# A family of Integrative and Conjugative Elements (ICEs) involved in carbohydrate metabolism

**Amelia G. Kelly**

**Thesis Advisor: Dr. Eric C. Martens**

**Honors Thesis**

**Program in Biology**

**University of Michigan**

**Winter 2013**

## I. Introduction

The Human Microbiota and Polysaccharide Utilization

The human gastrointestinal (GI) tract is home to trillions of diverse microorganisms, collectively known as the gut microbiota. Colonization of the GI tract occurs immediately after birth, however, the composition of microbial species within the gut shifts throughout one's lifetime (2). This dense, largely anaerobic, microbial community serves a variety of functions and provides many benefits to the host. These benefits include developing the immune system (4), producing essential vitamins for the host (5), preventing the colonization of pathogenic species (6), and aiding in the degradation of the complex carbohydrates consumed by the host (7). The composition of species within the GI tract is essential in maintaining this mutualistic relationship. Furthermore, unhealthy shifts in gut microbial composition, known as dysbiosis, have been postulated to lead to many harmful disease states such as inflammatory bowel disease (IBD) (8), colon cancer (9), obesity (10) and diabetes (11). Understanding the factors that alter the physiology or metabolic potential of this microbial community will be essential in developing methods to treat GI diseases associated with its behavior.

One prominent factor influencing the composition of the microbial community in the gut is the introduction of complex carbohydrates (glycans). The gut microbiota can degrade a diverse range of both dietary and host-mucosal glycans. While humans produce enzymes allowing them to degrade only lactose, sucrose and starch, the bacteria in the gut are capable of utilizing and degrading dozens, even hundreds, of complex glycans that humans are unable to break down themselves. Despite our inability to degrade many glycans, humans consume numerous indigestible polysaccharides that permit our gut bacteria to thrive. These polysaccharides, which come directly from the plants and animals we consume, are fermented into short-chain fatty

acids (SCFAs) such as acetate, butyrate, and propionate that can be absorbed by the host and used for energy. This process contributes up to 10% of our daily caloric intake, depending on diet. Both acetate and propionate are utilized in lipid and glucose metabolism (12), while butyrate provides nutrients directly to gut epithelial cells, such as colonocytes (13).

In addition to degrading a diverse range of dietary glycans, the gut microbiota can target host mucosal glycans and epithelial cells shed by the constantly-renewing mucosal layer in the GI tract. This potentially places the gut bacteria very close to host cells, which could significantly alter colonic health. The host endogenous polysaccharides that are available to the gut bacteria are either N-linked or O-linked glycans (the latter being dominant in secreted mucus) and provide a constant source of nutrients for the microbial community in the GI tract.

The bacteria inhabiting the gut vary in the number of polysaccharides that they can degrade. Furthermore, the wide array of polysaccharides available to the gut microbiota allows numerous bacterial species within the GI tract to fulfill a variety of niches. Some species are "generalists" that can target dozens of carbohydrates, while others are "specialists" that have evolved to degrade only a few substrates and fill very specific roles in the GI tract. Because of their ability to target a diverse range of glycans, "generalists" are able to adapt to changes or shifts in the polysaccharide composition of the gut. For example, *Bacteroides thetaiotamicron VPI-5482* (*B. theta*), a prominent gut bacteria, is a "generalist" that is able to survive in the host due to its ability to utilize a diverse range of both host and dietary glycans (14). On the other hand, "specialists" rely on a constant source of specific glycans and may not be able to adapt to changes in the polysaccharide composition of the gut. Little is known about how our gut microbiota have evolved these diverse repertoires of glycan utilization abilities.

Polysaccharide Utilization in the Bacteroidetes

The dominant Gram-negative phylum within the GI tract is the Bacteroidetes. The Bacteroidetes play key roles in degrading a diverse range of polysaccharides, many of which are important and abundant components of our diets. The Bacteroidetes employ a similar paradigm for degrading virtually all of the different glycans that they utilize as nutrient sources. This paradigm was initially characterized in *B. theta* and involves a single eight-gene locus to degrade starch (15) (16). This locus is known as the starch utilization system (Sus). The Sus system involves outer membrane (OM) and periplasmic glycan-degrading enzymes. These degradative enzymes work in concert with glycan-binding and transport proteins to systematically bind, degrade and import pieces of target glycans with molecular specificity (FIG. 1 from Koropatkin et al. 2012 (3)).

