinalis DSM 17393	BACINT00789	-	-
Bacteroides intestinalis DSM 17393	BACINT00790	-	-
Bacteroides intestinalis DSM 17393	BACINT00791	-	-
Bacteroides intestinalis DSM 17393	BACINT00792	PF01807	CHC2 zinc finger
Bacteroides intestinalis DSM 17393	BACINT00793	PF10626	Conjugative transposon protein TraO
Bacteroides intestinalis DSM 17393	BACINT00794	-	-
Bacteroides intestinalis DSM 17393	BACINT00795	-	-
Bacteroides intestinalis DSM 17393	BACINT00796	-	-
Bacteroides intestinalis DSM 17393	BACINT00797	-	-
Bacteroides intestinalis DSM 17393	BACINT00798	PF07863	Homologues of TraJ from Bacteroides conjugative transposon

Bacteroides intestinalis DSM 17393	BACINT00799	-	-
Bacteroides intestinalis DSM 17393	BACINT00800	-	-
Bacteroides intestinalis DSM 17393	BACINT00802	-	-
Bacteroides intestinalis DSM 17393	BACINT00803	-	-
Bacteroides intestinalis DSM 17393	BACINT00805	-	-
Bacteroides intestinalis DSM 17393	BACINT00808	-	-
Bacteroides intestinalis DSM 17393	BACINT00811	PF07693	KAP family P-loop domain
Bacteroides intestinalis DSM 17393	BACINT00812	-	-
Bacteroides intestinalis DSM 17393	BACINT00815	PF00158, PF00072, PF02954, PF07728	Sigma-54 interaction domain, Response regulator receiver domain, Bacterial regulatory protein, Fis family, ATPase family associated with various cellular activities (AAA)
Bacteroides intestinalis DSM 17393	BACINT00816	PF02518, PF00512, PF00072	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase, His Kinase A (phosphoacceptor) domain, Response regulator receiver domain
Bacteroides intestinalis DSM 17393	BACINT00819	PF08989	Domain of unknown function (DUF1896)
Bacteroides intestinalis DSM 17393	BACINT00820	PF01131, PF01751	DNA topoisomerase, Toprim domain
Bacteroides intestinalis DSM 17393	BACINT00821	-	-
Bacteroides intestinalis DSM 17393	BACINT00822	-	-
Bacteroides intestinalis DSM 17393	BACINT00823	-	-
Bacteroides intestinalis DSM 17393	BACINT00824	-	-

Bacteroides intestinalis DSM 17393	BACINT00825	-	-
Bacteroides intestinalis DSM 17393	BACINT00826	-	-
Bacteroides intestinalis DSM 17393	BACINT00827	-	-
Bacteroides intestinalis DSM 17393	BACINT00828	PF00589	Phage integrase family
Bacteroides intestinalis DSM 17393	BACINT00829	-	-
Bacteroides intestinalis DSM 17393	BACINT00830	-	-
Bacteroides intestinalis DSM 17393	BACINT00831	-	-
Bacteroides intestinalis DSM 17393	BACINT00832	PF00593, PF07715	TonB dependent receptor, TonB-dependent Receptor Plug Domain
Bacteroides intestinalis DSM 17393	BACINT00833	-	-
Bacteroides intestinalis DSM 17393	BACINT00834	-	-
Bacteroides intestinalis DSM 17393	BACINT00835	PF01928	CYTH domain
Bacteroides plebeius DSM 17135	BACPLE03530	PF00589	Phage integrase family
Bacteroides plebeius DSM 17135	BACPLE03531	PF00078, PF01348	Reverse transcriptase (RNA-dependent DNA polymerase), Type II intron maturase
Bacteroides plebeius DSM 17135	BACPLE03532	-	-
Bacteroides plebeius DSM 17135	BACPLE03533	-	-
Bacteroides plebeius DSM 17135	BACPLE03534	-	-
Bacteroides plebeius DSM 17135	BACPLE03535	-	-

Bacteroides plebeius DSM 17135	BACPLE03536	-	-
Bacteroides plebeius DSM 17135	BACPLE03537	-	-
Bacteroides plebeius DSM 17135	BACPLE03538	-	-
Bacteroides plebeius DSM 17135	BACPLE03539	PF01131, PF01751	DNA topoisomerase, Toprim domain
Bacteroides plebeius DSM 17135	BACPLE03540	PF08989	Domain of unknown function (DUF1896)
Bacteroides plebeius DSM 17135	BACPLE03543	PF02518, PF00512, PF00072	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase, His Kinase A (phosphoacceptor) domain, Response regulator receiver domain
Bacteroides plebeius DSM 17135	BACPLE03544	PF00158, PF00072, PF02954, PF07728	Sigma-54 interaction domain, Response regulator receiver domain, Bacterial regulatory protein, Fis family, ATPase family associated with various cellular activities (AAA)
Bacteroides plebeius DSM 17135	BACPLE03547	-	-
Bacteroides plebeius DSM 17135	BACPLE03548	PF07693	KAP family P-loop domain
Bacteroides plebeius DSM 17135	BACPLE03550	PF00078, PF08388	Reverse transcriptase (RNA-dependent DNA polymerase), Group II intron, maturase-specific domain
Bacteroides plebeius DSM 17135	BACPLE03553	-	-
Bacteroides plebeius DSM 17135	BACPLE03554	-	-
Bacteroides plebeius DSM 17135	BACPLE03557	-	-
Bacteroides plebeius DSM 17135	BACPLE03559	-	-
Bacteroides plebeius DSM 17135	BACPLE03560	-	-
Bacteroides plebeius DSM 17135	BACPLE03561	-	-

Bacteroides plebeius DSM 17135	BACPLE03563	-	-
Bacteroides plebeius DSM 17135	BACPLE03564	-	-
Bacteroides plebeius DSM 17135	BACPLE03565	PF07863	Homologues of TraJ from Bacteroides conjugative transposon
Bacteroides plebeius DSM 17135	BACPLE03566	-	-
Bacteroides plebeius DSM 17135	BACPLE03567	-	-
Bacteroides plebeius DSM 17135	BACPLE03568	-	-
Bacteroides plebeius DSM 17135	BACPLE03569	-	-
Bacteroides plebeius DSM 17135	BACPLE03570	PF10626	Conjugative transposon protein TraO
Bacteroides plebeius DSM 17135	BACPLE03571	PF01807	CHC2 zinc finger
Bacteroides plebeius DSM 17135	BACPLE03572	-	-
Bacteroides plebeius DSM 17135	BACPLE03573	-	-
Bacteroides plebeius DSM 17135	BACPLE03574	-	-
Bacteroides plebeius DSM 17135	BACPLE03575	-	-
Bacteroides plebeius DSM 17135	BACPLE03576	-	-
Bacteroides plebeius DSM 17135	BACPLE03577	-	-
Bacteroides plebeius DSM 17135	BACPLE03578	-	-
Bacteroides plebeius DSM 17135	BACPLE03579	-	-

Bacteroides plebeius DSM 17135	BACPLE03580	-	-
Bacteroides plebeius DSM 17135	BACPLE03581	-	-
Bacteroides plebeius DSM 17135	BACPLE03582	-	-
Bacteroides plebeius DSM 17135	BACPLE03583	-	-
Bacteroides eggerthii DSM 20697	BACEGG03049	PF00589	Phage integrase family
Bacteroides eggerthii DSM 20697	BACEGG03050	-	-
Bacteroides eggerthii DSM 20697	BACEGG03051	-	-
Bacteroides eggerthii DSM 20697	BACEGG03052	-	-
Bacteroides eggerthii DSM 20697	BACEGG03054	-	-
Bacteroides eggerthii DSM 20697	BACEGG03053	-	-
Bacteroides eggerthii DSM 20697	BACEGG03055	-	-
Bacteroides eggerthii DSM 20697	BACEGG03056	-	-
Bacteroides eggerthii DSM 20697	BACEGG03057	PF01131, PF01751	DNA topoisomerase, Toprim domain
Bacteroides eggerthii DSM 20697	BACEGG03058	PF08989	Domain of unknown function (DUF1896)
Bacteroides eggerthii DSM 20697	BACEGG03061	PF02518, PF00512, PF00072	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase, His Kinase A (phosphoacceptor) domain, Response regulator receiver domain
Bacteroides eggerthii DSM 20697	BACEGG03062	PF00158, PF00072, PF02954, PF07728	Sigma-54 interaction domain, Response regulator receiver domain, Bacterial regulatory protein, Fis family, ATPase family associated with various cellular activities (AAA)