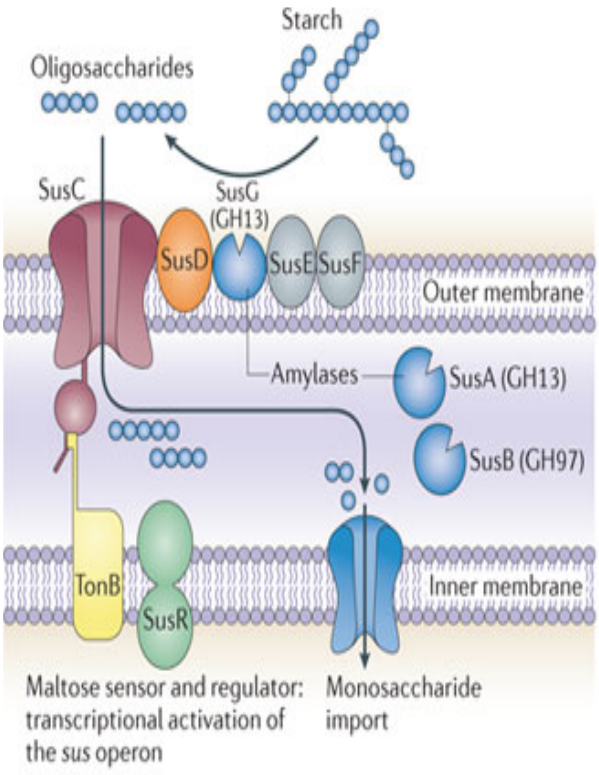


**Fig.1.** This illustration from Koropatikin et al. (2012) shows the current model of the *B. theta* starch utilization system (Sus). The SusC protein is a TonB-dependent transporter that transports oligosaccharides into the cell. SusC works with the starch-binding lipoproteins SusD, SusE, SusF, and SusG. SusD, SusE and SusF prompt starch binding, while SusG is important for the preliminary degradation of starch. SusG is an α-amylase in the glycoside hydrolase family 13 (GH13). After the initial degradation of starch by SusG, oligosaccharides are transported through SusC into the periplasm. In the periplasm, malto-oligosaccharides are further degraded by SusA and SusB to maltose and other sugars. SusA is another GH13 enzyme, while SusB is an α-glucosidase in the GH97 family. SusR, a protein spanning the inner-membrane, senses maltose and activates other Sus proteins. Simple sugars are then transported through the inner membrane by an undefined permease protein (3).

The sequence of the *B. theta* genome revealed that there are 87 additional gene clusters that are related to the *sus* system (17) (18). All of these gene clusters contain homologues of at least two of the genes (*susC* and *susD*) that are in the *sus* locus. These gene clusters are termed polysaccharide utilization loci (PULs) and their protein products are known as "Sus-like systems".

Genomic analyses suggest that individual gut Bacteroidetes, of which there are several dozen named species, dedicate as much as 20% of their open reading frames (ORFs) to coding for Sus-like systems (FIG. 2 from Martens et al. 2011 (1)). Moreover, transcriptional analyses suggest that many of the PULs that encode these Sus-like systems are composed of immediately adjacent genes and operons that are regulated in response to oligosaccharide cues.


**Fig. 2.** This illustration from Martens et al. (2011) shows the abundance of PULs in the sequenced genomes of the type strains of *B. theta* and *B. ovatus*. The two taxa dedicate a large portion of their genomes to glycan degradation. The dark green regions indicate homologous PULs and the light green regions are probably homologous PULs. Definite homologous PULs are labeled 1 through 28. The gold gene clusters are PULs that are specific to *B. theta* while the blue and pink gene clusters are specific to *B. ovatus* (1).

Studies on just a few Bacteroidetes from the gut microbiota have shown that there are Sus-like systems encoding PULs that degrade most of the glycans that would be expected to enter the human gut except for cellulose, including rare glycans such as porphyran, a polysaccharide present in red algae that has been consumed for several centuries in Asian cultures (FIG. 2 from Martens et al. 2011 (1)) (17) (19) (20). *B. theta* alone has nearly 300 glycoside hydrolases and polysaccharide lyases in its genome that aid in its ability to degrade a diverse range of carbohydrates within the GI tract (21). Additionally, several *sus*-like systems have been discovered that target O-linked glycans found in the host mucosa. Despite this deepening knowledge of Bacteroidetes glycan degradation strategies, the mechanisms of PUL evolution within and between genomes have not yet been investigated. One common mechanism of genome evolution employed by the Bacteroidetes to exchange genes, such as those involved in antibiotic resistance, is horizontal gene transfer via conjugative mechanisms.

## Horizontal Gene Transfer

Individual Bacteroidetes could have gained their glycan degrading abilities via horizontal gene transfer (HGT) mechanisms. The acquisition of glycan utilization abilities via HGT would be advantageous for the gut microbiota given the abundance of polysaccharides present in the GI tract. When HGT results in acquired genes that are advantageous for survivorship, these genes are likely to spread within a population. There are many forms of HGT that have been studied, including natural transformation, phage transduction, and conjugative transfer.

Natural transformation refers to the uptake, integration and expression of extracellular DNA, however, only ~1% of bacterial species are known to be naturally transformable (22). Natural transformation (competence) requires the expression of roughly 20 to 50 proteins and

refers to the bacterial ability to acquire extracellular DNA (23). Furthermore, competence only occurs under specific environmental conditions that involve a change in cell density by quorum sensing, altered growth conditions and nutrient access, or starvation. In order for successful HGT to occur via natural transformation, there must be competent bacterial cells, a continuous release of extracellular DNA, and the ability for the extracellular DNA to be integrated into the host genome.

Another known mechanism of HGT is via phage transduction. Transduction refers to packaging of bacterial DNA into a viral envelope, or phage, that can 'infect' another bacterial cell (24). Upon 'infection', the bacterial cell can integrate the novel DNA into its host chromosome. One type of transduction is generalized transduction, where any bacterial gene can be incorporated into the viral envelope. Generalized transduction is random and the amount of genetic material that is transferred depends on the size of the phage being utilized. Transduction can also be specialized where only a specific set of genes can be transferred by the phage.

The final known method of HGT, and the one that is central to this study, is conjugative transfer, which occurs via either conjugative plasmids or Integrative and Conjugative Elements (ICEs), also referred to as conjugative transposons (cTns). Conjugative plasmids are circular pieces of extra-chromosomal DNA that replicate independently from the genome. Plasmids can increase the fitness of microbes, especially if the plasmids carry a specific function that is known to augment the likelihood of survivorship. For example, it is known that in the presence of antibiotics, plasmids that harbor antibiotic resistance genes are much more likely to improve the fitness of bacteria. In spite of this, some plasmids have been shown to reduce the fitness of microbes in the absence of selection for a specific function (25).