Bacteroides eggerthii DSM 20697	BACEGG03064	-	-
Bacteroides eggerthii DSM 20697	BACEGG03066	-	-
Bacteroides eggerthii DSM 20697	BACEGG03067	PF07693	KAP family P-loop domain
Bacteroides eggerthii DSM 20697	BACEGG03070	-	-
Bacteroides eggerthii DSM 20697	BACEGG03071	-	-
Bacteroides eggerthii DSM 20697	BACEGG03074	-	-
Bacteroides eggerthii DSM 20697	BACEGG03076	-	-
Bacteroides eggerthii DSM 20697	BACEGG03077	-	-
Bacteroides eggerthii DSM 20697	BACEGG03079	-	-
Bacteroides eggerthii DSM 20697	BACEGG03080	-	-
Bacteroides eggerthii DSM 20697	BACEGG03081	PF07863	Homologues of TraJ from Bacteroides conjugative transposon
Bacteroides eggerthii DSM 20697	BACEGG03082	-	-
Bacteroides eggerthii DSM 20697	BACEGG03083	-	-
Bacteroides eggerthii DSM 20697	BACEGG03084	-	-
Bacteroides eggerthii DSM 20697	BACEGG03085	-	-
Bacteroides eggerthii DSM 20697	BACEGG03086	PF10626	Conjugative transposon protein TraO
Bacteroides eggerthii DSM 20697	BACEGG03087	PF01807	CHC2 zinc finger

Bacteroides eggerthii DSM 20697	BACEGG03088	-	-
Bacteroides eggerthii DSM 20697	BACEGG03089	-	-
Bacteroides eggerthii DSM 20697	BACEGG03090	-	-
Bacteroides eggerthii DSM 20697	BACEGG03091	-	-
Bacteroides eggerthii DSM 20697	BACEGG03092	-	-
Bacteroides eggerthii DSM 20697	BACEGG03093	-	-
Bacteroides eggerthii DSM 20697	BACEGG03094	-	-
Bacteroides eggerthii DSM 20697	BACEGG03095	-	-
Bacteroides eggerthii DSM 20697	BACEGG03096	-	-
Bacteroides eggerthii DSM 20697	BACEGG03097	-	-
Bacteroides eggerthii DSM 20697	BACEGG03098	-	-
Bacteroides eggerthii DSM 20697	BACEGG03099	-	-
Bacteroides eggerthii DSM 20697	BACEGG03100	PF00216	Bacterial DNA-binding protein
Bacteroides eggerthii DSM 20697	BACEGG03101	PF01694	Rhomboid family
Bacteroides eggerthii DSM 20697	BACEGG03102	PF01694	Rhomboid family
Bacteroides eggerthii DSM 20697	BACEGG03103	PF03372	Endonuclease/Exonuclease/phosphatase family
Bacteroides eggerthii DSM 20697	BACEGG03104	PF01432	Peptidase family M3

Bacteroides eggerthii DSM 20697	BACEGG03105	PF02355, PF07549	Protein export membrane protein, SecD/SecF GG Motif
Bacteroides eggerthii DSM 20697	BACEGG03106	PF00027	Cyclic nucleotide-binding domain
Bacteroides eggerthii DSM 20697	BACEGG03107	-	-
Bacteroides eggerthii DSM 20697	BACEGG03108	PF07715	TonB-dependent Receptor Plug Domain
Bacteroides eggerthii DSM 20697	BACEGG03109	PF07980	RagB, SusD and hypothetical proteins
Bacteroides eggerthii DSM 20697	BACEGG03110	-	-
Bacteroides eggerthii DSM 20697	BACEGG03111	-	-
Bacteroides eggerthii DSM 20697	BACEGG03112	PF00560	Leucine Rich Repeat
Bacteroides eggerthii DSM 20697	BACEGG03113	PF00560	Leucine Rich Repeat
Bacteroides eggerthii DSM 20697	BACEGG03114	PF00082, PF02368	Subtilase family, Bacterial Ig-like domain (group 2)
Bacteroides eggerthii DSM 20697	BACEGG03115	PF07610	Protein of unknown function (DUF1573)
Bacteroides eggerthii DSM 20697	BACEGG03116	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01768	PF04011	LemA family
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01769	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01770	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01772	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01771	-	-

Parabacteroides johnsonii DSM 18315	PRABACTJOHN01773	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01774	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01775	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01776	PF01751	Toprim domain
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01777	PF08989	Domain of unknown function (DUF1896)
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01780	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01782	PF00529	HlyD family secretion protein
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01783	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01784	PF00873	AcrB/AcrD/AcrF family
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01785	PF02321	Outer membrane efflux protein
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01786	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01787	PF02687	Predicted permease
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01788	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01790	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01792	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01793	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01794	-	-

Parabacteroides johnsonii DSM 18315	PRABACTJOHN01795	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01798	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01800	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01802	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01804	-	-
Parabacteroides johnsonii DSM 18315	PRABACTJOHN01806	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732569	PF07863	Homologues of TraJ from Bacteroides conjugative transposon
Bacteroides thetaiotaomicron 3731	Bthe3732570	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732571	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732572	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732574	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732575	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732578	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732580	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732581	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732584	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732585	-	-

Bacteroides thetaiotaomicron 3731	Bthe3732588	PF00158, PF00072, PF02954, PF07728	Sigma-54 interaction domain, Response regulator receiver domain, Bacterial regulatory protein, Fis family, ATPase family associated with various cellular activities (AAA)
Bacteroides thetaiotaomicron 3731	Bthe3732589	PF02518, PF00512, PF00072	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase, His Kinase A (phosphoacceptor) domain, Response regulator receiver domain
Bacteroides thetaiotaomicron 3731	Bthe3732592	PF08989	Domain of unknown function (DUF1896)
Bacteroides thetaiotaomicron 3731	Bthe3732593	PF01131, PF01751	DNA topoisomerase, Toprim domain
Bacteroides thetaiotaomicron 3731	Bthe3732594	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732595	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732596	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732597	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732598	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732599	-	-
Bacteroides thetaiotaomicron 3731	Bthe3732600	PF00589	Phage integrase family
Bacteroides thetaiotaomicron 3731	Bthe3735212	-	-
Bacteroides thetaiotaomicron 3731	Bthe3735213	-	-
Bacteroides thetaiotaomicron 3731	Bthe3735214	-	-
Bacteroides thetaiotaomicron 3731	Bthe3735215	-	-
Bacteroides thetaiotaomicron 3731	Bthe3735216	-	-

Bacteroides thetaiotaomicron 3731	Bthe3735217	-	-
Bacteroides thetaiotaomicron 3731	Bthe3735218	-	-
Bacteroides thetaiotaomicron 3731	Bthe3735219	PF10771	Protein of unknown function (DUF2582)

CHAPTER 4. Conclusions

From pathogen resistance, metabolism, and modulation of the immune system, the impact the gut microbiome has in the health of its respective hosts is undeniable (Bull & Plummer 2014). HGT is major contributor to the general fitness of bacteria that colonize and persist in the gut, allowing the rapid spread of new genes that may be favorable to natural selection (Vogan & Higgs 2011). Research pertaining to HGT has truly advanced since the discovery of the first recorded instance of HGT in *Diplococcus pneumoniae* in 1928 by Frederick Griffith, where we now know the tremendous diversity of HGT, from their method of mobilization and the accessory genes they can bring along with them (Griffith 1928, Boto 2009, Jain et al. 2003, Niehus et al. 2015 Jiang et al. 2019). However, the majority of these MGEs lack thorough characterization concerning their mobility and potential fitness effects. Also, while various computational methods exist for identifying potential MGEs, confirming their transferability requires additional experimental approaches.

In chapter 2, we developed the transposon mutagenesis mobilization method (TMMM) to immediately identify and confirm the mobility of MGEs without the need for targeted mutations. We demonstrated the efficacy of this method by identifying two CTns: *Bo*CTn previously discovered in *B. ovatus* and a novel CTn, *Pv*CTn, identified in *Phocaeicola vulgatus*. The discovery of *Pv*CTn led us to observe that the transposon-induced mutations all inserted into *BVU3433*, a gene encoding a helix-turn-helix-structural motif known for regulate gene expression through repression. Our *in vitro*

conjugation experiments confirmed the increased conjugation efficiency of *Pv*CTn upon *BVU3433* deletion. Additionally, our transcriptomics data revealed the general downregulation throughout *Pv*CTn caused by BVU3433, including genes involved in forming the T4SS. Finally, the analysis of individual strain genomes and community metagenomes demonstrated how common *Pv*CTn-like elements with putative BVU3433 homologues are in human gut-associated bacteria. Overall, our findings introduce a novel strategy for detecting transfer events *in vitro*, which can be slightly modified for application to other bacterial systems and our discovery of the persistence of *Pv*CTn across community metagenomes suggests its importance as a fitness determinant.

In chapter 3, we characterized the sequence, structural, and functional diversity of CTnDOT-like elements. We identified the *in vitro* transfer of a novel CTnDOT-like element, PmDOT, and determined that both CTnDOT and PmDOT possess conserved genes responsible for tetracycline resistance (*tetQ*), the generation of the T4SS (*tra*), and the excision and conjugative transfer of the CTn (*rteB, rteC, xis2c, xis2d, exc, intDOT*). We observed 9 differences in the amino acid contents of intDOT, with 8/9 being conservative amino acid changes that do not alter the predicted structure, suggesting similar functionality throughout. However, PmDOT lacks an integron adjacent to *intDOT* containing genes for streptomycin (*aadS*) and erythromycin (*ermF*) resistance. Furthermore, we investigated the regulatory mechanisms of CTnDOT-like elements, finding that 7/9 CTns regulate similarly to CTnDOT, except for those in *Ph. dorei* and *P. johnsonii*. Notably, *Ph. dorei* lacks the entire excision operon and intDOT, indicating a deficiency in mobilization genes. Similarly, *P. johnsonii* lacks essential genes for

complete transfer (e.g. *intDOT*, *rteA*, *rteB*). Despite this, *in vitro* conjugations demonstrated the transfer of *P. johnsonii* into *B. thetaiotaomicron* VPI-5482, suggesting distinct regulatory mechanisms compared to CTnDOT. Additionally, we defined 10 universally conserved core genes within the CTn family, centrally localized within the 25kb region of the 64kb element. Utilizing this core gene panel, we identified 77 predicted CTnDOT-like elements from 124 Bacteroidota genomes. Our analysis revealed the persistence of conserved *tetQ* genes and the circulation of most CTnDOT-like clusters, including CTnDOT and PmDOT, among gut microbe strains for over five decades. This persistence across various time points and geographic locations underscores the importance of CTnDOT-like elements in shaping the fitness of Bacteroidota in the human gut.

REFERENCES

Acman, M., Wang, R., van Dorp, L., Shaw, L. P., Wang, Q., Luhmann, N., et al. (2022). Role of mobile genetic elements in the global dissemination of the carbapenem resistance gene blaNDM. *Nat. Commun.* 13, 1131. doi: 10.1038/s41467-022-28819-2.

Akhter, S., Aziz, R. K., and Edwards, R. A. (2012). PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* 40, e126–e126. doi: 10.1093/nar/gks406.

Amor, J. C., Swails, J., Zhu, X., Roy, C. R., Nagai, H., Ingmundson, A., et al. (2005). The structure of RalF, an ADP-ribosylation factor guanine nucleotide exchange factor from Legionella pneumophila, reveals the presence of a cap over the active site. *J. Biol. Chem.* 280, 1392–1400. doi: 10.1074/jbc.M410820200.

Anthony, W. E., Wang, B., Sukhum, K. V., D'Souza, A. W., Hink, T., Cass, C., et al. (2022). Acute and persistent effects of commonly used antibiotics on the gut microbiome and resistome in healthy adults. *Cell Rep.* 39, 110649. doi: 10.1016/j.celrep.2022.110649.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M., and Iyer, L. M. (2005). The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.* 29, 231–262. doi: 10.1016/j.femsre.2004.12.008.

Aravind, L., Iyer, L. M., Leipe, D. D., and Koonin, E. V. (2004). A novel family of P-loop NTPases with an unusual phyletic distribution and transmembrane segments inserted within the NTPase domain. *Genome Biol.* 5, R30. doi: 10.1186/gb-2004-5-5-r30.

Auchtung, J. M., Lee, C. A., Monson, R. E., Lehman, A. P., and Grossman, A. D. (2005). Regulation of a Bacillus subtilis mobile genetic element by intercellular signaling and the global DNA damage response. *Proc. Natl. Acad. Sci. U. S. A.* 102, 12554–12559. doi: 10.1073/pnas.0505835102.

Avery, O. T., Macleod, C. M., and McCarty, M. (1944). 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.* 79, 137–158. doi: 10.1084/jem.79.2.137.

Ayoubi, P., Kilic, A. O., and Vijayakumar, M. N. (1991). Tn5253, the pneumococcal omega (cat tet) BM6001 element, is a composite structure of two conjugative transposons, Tn5251 and Tn5252. *J. Bacteriol.* 173, 1617–1622. doi: 10.1128/jb.173.5.1617-1622.1991.

Bachran, C., Sutherland, M., Bachran, D., and Fuchs, H. (2007). Quantification of diphtheria toxin mediated ADP-ribosylation in a solid-phase assay. *Clin. Chem.* 53, 1676–1683. doi: 10.1373/clinchem.2007.085365.

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Res.* 43, W39–49. doi: 10.1093/nar/gkv416.

Bedzyk, L. A., Shoemaker, N. B., Young, K. E., and Salyers, A. A. (1992). Insertion and excision of Bacteroides conjugative chromosomal elements. *J. Bacteriol.* 174, 166–172. doi: 10.1128/jb.174.1.166-172.1992.

Bellanger, X., Payot, S., Leblond-Bourget, N., and Guédon, G. (2014). Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol. Rev.* 38, 720–760. doi: 10.1111/1574-6976.12058.

Bobrovskyy, M., and Vanderpool, C. K. (2016). Diverse mechanisms of post-transcriptional repression by the small RNA regulator of glucose-phosphate stress. *Mol. Microbiol.* 99, 254–273. doi: 10.1111/mmi.13230.

Boiten, K. E., Kuijper, E. J., Schuele, L., van Prehn, J., Bode, L. G. M., Maat, I., et al. (2023). Characterization of mobile genetic elements in multidrug-resistant Bacteroides fragilis isolates from different hospitals in the Netherlands. *Anaerobe* 81, 102722. doi: 10.1016/j.anaerobe.2023.102722.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170.

Bookout, A. L., Cummins, C. L., Mangelsdorf, D. J., Pesola, J. M., and Kramer, M. F. (2006). High-throughput real-time quantitative reverse transcription PCR. *Curr. Protoc. Mol. Biol.* Chapter 15, Unit 15.8. doi: 10.1002/0471142727.mb1508s73.