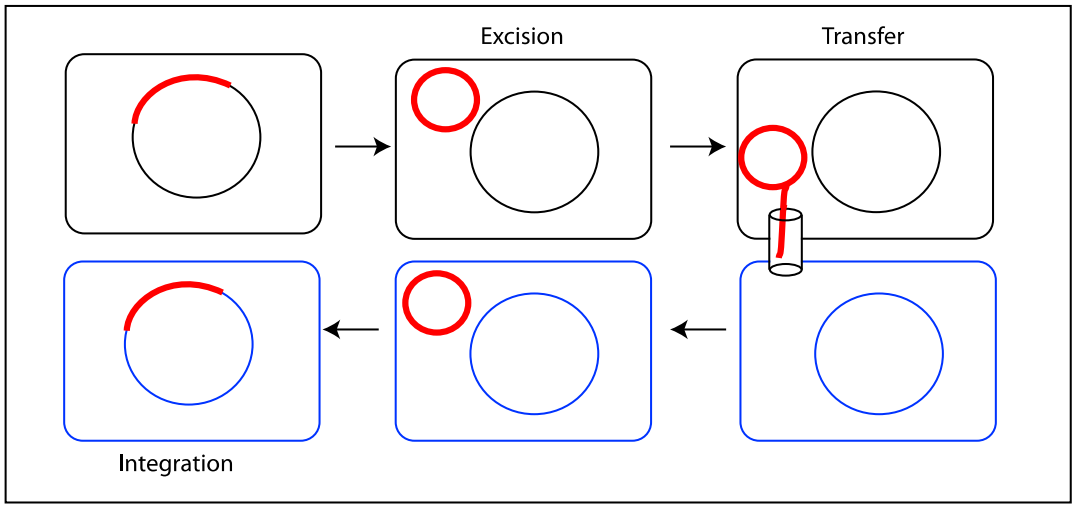


**Fig 3.** The ICE, shown is red, is normally integrated into the host chromosome. However, under certain conditions, the ICE will excise from the host chromosome, shown in black, and circularize. This circular element undergoes "rolling circle replication" in which single stranded DNA copies are generated. The replicated element is transferred through a mating pore known as a pilus. The recipient cell, shown in blue, can then integrate the ICE into its chromosome and retain the element.

Unlike plasmids, ICEs are a group of mobile genetic elements that integrate into and replicate with the host chromosome (25). ICEs are abundant in both Gram-negative and Gram-positive bacteria and they require the ability to integrate, excise, and mobilize for transfer to occur, but mechanisms for these functions vary (FIG. 3). Due to the nature of their insertion sites, the elements will frequently, although not always, insert at a specific target sequence(s) within the genome (26). Genomic insertion typically creates flanking direct repeats. These repeats could be as short as five base pairs and they delineate the entire ICE. Furthermore, many ICEs in Gram-negative bacteria contain type IV secretion systems that form a pilus or other conduit for conjugative DNA transfer (23). This structure is essential in generating a junction between the bacteria in order to successfully transfer these conjugative elements and the gene cassettes that they harbor.

Thesis Plan: The Evolution of Polysaccharide Utilization within the Bacteroidetes

The glycan degradation strategies of the Bacteroidetes phylum are being intensely investigated, however, little is known about the mechanisms of PUL evolution within and between genomes. Determining how members of the gut microbiota gained their abilities to degrade many of the carbohydrates that we consume is essential in understanding the plasticity of the microbiome. To investigate PUL evolution, I began by exploring the genetic basis of two rare phenotypes, fungal cell wall α-mannan degradation and algal porphyran degradation. I reasoned that rare polysaccharide utilization traits would be those that are most likely to have been recently acquired. Each of these glycans is present in the human diet, but consumption is either restricted to certain populations or historically recent (*i.e.* within the last few centuries or millennia). Using comparative genomic analysis between related strains that vary in these phenotypes, we discovered that the PULs encoding each behavior are contained on homologous ICEs that are closely related to Bacteroidetes cTns involved in transfer of antibiotic resistance. After examining the genetic makeup of two ICEs harboring PULs, this study elucidated the diversity of PULs contained on ICEs by building a phylogenetic tree based off of conserved sequences that are necessary for the conjugative elements to function.

The phylogeny obtained is based off of close to one hundred Bacteroidetes genomes. Many of the genomes utilized have only been recently isolated and sequenced, ensuring that new genetic diversity was accounted for. As numerous human and animal gut Bacteroidetes genomes continue to be sequenced, there is a growing need for a genetic analysis of important PULs within the genomes of the gut microbiota. The current availability of hundreds of Bacteroidetes genomes, along with connections between their genetic sequences and phenotypes of a few

model strains, provides a unique opportunity to investigate the mechanism of evolution of glycan metabolic traits in the gut bacteria.

## II. The evolution of polysaccharide utilization within the Bacteroidetes: A family of Integrative and Conjugative Elements (ICEs) involved in carbohydrate metabolism

*Portions of this chapter were published in Hehemann et al. in 2012 (27).

Abstract

Bacteria inhabiting the human gastrointestinal (GI) tract appear to have gained glycan utilization abilities via Horizontal Gene Transfer (HGT). Here, two polysaccharide utilization loci (PULs), enabling α-mannan utilization by *B. thetaiotaomicron VPI-3731* (*B. theta*) and porphyran utilization by *B. plebeius DSM 17135*, were identified that were contained on homologous Integrative and Conjugative Elements (ICEs) that are likely products of HGT. In this study, we showed that *B. plebeius* expresses its porphyran PUL in the presence of porphyran as its sole carbon source and that both the *B. plebeius* and *B. theta* ICEs are capable of excision, although at a low frequency.

In order to discover the connections between Bacteroidetes ICEs and the evolution of new polysaccharide metabolism traits, a Maximum Parsimony (MP) phylogenetic tree was constructed based off of the protein sequences of a conserved gene that is required for the ICEs to function. Interestingly, all elements on the tree were from Bacteroidetes isolates and they tended to group by the isolation source of the bacterial genome in which they are located. Examining the branch of the tree with the *B. theta* and *B. plebeius* elements revealed that closely related elements tend to use the same insertion site to enter into their respective host chromosomes. Furthermore, an analysis of the genetic cargo of these elements elucidated more ICEs that carry PULs, as well as many other elements that harbor a diverse range of cargo genes,

including capsule synthesis genes and genetic clusters conferring putative antibiotic resistance. Therefore, ICEs appear to be one important mechanism by which PULs and other important gene clusters have entered the genomes of the gut bacteria. As more sequencing data continues to become available, it will be imperative to study more PULs on ICEs to gain a more complete understanding of how the bacteria inhabiting the GI tract acquire their polysaccharide utilization traits and the extent of the adaptability of the human microbiome.

## Introduction

In 2010 it was discovered by Hehemann et al. that *B. plebeius*, a gut symbiont isolated from a healthy Japanese person, encodes an enzyme that in pure form can degrade porphyran, a red algal carbohydrate found in nori (20). Nori is commonly consumed in Japanese culture as a wrapping for sushi. Porphyran from *Porphyra* spp. is a member of the agar-family of polysaccharides. These polysaccharides are all polymers of galactose and galactose derivitives. Porphyran differs from agarose in that the majority of the 3,6-anhydro-L-galactose monomers found in agarose are replaced by L-galactose-6-sulfate (L6S) monomers in porphyran. Furthermore, C-6-methylations of the D-galactose are common in porphyran (28). We showed that the type strain of *B. plebeius* (DSM 17135) can degrade this novel substrate and currently this strain is the only known isolated gut bacterium with the ability to utilize porphyran as its sole carbon source (FIG. 4). All other bacterial species that are known to degrade porphyran live in a marine environment, suggesting that this could be the source of the genes found in *B. plebeius*. Understanding how *B. plebeius* acquired its ability to degrade porphyran became a central question to my research.

Another novel polysaccharide utilization trait is the degradation of α-mannan by *B. theta*. *B. theta* has acquired several gene clusters that enable it to degrade this polysaccharide, a fungal cell wall glycan that has the same α1,2, α1,3 and α1,6 mannosidic linkages contained in animal *N*-glycans. Comparative genomic analysis between the *B. theta* PUL encoding α-mannan consumption and the *B. plebeius* PUL encoding porphyran utilization revealed that the PULs encoding each unique behavior (porphyran or mannan utilization) are located on putative ICEs that share regions of homology on both their left and right sides. Exploring these ICEs, as well as finding other related ICEs with polysaccharide utilization abilities has provided us with new knowledge about how the Bacteroidetes phylum can gain new polysaccharide utilization traits and established one genetic mechanism through which PULs are transferred between genomes.

Results and Discussion

***B. plebeius* grows on porphyran as a sole carbon source**

Hehemann et al. showed that *B. plebeius* encodes genes that may enable it to degrade porphyran. Specifically, they identified a ~50kb cluster containing 40 genes that is uniquely present in the genome of the *B. plebeius* type strain (20). Within this cluster, Hehemann et al. located two glycoside hydrolases (BACPLE 01670 and BACPLE 01689) that share amino acid sequence identity with GH16 homologues from microbes that degrade

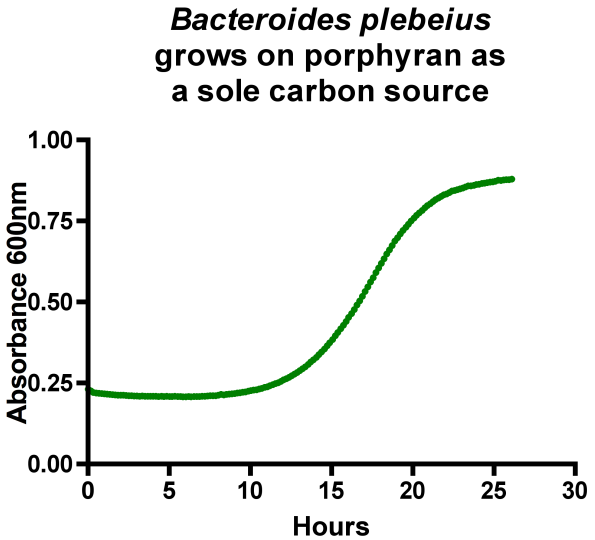


**Fig. 4.** *B. plebeius* can grow anaerobically with porphyran as a sole carbon source. *B. plebeius* is the only gut symbiont known to have this ability.

marine seaweed, indicating that they may be involved in the degradation of porphyran. Enzymatic tests on purified forms of these enzymes confirmed this hypothesis, but did not resolve whether or not this microorganism could degrade this polysaccharide.

To test if *B. plebeius* can in fact degrade porphyran, I inoculated *B. plebeius* in a custom chopped meat-supplemented minimal medium (see Methods) with porphyran as its sole carbon source. In an initial stage of this project, I determined that the addition of 0.5% chopped meat extract was required for this species to grow in defined carbon source conditions. Previous inability to grow this organism under such conditions had been a limitation to testing its metabolic ability directly. Growth curves illustrate that *B. plebeius* can utilize porphyran as its sole substrate (FIG. 4). No growth was seen in the negative control curves when *B. plebeius* was inoculated with the custom chopped meat-supplemented minimal medium and no carbon source.

**Porphyran activates the *B. plebeius* PUL**

After confirming that *B. plebeius* can in fact degrade porphyran as a nutrient, I wanted to determine if the bacterium was indeed expressing the same gene cluster that was discovered by Hehemann et al. (note that genetic techniques have not yet been tested in this recently isolated organism). Furthermore, I wanted to characterize the extent of the *B. plebeius* putative porphyran PUL and determine what genes are necessary for porphyran degradation.

Bacteroidetes typically organize the genes required for degrading a single polysaccharide into a cohesive polysaccharide utilization locus (PUL). When examining the putative porphyran PUL on the Integrated Microbial Genomes (IMG) database (29), we noticed that this PUL had surrounding regions that resembled a different PUL in *B. theta* contained on an ICE. I estimated the ends of the putative *B. plebeius* porphyran PUL and ICE by comparing this element to the

element in *B. theta* (FIG. 5A). Within the *B. plebeius* ICE, there are 12 putative glycoside hydrolases, 2 SusC/SusD-like protein pairs (BACPLE 01697-01698 and BACPLE 01704-01705) and a hybrid two-component system (HTCS) sensor/regulator (30) (BACPLE 01699) (FIG. 5A). Two of the glycoside hydrolases (BACPLE 01670 and BACPLE 01689) are homologues of GH16 from microbes that degrade marine seaweed. The putative functions of these genes were initially determined via a basic local alignment search tool (BLAST) analysis using the National Center for Biotechnology Information database (31).