Bose, B., Auchtung, J. M., Lee, C. A., and Grossman, A. D. (2008). A conserved anti-repressor controls horizontal gene transfer by proteolysis. *Mol. Microbiol.* 70, 570–582. doi: 10.1111/j.1365-2958.2008.06414.x.

Boto, L. (2010). Horizontal gene transfer in evolution: facts and challenges. *Proc. Biol. Sci.* 277, 819–827. doi: 10.1098/rspb.2009.1679.

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., et al. (2018). Ten things you should know about transposable elements. *Genome Biol*. 19, 199. doi: 10.1186/s13059-018-1577-z.

Brennan, R. G., and Matthews, B. W. (1989). The helix-turn-helix DNA binding motif. *J. Biol. Chem*. 264, 1903–1906. doi: 10.1016/S0021-9258(18)94115-3.

Brook, I. (1989). Anaerobic bacterial bacteremia: 12-year experience in two military hospitals. *J. Infect. Dis*. 160, 1071–1075. doi: 10.1093/infdis/160.6.1071.

Bull, M. J., and Plummer, N. T. (2014). Part 1: The Human Gut Microbiome in Health and Disease. *Integr. Med*. 13, 17–22. Available at: https://www.ncbi.nlm.nih.gov/pubmed/26770121.

Burrus, V., Pavlovic, G., Decaris, B., and Guédon, G. (2002). The ICESt1 element of Streptococcus thermophilus belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. *Plasmid* 48, 77–97. doi: 10.1016/s0147-619x(02)00102-6.

Burrus, V., and Waldor, M. K. (2003). Control of SXT integration and excision. *J. Bacteriol*. 185, 5045–5054. doi: 10.1128/JB.185.17.5045-5054.2003.

Cain, A. K., Barquist, L., Goodman, A. L., Paulsen, I. T., Parkhill, J., and van Opijnen, T. (2020). A decade of advances in transposon-insertion sequencing. *Nat. Rev. Genet*. 21, 526–540. doi: 10.1038/s41576-020-0244-x.

Campbell, D. E., Ly, L. K., Ridlon, J. M., Hsiao, A., Whitaker, R. J., and Degnan, P. H. (2020). Infection with Bacteroides Phage BV01 Alters the Host Transcriptome and Bile Acid Metabolism in a Common Human Gut Microbe. *Cell Rep*. 32, 108142. doi: 10.1016/j.celrep.2020.108142.

Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., et al. (2014). In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother*. 58, 3895–3903. doi: 10.1128/AAC.02412-14.

Carrow, H. C., Batachari, L. E., and Chu, H. (2020). Strain diversity in the microbiome: Lessons from Bacteroides fragilis. *PLoS Pathog*. 16, e1009056. doi: 10.1371/journal.ppat.1009056.

Cavalli, L. L., Lederberg, J., and Lederberg, E. M. (1953). An infective factor controlling sex compatibility in Bacterium coli. *J. Gen. Microbiol*. 8, 89–103. doi: 10.1099/00221287-8-1-89.

CDC (2023). Antimicrobial Resistance. *Centers for Disease Control and Prevention*. Available at: https://www.cdc.gov/drugresistance/index.html

Chatzidaki-Livanis, M., Geva-Zatorsky, N., and Comstock, L. E. (2016). Bacteroides fragilis type VI secretion systems use novel effector and immunity proteins to antagonize human gut Bacteroidales species. *Proc. Natl. Acad. Sci. U. S. A.* 113, 3627–3632. doi: 10.1073/pnas.1522510113.

Cheng, Q., Paszkiet, B. J., Shoemaker, N. B., Gardner, J. F., and Salyers, A. A. (2000). Integration and excision of a Bacteroides conjugative transposon, CTnDOT. *J. Bacteriol.* 182, 4035–4043. doi: 10.1128/JB.182.14.4035-4043.2000.

Cheng, Q., Sutanto, Y., Shoemaker, N. B., Gardner, J. F., and Salyers, A. A. (2001). Identification of genes required for excision of CTnDOT, a Bacteroides conjugative transposon. *Mol. Microbiol.* 41, 625–632. doi: 10.1046/j.1365-2958.2001.02519.x.

Chiang, Y. N., Penadés, J. R., and Chen, J. (2019). Genetic transduction by phages and chromosomal islands: The new and noncanonical. *PLoS Pathog.* 15, e1007878. doi: 10.1371/journal.ppat.1007878.

Ciric, L., Jasni, A., de Vries, L. E., Agersø, Y., Mullany, P., and Roberts, A. P. (2013). *The Tn916/Tn1545 Family of Conjugative Transposons*. Landes Bioscience Available at: https://www.ncbi.nlm.nih.gov/books/NBK63531/ [Accessed February 8, 2024].

Clewell, D. B., Flannagan, S. E., and Jaworski, D. D. (1995). Unconstrained bacterial promiscuity: the Tn916-Tn1545 family of conjugative transposons. *Trends Microbiol.* 3, 229–236. doi: 10.1016/s0966-842x(00)88930-1.

Coyne Michael J., Zitomersky Naamah Levy, McGuire Abigail Manson, Earl Ashlee M., and Comstock Laurie E. (2014). Evidence of Extensive DNA Transfer between Bacteroidales Species within the Human Gut. *MBio* 5, e01305–14. doi: 10.1128/mBio.01305-14.

Cullen, T. W., Schofield, W. B., Barry, N. A., Putnam, E. E., Rundell, E. A., Trent, M. S., et al. (2015). Gut microbiota. Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation. *Science* 347, 170–175. doi: 10.1126/science.1260580.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, giab008. doi: 10.1093/gigascience/giab008.

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 14, 1394–1403. doi: 10.1101/gr.2289704.

Daubin, V., and Ochman, H. (2004). Bacterial genomes as new gene homes: the genealogy of ORFans in E. coli. *Genome Res*. 14, 1036–1042. doi: 10.1101/gr.2231904.

Declerck, N., Bouillaut, L., Chaix, D., Rugani, N., Slamti, L., Hoh, F., et al. (2007). Structure of PlcR: Insights into virulence regulation and evolution of quorum sensing in Gram-positive bacteria. *Proceedings of the National Academy of Sciences* 104, 18490–18495. doi: 10.1073/pnas.0704501104.

Degnan, P. H., Michalowski, C. B., Babić, A. C., Cordes, M. H. J., and Little, J. W. (2007). Conservation and diversity in the immunity regions of wild phages with the immunity specificity of phage lambda. *Mol. Microbiol*. 64, 232–244. doi: 10.1111/j.1365-2958.2007.05650.x.

Degnan, P. H., Taga, M. E., and Goodman, A. L. (2014). Vitamin B12 as a modulator of gut microbial ecology. *Cell Metab*. 20, 769–778. doi: 10.1016/j.cmet.2014.10.002.

Delaney, S., Murphy, R., and Walsh, F. (2018). A Comparison of Methods for the Extraction of Plasmids Capable of Conferring Antibiotic Resistance in a Human Pathogen From Complex Broiler Cecal Samples. *Front. Microbiol*. 9, 1731. doi: 10.3389/fmicb.2018.01731.

Dempwolff, F., Sanchez, S., and Kearns, D. B. (2020). TnFLX: a Third-Generation mariner-Based Transposon System for Bacillus subtilis. *Appl. Environ. Microbiol*. 86. doi: 10.1128/AEM.02893-19.

Deng, H., Yang, S., Zhang, Y., Qian, K., Zhang, Z., Liu, Y., et al. (2018). Bacteroides fragilis Prevents Clostridium difficile Infection in a Mouse Model by Restoring Gut Barrier and Microbiome Regulation. *Front. Microbiol*. 9, 2976. doi: 10.3389/fmicb.2018.02976.

Duerkop, B. A., Clements, C. V., Rollins, D., Rodrigues, J. L. M., and Hooper, L. V. (2012). A composite bacteriophage alters colonization by an intestinal commensal bacterium. *Proceedings of the National Academy of Sciences* 109, 17621–17626. doi: 10.1073/pnas.1206136109.

Durrant, M. G., Li, M. M., Siranosian, B. A., Montgomery, S. B., and Bhatt, A. S. (2020). A Bioinformatic Analysis of Integrative Mobile Genetic Elements Highlights Their Role in Bacterial Adaptation. *Cell Host Microbe* 27, 140–153.e9. doi: 10.1016/j.chom.2019.10.022.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32, 1792–1797. doi: 10.1093/nar/gkh340.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res*. 47, D427–D432. doi: 10.1093/nar/gky995.

Enright, A. J., Van Dongen, S., and Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res*. 30, 1575–1584. doi: 10.1093/nar/30.7.1575.

Eren, A. M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S. E., Schechter, M. S., et al. (2021). Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* 6, 3–6. doi: 10.1038/s41564-020-00834-3.

Feng, Q., Liang, S., Jia, H., Stadlmayr, A., Tang, L., Lan, Z., et al. (2015). Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat. Commun*. 6, 6528. doi: 10.1038/ncomms7528.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., et al. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* 269, 496–512. doi: 10.1126/science.7542800.

Franke, A. E., and Clewell, D. B. (1981). Evidence for a chromosome-borne resistance transposon (Tn916) in Streptococcus faecalis that is capable of "conjugal" transfer in the absence of a conjugative plasmid. *J. Bacteriol*. 145, 494–502. doi: 10.1128/jb.145.1.494-502.1981.

Frye, K. A., Piamthai, V., Hsiao, A., and Degnan, P. H. (2021a). Mobilization of vitamin B12 transporters alters competitive dynamics in a human gut microbe. *Cell Rep*. 37, 110164. doi: 10.1016/j.celrep.2021.110164.

Frye, K. A., Piamthai, V., Hsiao, A., and Degnan, P. H. (2021b). Mobilization of vitamin B12 transporters alters competitive dynamics in a human gut microbe. *Cell Rep*. 37, 110164. doi: 10.1016/j.celrep.2021.110164.

García-Bayona Leonor, and Comstock Laurie E. (2019). Streamlined Genetic Manipulation of Diverse Bacteroides and Parabacteroides Isolates from the Human Gut Microbiota. *MBio* 10, e01762–19. doi: 10.1128/mBio.01762-19.

Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T., and Kado, C. I. (1985). Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J. Bacteriol*. 164, 918–921. doi: 10.1128/jb.164.2.918-921.1985.

Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., Mitra, R. D., Lozupone, C. A., et al. (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* 6, 279–289. doi: 10.1016/j.chom.2009.08.003.
Griffith, F. (1928). The Significance of Pneumococcal Types. *J. Hyg.* 27, 113–159. doi: 10.1017/s0022172400031879.

Guzman-Bautista, E. R., Suzuki, K., Asami, S., and Fagarasan, S. (2020). Bacteria-immune cells dialog and the homeostasis of the systems. *Curr. Opin. Immunol.* 66, 82–89. doi: 10.1016/j.coi.2020.05.010.

Hacker, J., Bender, L., Ott, M., Wingender, J., Lund, B., Marre, R., et al. (1990). Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal Escherichia coli isolates. *Microb. Pathog.* 8, 213–225. doi: 10.1016/0882-4010(90)90048-u.

Haft, D. H., Selengut, J. D., and White, O. (2003). The TIGRFAMs database of protein families. *Nucleic Acids Res.* 31, 371–373. doi: 10.1093/nar/gkg128.

Herold, S., Karch, H., and Schmidt, H. (2004). Shiga toxin-encoding bacteriophages--genomes in motion. *Int. J. Med. Microbiol.* 294, 115–121. doi: 10.1016/j.ijmm.2004.06.023.

Hopp, C. M., Gardner, J. F., and Salyers, A. A. (2015). The Xis2d protein of CTnDOT binds to the intergenic region between the mob and tra operons. *Plasmid* 81, 63–71. doi: 10.1016/j.plasmid.2015.07.002.

Hugon, P., Dufour, J.-C., Colson, P., Fournier, P.-E., Sallah, K., and Raoult, D. (2015). A comprehensive repertoire of prokaryotic species identified in human beings. *Lancet Infect. Dis.* 15, 1211–1219. doi: 10.1016/S1473-3099(15)00293-5.

Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. doi: 10.1038/nature11234.

Jain, R., Rivera, M. C., Moore, J. E., and Lake, J. A. (2003). Horizontal gene transfer accelerates genome innovation and evolution. *Mol. Biol. Evol.* 20, 1598–1602. doi: 10.1093/molbev/msg154.

Jeters, R. T., Wang, G.-R., Moon, K., Shoemaker, N. B., and Salyers, A. A. (2009). Tetracycline-associated transcriptional regulation of transfer genes of the Bacteroides conjugative transposon CTnDOT. *J. Bacteriol.* 191, 6374–6382. doi: 10.1128/JB.00739-09.

Jiang, X., Hall, A. B., Xavier, R. J., and Alm, E. J. (2019). Comprehensive analysis of chromosomal mobile genetic elements in the gut microbiome reveals phylum-level niche-adaptive gene pools. *PLoS One* 14, e0223680. doi: 10.1371/journal.pone.0223680.

Johansson, M. H. K., Bortolaia, V., Tansirichaiya, S., Aarestrup, F. M., Roberts, A. P., and Petersen, T. N. (2020). Detection of mobile genetic elements associated with antibiotic resistance in Salmonella enterica using a newly developed web tool: MobileElementFinder. *J. Antimicrob. Chemother.* 76, 101–109. doi: 10.1093/jac/dkaa390.

Johansson, M. H. K., Bortolaia, V., Tansirichaiya, S., Aarestrup, F. M., Roberts, A. P., and Petersen, T. N. (2021). Detection of mobile genetic elements associated with antibiotic resistance in Salmonella enterica using a newly developed web tool: MobileElementFinder. *J. Antimicrob. Chemother.* 76, 101–109. doi: 10.1093/jac/dkaa390.

Johnson, C. M., and Grossman, A. D. (2015). Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Annu. Rev. Genet.* 49, 577–601. doi: 10.1146/annurev-genet-112414-055018.

Jones, B. V., and Marchesi, J. R. (2006). Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. *Nat. Methods* 4, 55–61. doi: 10.1038/nmeth964.

Joshi, C. J., Ke, W., Drangowska-Way, A., O'Rourke, E. J., and Lewis, N. E. (2022). What are housekeeping genes? *PLoS Comput. Biol.* 18, e1010295. doi: 10.1371/journal.pcbi.1010295.

Ka, D., Oh, H., Park, E., Kim, J.-H., and Bae, E. (2020). Structural and functional evidence of bacterial antiphage protection by Thoeris defense system via NAD+ degradation. *Nat. Commun.* 11, 2816. doi: 10.1038/s41467-020-16703-w.

Keeton, C. M., Park, J., Wang, G.-R., Hopp, C. M., Shoemaker, N. B., Gardner, J. F., et al. (2013). The excision proteins of CTnDOT positively regulate the transfer operon. *Plasmid* 69, 172–179. doi: 10.1016/j.plasmid.2012.12.001.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053.

Kers, J. A., Cameron, K. D., Joshi, M. V., Bukhalid, R. A., Morello, J. E., Wach, M. J., et al. (2005). A large, mobile pathogenicity island confers plant pathogenicity on Streptomyces species. *Mol. Microbiol.* 55, 1025–1033. doi: 10.1111/j.1365-2958.2004.04461.x.