I then asked whether or not *B. plebeius* is indeed activating this PUL by utilizing quantitative PCR (qPCR). I probed expression of 35/40 genes in this locus with cells grown on porphyran versus galactose to determine the extent of the locus that is induced in response to

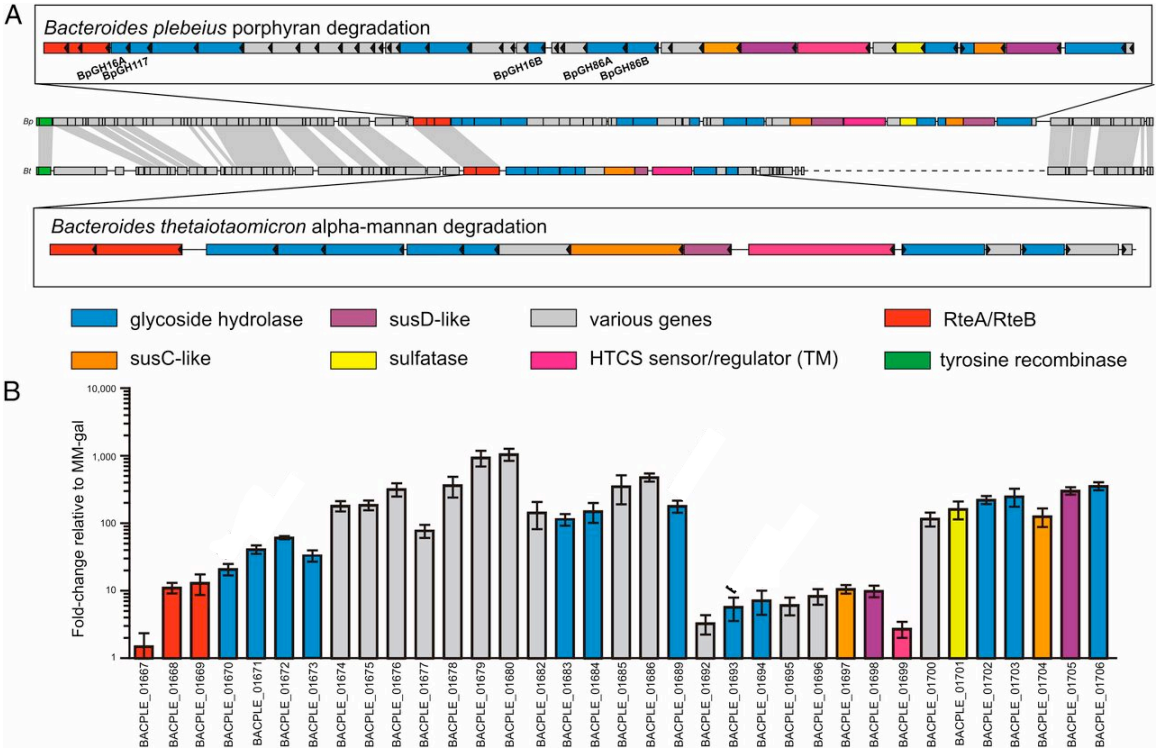


**Fig. 5.** (A) Synteny between the adjacent genomic region surrounding the porphyran PUL to a fungal α-mannan specific PUL in *B. theta*. (B) The expression of 35 genes within the PUL was monitored by RT-PCR upon incubation with porphyran compared to the control incubated with galactose. Five additional genes ranging in size from 126 to 489 bp (BACPLE 01681, 01687, 01688, 01690, and 01691) are present within this locus but were not probed by qPCR. Data in (B) are mean and s.d. of independent biological replicates.

porphyran as a sole carbohydrate source (FIG. 5B). Five additional genes ranging in size from 126 to 489 bp (BACPLE 01681, 01687, 01688, 01690, and 01691) are present within this locus but were not probed due to their small size. Compared to a galactose reference, I measured 15-1,000-fold increases in gene expression when *B. plebeius* was grown on porphyran, depending on the operon and the position of the gene (FIG. 5B).

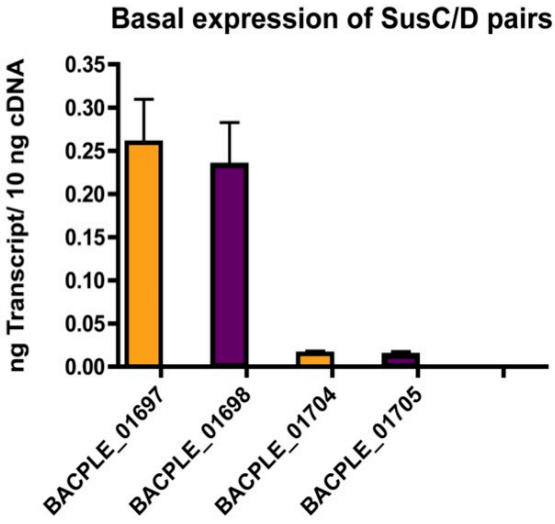


**Fig. 6.** Basal expression of both SusC/D pairs. BACPLE 01697 and BACPLE 01698 are 16-18 times more highly expressed basally, indicating that they are important for polysaccharide utilization.