Khedkar, S., Smyshlyaev, G., Letunic, I., Maistrenko, O. M., Coelho, L. P., Orakov, A., et al. (2022). Landscape of mobile genetic elements and their antibiotic resistance cargo in prokaryotic genomes. *Nucleic Acids Res*. 50, 3155–3168. doi: 10.1093/nar/gkac163.
Kiljunen, S., Pajunen, M. I., and Savilahti, H. (2017). Transposon Insertion Mutagenesis for Archaeal Gene Discovery. *Methods Mol. Biol*. 1498, 309–320. doi: 10.1007/978-1-4939-6472-7_20.

Knapp, J. S., Johnson, S. R., Zenilman, J. M., Roberts, M. C., and Morse, S. A. (1988). High-level tetracycline resistance resulting from TetM in strains of Neisseria spp., Kingella denitrificans, and Eikenella corrodens. *Antimicrob. Agents Chemother*. 32, 765–767. doi: 10.1128/AAC.32.5.765.

Koropatkin, N. M., Martens, E. C., Gordon, J. I., and Smith, T. J. (2008). Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* 16, 1105–1115. doi: 10.1016/j.str.2008.03.017.

Lampe, D. J., Akerley, B. J., Rubin, E. J., Mekalanos, J. J., and Robertson, H. M. (1999). Hyperactive transposase mutants of the *Himar1* mariner transposon. *Proc. Natl. Acad. Sci. U. S. A*. 96, 11428–11433. doi: 10.1073/pnas.96.20.11428.

Lampe, D. J., Churchill, M. E., and Robertson, H. M. (1996a). A purified mariner transposase is sufficient to mediate transposition in vitro. *EMBO J*. 15, 5470–5479. doi: 10.1002/j.1460-2075.1996.tb00930.x.

Lampe, D. J., Churchill, M. E., and Robertson, H. M. (1996b). A purified mariner transposase is sufficient to mediate transposition in vitro. *EMBO J*. 15, 5470–5479. doi: 10.1002/j.1460-2075.1996.tb00930.x.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923.

Lederberg, J. (1952). Cell genetics and hereditary symbiosis. *Physiol. Rev*. 32, 403–430. doi: 10.1152/physrev.1952.32.4.403.

Lederberg, J., and Tatum, E. L. (1946). Gene recombination in Escherichia coli. *Nature* 158, 558. doi: 10.1038/158558a0.

Leigh, J. (2021). *IGBIllinois/VICSIN: v1.0.1*. doi: 10.5281/zenodo.5138177.

Li, L. Y., Shoemaker, N. B., and Salyers, A. A. (1995). Location and characteristics of the transfer region of a Bacteroides conjugative transposon and regulation of transfer genes. *J. Bacteriol*. 177, 4992–4999. doi: 10.1128/jb.177.17.4992-4999.1995.

Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., et al. (2019). ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res*. 47, D660–D665. doi: 10.1093/nar/gky1123.

Lorenz, M. G., and Wackernagel, W. (1994). Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Rev*. 58, 563–602. doi: 10.1128/mr.58.3.563-602.1994.

Martens, E. C., Chiang, H. C., and Gordon, J. I. (2008). Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* 4, 447–457. doi: 10.1016/j.chom.2008.09.007.

Mazurkiewicz, P., Tang, C. M., Boone, C., and Holden, D. W. (2006). Signature-tagged mutagenesis: barcoding mutants for genome-wide screens. *Nat. Rev. Genet*. 7, 929–939. doi: 10.1038/nrg1984.

McClintock, B. (1931). The Order of the Genes C, Sh and Wx in Zea Mays with Reference to a Cytologically Known Point in the Chromosome. *Proc. Natl. Acad. Sci. U. S. A*. 17, 485–491. doi: 10.1073/pnas.17.8.485.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences* 36, 344–355. doi: 10.1073/pnas.36.6.344.

McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C. A., et al. (2013). Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res*. 41, e140. doi: 10.1093/nar/gkt444.

Messerer, M., Fischer, W., and Schubert, S. (2017). Investigation of horizontal gene transfer of pathogenicity islands in Escherichia coli using next-generation sequencing. *PLoS One* 12, e0179880. doi: 10.1371/journal.pone.0179880.

Mikkelsen, H., Hui, K., Barraud, N., and Filloux, A. (2013). The pathogenicity island encoded PvrSR/RcsCB regulatory network controls biofilm formation and dispersal in Pseudomonas aeruginosa PA14. *Mol. Microbiol*. 89, 450–463. doi: 10.1111/mmi.12287.

Mikolčević, P., Hloušek-Kasun, A., Ahel, I., and Mikoč, A. (2021). ADP-ribosylation systems in bacteria and viruses. *Comput. Struct. Biotechnol. J*. 19, 2366–2383. doi: 10.1016/j.csbj.2021.04.023.

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. in *2010 Gateway Computing Environments Workshop (GCE)* (IEEE), 1–8. doi: 10.1109/GCE.2010.5676129.

Moon, K., Sonnenburg, J., and Salyers, A. A. (2007). Unexpected effect of a Bacteroides conjugative transposon, CTnDOT, on chromosomal gene expression in its bacterial host. *Mol. Microbiol.* 64, 1562–1571. doi: 10.1111/j.1365-2958.2007.05756.x.

Mullish, B. H., McDonald, J. A. K., Pechlivanis, A., Allegretti, J. R., Kao, D., Barker, G. F., et al. (2019). Microbial bile salt hydrolases mediate the efficacy of faecal microbiota transplant in the treatment of recurrent Clostridioides difficile infection. *Gut* 68, 1791–1800. doi: 10.1136/gutjnl-2018-317842.

Murray, N. E. (2000). Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.* 64, 412–434. doi: 10.1128/MMBR.64.2.412-434.2000.

Nayfach, S., Shi, Z. J., Seshadri, R., Pollard, K. S., and Kyrpides, N. C. (2019). New insights from uncultivated genomes of the global human gut microbiome. *Nature* 568, 505–510. doi: 10.1038/s41586-019-1058-x.

Niehus, R., Mitri, S., Fletcher, A. G., and Foster, K. R. (2015). Migration and horizontal gene transfer divide microbial genomes into multiple niches. *Nat. Commun.* 6, 8924. doi: 10.1038/ncomms9924.

Ochman, H., and Selander, R. K. (1984). Standard reference strains of Escherichia coli from natural populations. *J. Bacteriol.* 157, 690–693. doi: 10.1128/jb.157.2.690-693.1984.

Ochman, H., Worobey, M., Kuo, C.-H., Ndjango, J.-B. N., Peeters, M., Hahn, B. H., et al. (2010). Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biol.* 8, e1000546. doi: 10.1371/journal.pbio.1000546.

Ortañez, J., and Degnan, P.H. (2024). Tracking and Characterization of a Novel Conjugative Transposon Identified by Shotgun Transposon Mutagenesis. *Front. Microbiol.* 15. doi: 10.3389/fmicb.2024.1241582

Osborn, A. M., and Böltner, D. (2002). When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. *Plasmid* 48, 202–212. doi: 10.1016/s0147-619x(02)00117-8.

Ostblom, A., Adlerberth, I., Wold, A. E., and Nowrouzian, F. L. (2011). Pathogenicity island markers, virulence determinants malX and usp, and the capacity of Escherichia coli to persist in infants' commensal microbiotas. *Appl. Environ. Microbiol.* 77, 2303–2308. doi: 10.1128/AEM.02405-10.

Ozer, E. A., Allen, J. P., and Hauser, A. R. (2014). Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools Spine and AGEnt. *BMC Genomics* 15, 737. doi: 10.1186/1471-2164-15-737.

Panwar, S., Kumari, S., Verma, J., Bakshi, S., Narendrakumar, L., Paul, D., et al. (n.d.). Toxin-linked mobile genetic elements in major enteric bacterial pathogens. doi: 10.1017/gmb.2023.2.