Furthermore, I noticed that one cluster of genes in a potential operon (BACPLE 01692-01699) appeared to have much lower expression when measured by fold-change than the other genes in this locus. I suspected that this pattern might be caused by higher basal expression when *B. plebeius* is grown on galactose. To test this, I made a standard curve using qPCR cycle threshold values generated from known amounts of *B. plebeius* genomic DNA, assuming that it had equal copies of each gene. I then used qPCR to probe expression levels of the two SusC/D pairs using the same cDNA from the cells grown in the presence of galactose that were used for the fold-change analysis. I found that the porphyran utilization operon that appears to be lowly expressed, in fact, contains 16-18 times more cDNA, and therefore transcript, than the other porphyran utilization operon when *B. plebeius* is grown on galactose (FIG. 6). These results indicate that this operon is actually more highly expressed when grown in the absence of

porphyran and may be important for the bacterium to initially "sense" the presence of carbohydrate or has residual activation in response to galactose alone. Thus, these genes, which appear to be lowly expressed in Figure 5B, may be quite important for the degradation of porphyran.

The final three genes on the ICE that contains the porphyran utilization PUL are RteA and RteB homologues (BACPLE 01667-01669), a feature that is shared with the homologous ICE in *B. theta*. RteA and RteB are sensor histidine kinase and response regulator proteins that are important for the excision and mobilization of Bacteroides cTns carrying antibiotic resistance (32). I confirmed via resequencing that there is a frameshift mutation between BACPLE 01667 and 01668 where there normally is one RteB gene in *B. theta*. Therefore, these genes in *B. plebeius* are only partial RteB homologs that are likely to be no longer functional. This indicates that this locus, contained on an ICE, has lost its ability to excise readily from the *B. plebeius* genome, at least if it employs a similar mechanism as related elements (cTnDOT and cTnERL). This may be evolutionarily advantageous for the bacterium if porphyran is frequently present in the diet. Due to the competitive nature of the gut environment, it would be beneficial for *B. plebeius* to not share the genes that are required for porphyran utilization via HGT with other bacterial strains that are also attempting to thrive in the GI tract.

**The *B. plebeius* and *B. theta* PULs are contained on ICES and may have entered into their host chromosomes via HGT**

The fact that both the *B. plebeius* PUL and the *B. theta* PUL are contained on ICEs indicates that they may have entered the host genome via HGT. The *B. plebeius* ICE encodes a group of mobilization proteins as well as an integrase tyrosine recombinase that is directly

adjacent to a $tRNA^{Lys}$ (FIG. 5A). I successfully located a pair of identical 18 bp direct repeats on both sides of the ICE (see FIG. 7 and Table 1). One of the direct repeats overlaps with the 3' end of the $tRNA^{Lys}$, suggesting that this tRNA might be the integration site where the PUL entered the genome. Another copy of this repeat sequence is located 107,705 bp upstream. However, related genomes of *B. plebeius* are not available to determine, via comparative genomic analysis, if a corresponding version of this site exists in other strains that lack this element.

It was determined via a BLAST analysis that the *B. theta* ICE contains many genes that are homologous to the *B. plebeius* ICE (homologues connected by gray bars in FIG. 5A) (31). The *B. theta* ICE has homologous mobilization genes and a homologous integrase tyrosine recombinase. Additionally, the *B. theta* ICE also contains homologues of RteA and RteB. Therefore, these two ICEs have very similar organization, however they contain unique PUL "cargo".

Unlike the *B. plebeius* ICE, the one in *B. theta* is located adjacent to a $tRNA^{Phe}$, suggesting that this is its integration site. I located a 22bp direct repeat sequence that define the "left" and "right" ends of this putative ICE (see FIG. 7 and Table 1). These direct repeats are 90,450 bp apart. Unlike the case in *B. plebeius*, there are several other related *B. theta* genomes to compare the architecture of the putative insertion site to. When the surrounding genomic regions are compared to two other sequenced *B. theta* strains, these 90,450 bp are absent, suggesting that an HGT event with the ICE containing the α-mannan PUL inserted into the type-strain but not other ancestral strains.

Nested PCR provided evidence of the ability of these two ICEs to excise from the genome in order to transfer their element to other species, albeit at low frequency (FIG. 8). Primers were used that flanked the putative ICE insertion sites for both species (FIG. 7). Using

PCR, we detected excision of these elements (FIG. 8). Sequencing of the recombined amplification products validated that the 18 and 22 bp direct repeat sequences do indeed mediate recombination as the excised ICE junction and residual genomic site contained the precise sequences that would be predicted if homologous recombination occurred within these sites (FIG. 7). However, excision via this nested PCR analysis was only successful after 70 rounds of PCR, indicating that these excision events are extremely rare. It is worth noting that an antibiotic marked version of the B. theta α-mannan ICE has been shown to mobilize between isogenic versions *B. theta* type strain, but only when a mobilization-competent ICE, cTnERL, is triggered to mobilize *in trans* (33).