Partridge Sally R., Kwong Stephen M., Firth Neville, and Jensen Slade O. (2018). Mobile Genetic Elements Associated with Antimicrobial Resistance. *Clin. Microbiol. Rev.* 31, e00088–17. doi: 10.1128/CMR.00088-17.

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., et al. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* 176, 649–662.e20. doi: 10.1016/j.cell.2019.01.001.

Poyart-Salmeron, C., Trieu-Cuot, P., Carlier, C., and Courvalin, P. (1990). The integration-excision system of the conjugative transposon Tn 1545 is structurally and functionally related to those of lambdoid phages. *Mol. Microbiol.* 4, 1513–1521. doi: 10.1111/j.1365-2958.1990.tb02062.x.

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 5, e9490. doi: 10.1371/journal.pone.0009490.

Price-Carter, M., Tingey, J., Bobik, T. A., and Roth, J. R. (2001). The alternative electron acceptor tetrathionate supports B12-dependent anaerobic growth of Salmonella enterica serovar typhimurium on ethanolamine or 1,2-propanediol. *J. Bacteriol.* 183, 2463–2475. doi: 10.1128/JB.183.8.2463-2475.2001.

Ramsey, M. E., Woodhams, K. L., and Dillard, J. P. (2011). The Gonococcal Genetic Island and Type IV Secretion in the Pathogenic Neisseria. *Front. Microbiol.* 2, 61. doi: 10.3389/fmicb.2011.00061.

Rankin, D. J., Rocha, E. P. C., and Brown, S. P. (2011). What traits are carried on mobile genetic elements, and why? *Heredity* 106, 1–10. doi: 10.1038/hdy.2010.24.

Rendueles, O., de Sousa, J. A. M., Bernheim, A., Touchon, M., and Rocha, E. P. C. (2018). Genetic exchanges are more frequent in bacteria encoding capsules. *PLoS Genet.* 14, e1007862. doi: 10.1371/journal.pgen.1007862.

Reyes, A., Wu, M., McNulty, N. P., Rohwer, F. L., and Gordon, J. I. (2013). Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20236–20241. doi: 10.1073/pnas.1319470110.

Rice, L. B., and Carias, L. L. (1994). Studies on excision of conjugative transposons in enterococci: evidence for joint sequences composed of strands with unequal numbers of nucleotides. *Plasmid* 31, 312–316. doi: 10.1006/plas.1994.1034.

Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, G. A. D., Gasbarrini, A., et al. (2019). What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms* 7. doi: 10.3390/microorganisms7010014.

Riva, A., Borgo, F., Lassandro, C., Verduci, E., Morace, G., Borghi, E., et al. (2017). Pediatric obesity is associated with an altered gut microbiota and discordant shifts in Firmicutes populations. *Environ. Microbiol.* 19, 95–105. doi: 10.1111/1462-2920.13463.

Roux, S., Enault, F., Hurwitz, B. L., and Sullivan, M. B. (2015). VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3, e985. doi: 10.7717/peerj.985.

Saak, C. C., Dinh, C. B., and Dutton, R. J. (2020). Experimental approaches to tracking mobile genetic elements in microbial communities. *FEMS Microbiol. Rev.* 44, 606–630. doi: 10.1093/femsre/fuaa025.

Salyers, A. A., Shoemaker, N. B., Stevens, A. M., and Li, L. Y. (1995a). Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol. Rev.* 59, 579–590. doi: 10.1128/mr.59.4.579-590.1995.

Salyers, A. A., Shoemaker, N. B., Stevens, A. M., and Li, L. Y. (1995b). Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol. Rev.* 59, 579–590. Available at: https://www.ncbi.nlm.nih.gov/pubmed/8531886.

Schlesinger, D. J., Shoemaker, N. B., and Salyers, A. A. (2007). Possible origins of CTnBST, a conjugative transposon found recently in a human colonic Bacteroides strain. *Appl. Environ. Microbiol.* 73, 4226–4233. doi: 10.1128/AEM.00455-07.

Schmidt, H., Bielaszewska, M., and Karch, H. (1999). Transduction of enteric Escherichia coli isolates with a derivative of Shiga toxin 2-encoding bacteriophage phi3538 isolated from Escherichia coli O157:H7. *Appl. Environ. Microbiol.* 65, 3855–3861. doi: 10.1128/AEM.65.9.3855-3861.1999.

Schröder, G., and Lanka, E. (2005). The mating pair formation system of conjugative plasmids-A versatile secretion machinery for transfer of proteins and DNA. *Plasmid* 54, 1–25. doi: 10.1016/j.plasmid.2005.02.001.

Schubert, S., Rakin, A., and Heesemann, J. (2004). The Yersinia high-pathogenicity island (HPI): evolutionary and functional aspects. *Int. J. Med. Microbiol.* 294, 83–94. doi: 10.1016/j.ijmm.2004.06.026.

Sentchilo, V., Ravatn, R., Werlen, C., Zehnder, A. J. B., and van der Meer, J. R. (2003). Unusual integrase gene expression on the clc genomic island in Pseudomonas sp. strain B13. *J. Bacteriol.* 185, 4530–4538. doi: 10.1128/JB.185.15.4530-4538.2003.
Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303.

Shoemaker, N. B., Vlamakis, H., Hayes, K., and Salyers, A. A. (2001). Evidence for extensive resistance gene transfer among Bacteroides spp. and among Bacteroides and other genera in the human colon. *Appl. Environ. Microbiol.* 67, 561–568. doi: 10.1128/AEM.67.2.561-568.2001.

Shoemaker N. B., Vlamakis H., Hayes K., and Salyers A. A. (2001). Evidence for Extensive Resistance Gene Transfer amongBacteroides spp. and among Bacteroides and Other Genera in the Human Colon. *Appl. Environ. Microbiol.* 67, 561–568. doi: 10.1128/AEM.67.2.561-568.2001.

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–6. doi: 10.1093/nar/gkj014.

Simon, N. C., Aktories, K., and Barbieri, J. T. (2014). Novel bacterial ADP-ribosylating toxins: structure and function. *Nat. Rev. Microbiol.* 12, 599–611. doi: 10.1038/nrmicro3310.

Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., and Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480, 241–244. doi: 10.1038/nature10571.

Smillie Chris, Garcillán-Barcia M. Pilar, Francia M. Victoria, Rocha Eduardo P. C., and de la Cruz Fernando (2010). Mobility of Plasmids. *Microbiol. Mol. Biol. Rev.* 74, 434–452. doi: 10.1128/MMBR.00020-10.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033.

Stingl, K., Müller, S., Scheidgen-Kleyboldt, G., Clausen, M., and Maier, B. (2010). Composite system mediates two-step DNA uptake into *Helicobacter pylori*. *Proc. Natl. Acad. Sci. U. S. A.* 107, 1184–1189. doi: 10.1073/pnas.0909955107.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 102, 15545–15550. doi: 10.1073/pnas.0506580102.

Takeda, Y., Folkmanis, A., and Echols, H. (1977). Cro regulatory protein specified by bacteriophage lambda. Structure, DNA-binding, and repression of RNA synthesis. *J. Biol. Chem.* 252, 6177–6183. doi: 10.1016/S0021-9258(17)40047-0.

Tansirichaiya Supathep, Goodman Richard N., Guo Xinyu, Bulgasim Issra, Samuelsen Ørjan, Al-Haroni Mohammed, et al. (2022). Intracellular Transposition and Capture of Mobile Genetic Elements following Intercellular Conjugation of Multidrug Resistance Conjugative Plasmids from Clinical Enterobacteriaceae Isolates. *Microbiology Spectrum* 10, e02140–21. doi: 10.1128/spectrum.02140-21.