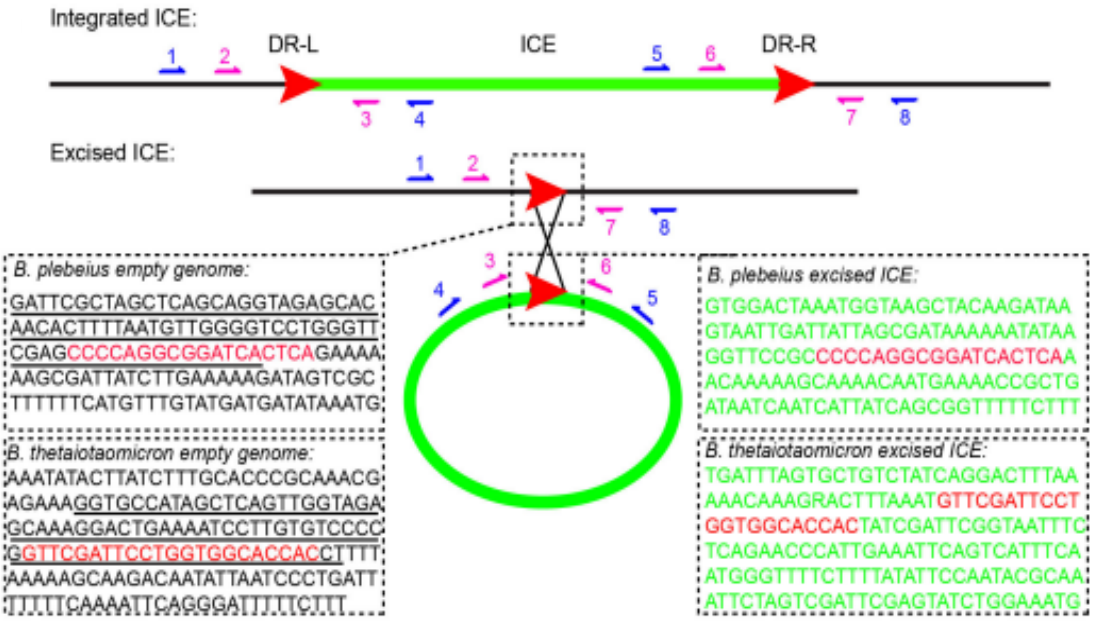


**Fig. 7** A schematic illustrating the primers used for nested PCR is shown. If excision occurred, PCR products connecting primers 1 and 8 as well as 4 and 5 were seen. Additionally, products from the reactions with primers 2 and 7 and 3 and 6 were seen, indicating the ICE properly excised. Predicted products were verified via sequencing. Red arrows indicate the "left" direct repeat (DR-L) and the "right" direct repeat (DR-R). Primers were designed for both the *B. plebeius* and *B. theta* ICEs and are listed in Table 7 in the Methods section.

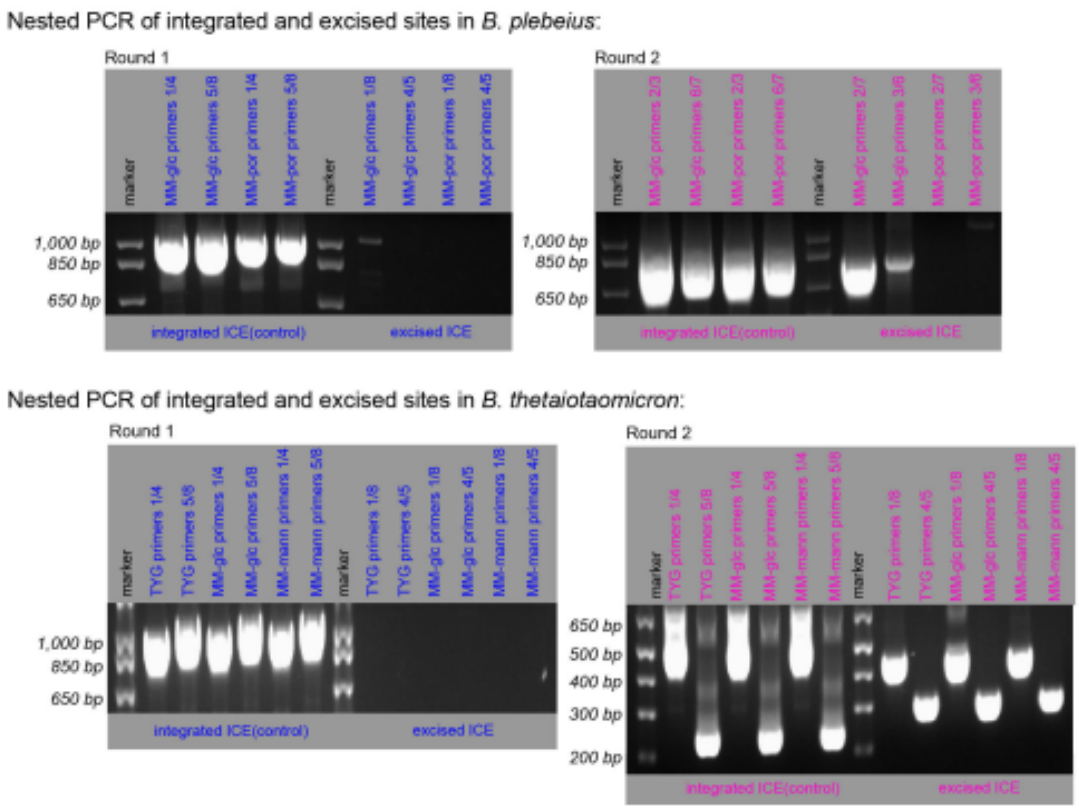


**Fig. 8** Two rounds of PCR were performed. The gels on the left are after one round of PCR. The gels on the right demonstrate the status of the ICEs after two rounds of PCR. After the first round of PCR, a weak *B. plebeius* ICE excision product was formed in the reaction with primers 1 and 8 (see FIG. 7 above for primer placements) in the Minimal Media-glucose (MM-glc) condition. After the second round of PCR, excision of each ICE was detected. Furthermore, the excision products for each species were confirmed via sequencing the products from the 2/7 and 3/6 reactions for *B. plebeius* and sequencing the products from the 1/8 and 4/5 reactions for *B. theta* (see FIG. 7 above). Approximately 250 ng of genomic DNA was utilized in each PCR reaction from cells grown in stationary phase in each condition (minimal media (MM), plus either glucose (glc), porphyran (por), α-mannan (mann) or tryptone-yeast extract medium (TYG)). For the second round of PCR, the DNA products from the first round were utilized after a PCR cleanup (Qiagen). The samples were eluted in the same volume of sterile water as the original amplification reaction. Each PCR reaction contained: 4mm of each primer plus Platinum Taq polymerase (Invitrogen) according to manufacturer's suggested reaction conditions. Reactions were cycled 35X with a 1.5 minute extension time at 72 °C and an annealing temperature of 59 °C.