Tantoso, E., Eisenhaber, B., Kirsch, M., Shitov, V., Zhao, Z., and Eisenhaber, F. (2022). To kill or to be killed: pangenome analysis of Escherichia coli strains reveals a tailocin specific for pandemic ST131. *BMC Biol.* 20, 146. doi: 10.1186/s12915-022-01347-7.

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome." *Proc. Natl. Acad. Sci. U. S. A.* 102, 13950–13955. doi: 10.1073/pnas.0506758102.

Turner, N. A., Sharma-Kuinkel, B. K., Maskarinec, S. A., Eichenberger, E. M., Shah, P. P., Carugati, M., et al. (2019). Methicillin-resistant Staphylococcus aureus: an overview of basic and clinical research. *Nat. Rev. Microbiol.* 17, 203–218. doi: 10.1038/s41579-018-0147-4.

Usyk, M., Pandey, A., Hayes, R. B., Moran, U., Pavlick, A., Osman, I., et al. (2021). Bacteroides vulgatus and Bacteroides dorei predict immune-related adverse events in immune checkpoint blockade treatment of metastatic melanoma. *Genome Med.* 13, 160. doi: 10.1186/s13073-021-00974-z.

van Opijnen, T., Bodi, K. L., and Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods* 6, 767–772. doi: 10.1038/nmeth.1377.

van Schaik, W. (2015). The human gut resistome. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20140087. doi: 10.1098/rstb.2014.0087.

Vineis, J. H., Reznikoff, W. S., Antonopoulos, D. A., Koval, J., Chang, E., Fallon, B. R., et al. (2024). A novel conjugative transposon carrying an autonomously amplified plasmid. *MBio* 15, e0278723. doi: 10.1128/mbio.02787-23.

Vogan, A. A., and Higgs, P. G. (2011). The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. *Biol. Direct* 6, 1. doi: 10.1186/1745-6150-6-1.

Vogtmann, E., Hua, X., Zeller, G., Sunagawa, S., Voigt, A. Y., Hercog, R., et al. (2016). Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing. *PLoS One* 11, e0155362. doi: 10.1371/journal.pone.0155362.

Wallace, M. J., Jean, S., Wallace, M. A., Burnham, C.-A. D., and Dantas, G. (2022). Comparative Genomics of Bacteroides fragilis Group Isolates Reveals Species-Dependent Resistance Mechanisms and Validates Clinical Tools for Resistance Prediction. *MBio* 13, e0360321. doi: 10.1128/mbio.03603-21.

Wallden, K., Rivera-Calzada, A., and Waksman, G. (2010). Type IV secretion systems: versatility and diversity in function. *Cell. Microbiol.* 12, 1203–1212. doi: 10.1111/j.1462-5822.2010.01499.x.

Wang, Y., Rotman, E. R., Shoemaker, N. B., and Salyers, A. A. (2005). Translational control of tetracycline resistance and conjugation in the Bacteroides conjugative transposon CTnDOT. *J. Bacteriol.* 187, 2673–2680. doi: 10.1128/JB.187.8.2673-2680.2005.

Wang, Y., Shoemaker, N. B., and Salyers, A. A. (2004). Regulation of a Bacteroides operon that controls excision and transfer of the conjugative transposon CTnDOT. *J. Bacteriol.* 186, 2548–2557. doi: 10.1128/JB.186.9.2548-2557.2004.

Waters, J. L., and Salyers, A. A. (2012). The small RNA RteR inhibits transfer of the Bacteroides conjugative transposon CTnDOT. *J. Bacteriol.* 194, 5228–5236. doi: 10.1128/JB.00941-12.

Waters, J. L., and Salyers, A. A. (2013). Regulation of CTnDOT conjugative transfer is a complex and highly coordinated series of events. *MBio* 4, e00569–13. doi: 10.1128/mBio.00569-13.

Waters, V. L. (2001). Conjugation between bacterial and mammalian cells. *Nat. Genet.* 29, 375–376. doi: 10.1038/ng779.

Welch, R. A., Burland, V., Plunkett, G., 3rd, Redford, P., Roesch, P., Rasko, D., et al. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* 99, 17020–17024. doi: 10.1073/pnas.252529799.

Wensel, C. R., Pluznick, J. L., Salzberg, S. L., and Sears, C. L. (2022). Next-generation sequencing: insights to advance clinical investigations of the microbiome. *J. Clin. Invest.* 132. doi: 10.1172/JCI154944.

Wexler, H. M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* 20, 593–621. doi: 10.1128/CMR.00008-07.

Whittle, G., Hamburger, N., Shoemaker, N. B., and Salyers, A. A. (2006). A bacteroides conjugative transposon, CTnERL, can transfer a portion of itself by conjugation without excising from the chromosome. *J. Bacteriol.* 188, 1169–1174. doi: 10.1128/JB.188.3.1169-1174.2006.

Whittle, G., Hund, B. D., Shoemaker, N. B., and Salyers, A. A. (2001). Characterization of the 13-kilobase ermF region of the Bacteroides conjugative transposon CTnDOT. *Appl. Environ. Microbiol.* 67, 3488–3495. doi: 10.1128/AEM.67.8.3488-3495.2001.

Whittle, G., Shoemaker, N. B., and Salyers, A. A. (2002a). Characterization of genes involved in modulation of conjugal transfer of the Bacteroides conjugative transposon CTnDOT. *J. Bacteriol.* 184, 3839–3847. doi: 10.1128/JB.184.14.3839-3847.2002.

Whittle, G., Shoemaker, N. B., and Salyers, A. A. (2002b). The role of Bacteroides conjugative transposons in the dissemination of antibiotic resistance genes. *Cell. Mol. Life Sci.* 59, 2044–2054. doi: 10.1007/s000180200004.

WHO (2023). Antimicrobial resistance. Available at: https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance [Accessed December 21, 2023].

Wirbel, J., Pyl, P. T., Kartal, E., Zych, K., Kashani, A., Milanese, A., et al. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* 25, 679–689. doi: 10.1038/s41591-019-0406-6.

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 257. doi: 10.1186/s13059-019-1891-0.

Wood, M. M., and Gardner, J. F. (2015). The Integration and Excision of CTnDOT. *Microbiol Spectr* 3, MDNA3–0020–2014. doi: 10.1128/microbiolspec.MDNA3-0020-2014.

Wozniak, R. A. F., and Waldor, M. K. (2010). Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.* 8, 552–563. doi: 10.1038/nrmicro2382.

Wright, L. D., and Grossman, A. D. (2016). Autonomous Replication of the Conjugative Transposon Tn916. *J. Bacteriol.* 198, 3355–3366. doi: 10.1128/JB.00639-16.

Xie, F., Jin, W., Si, H., Yuan, Y., Tao, Y., Liu, J., et al. (2021). An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants. *Microbiome* 9, 137. doi: 10.1186/s40168-021-01078-x.

Yoshida, N., Emoto, T., Yamashita, T., Watanabe, H., Hayashi, T., Tabata, T., et al. (2018). Bacteroides vulgatus and Bacteroides dorei Reduce Gut Microbial Lipopolysaccharide Production and Inhibit Atherosclerosis. *Circulation* 138, 2486–2498. doi: 10.1161/CIRCULATIONAHA.118.033714.

Yu, J., Feng, Q., Wong, S. H., Zhang, D., Liang, Q. Y., Qin, Y., et al. (2017). Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* 66, 70–78. doi: 10.1136/gutjnl-2015-309800.

Zhang, H., Sparks, J. B., Karyala, S. V., Settlage, R., and Luo, X. M. (2015). Host adaptive immunity alters gut microbiota. *ISME J*. 9, 770–781. doi: 10.1038/ismej.2014.165.

Zhou, Z., Charlesworth, J., and Achtman, M. (2020). Accurate reconstruction of bacterial pan- and core genomes with PEPPAN. *Genome Res*. 30, 1667–1679. doi: 10.1101/gr.260828.120.