## Identification of additional ICEs that may contain PULs

To understand the potential connections between Bacteroidetes ICEs and the evolution of new polysaccharide metabolism traits, I decided to make a phylogenetic tree based on the protein sequence encoded by the conserved transfer gene *traJ*, which is a homologous gene contained on both the *B. theta* and *B. plebeius* ICEs. The *traJ* gene is within the transfer (*tra)* operon, which is

essential for the mobilization of these elements and is therefore important for the HGT of ICEs (34). I conducted a thorough BLAST- based search for TraJ proteins using the National Center for Biotechnology Information (31) and Broad Institute of MIT and Harvard genome sequence databases (35) (36), in conjunction with a Hidden Markov Model-based TraJ classification scheme present in the pfam database (37). This effort yielded 501 TraJ amino acid sequences. The sequences were aligned using the default settings in the program MUSCLE (38) and the alignment was trimmed so that all the sequences were the same length. A maximum parsimony (MP) tree was constructed with this aligned file using MEGA 5 (39) and color-coded by the isolation source of the bacterial genome in which the various elements are located (FIG. 9).

Interestingly, all of the ICEs contained on this tree were Bacteroidetes isolates and the taxa tended to group by their isolation source. Furthermore, it was curious that the ICEs from marine isolates were not closely related to the *B. plebeius* ICE because the only other microbes that we know can degrade porphyran are marine bacteria. This observation suggests that the mechanism in which *B. plebeius* acquired its ability to degrade porphyran could have been more complex than a "one-step" HGT between a marine isolate and *B. plebeius.*

I decided to focus on the branch of this tree that contained the *B. theta* and *B. plebeius* elements, reasoning that this branch would be the most likely place to find more PULs. Interestingly, all of the species on this branch are gut isolates. To validate my first phylogenetic tree, I constructed another MP tree using DNA sequences of the ICEs from the bacteria on this branch. I constructed this tree using 2.4 kb of DNA sequence from the *tra* operon from these species. To obtain these sequences for the MP tree, I concatenated the sequences of 10 *tra* genes, *traH-traQ*. These concatenated files were entered into the program G-Blocks and trimmed using the default settings (40). Finally, I confirmed that all the sequences were the same length and

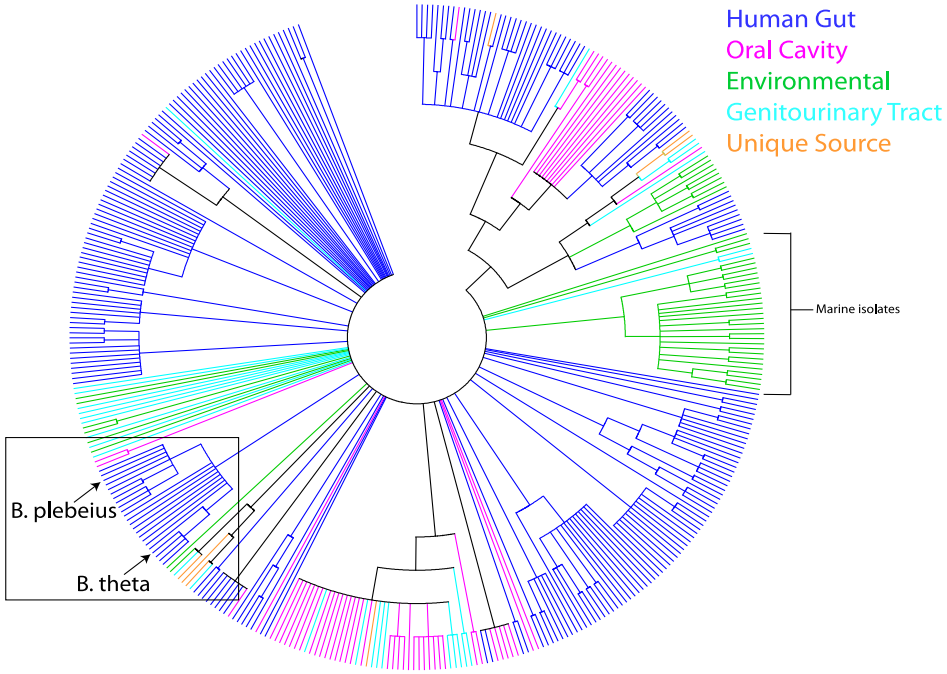


**Fig. 9** This evolutionary history was inferred using 501 amino acid sequences with the Maximum Parsimony method. The bootstrap consensus tree was inferred from 1000 replicates and the branches corresponding to partitions reproduced in less than 50% of the bootstrap replicates are condensed. The taxa are color-coded by the isolation source of the bacterial genome in which they are located. All ICEs were Bacteroidetes isolates that tended to group by their isolation source. The branch with the *B. plebeius* and *B. theta* ICEs is boxed. It is curious that the ICEs from marine isolates, highlighted above, were not closely related to the *B. plebeius* ICE because the only other microbes that we know can degrade porphyran are marine bacteria.

I used this file to make a MP tree with MEGA 5 (39). The second MP tree confirmed the relationships on the branch with the *B. theta* and *B. plebeius* ICEs from the first tree (FIG. 10).

I then looked for the direct repeats and insertion sites for the ICEs on this tree. For eight of the twelve ICEs that form a monophyletic group with *B. plebeius*, I located a pair of direct repeat sequences that define the "left" and "right" ends of these putative transposons. The direct repeats were the same for all eight of the sequences and they matched the direct repeat sequence for the *B. plebeius* ICE as well (Table 1). Furthermore, these ICEs all contain an integrase tyrosine recombinase that is directly adjacent to or near a $tRNA^{Lys}$, indicating that the ICEs

related to the *B. plebeius* ICE insert into the host genome at a $tRNA^{Lys}$ (FIG. 10). The longest possible direct repeats found at each end of the ICE are listed in Table 1, however, it is likely that only a minimum sequence is required for integration to occur. For the $tRNA^{Lys}$ insertion sequence, it appears that 13 bp are conserved at every insertion site.

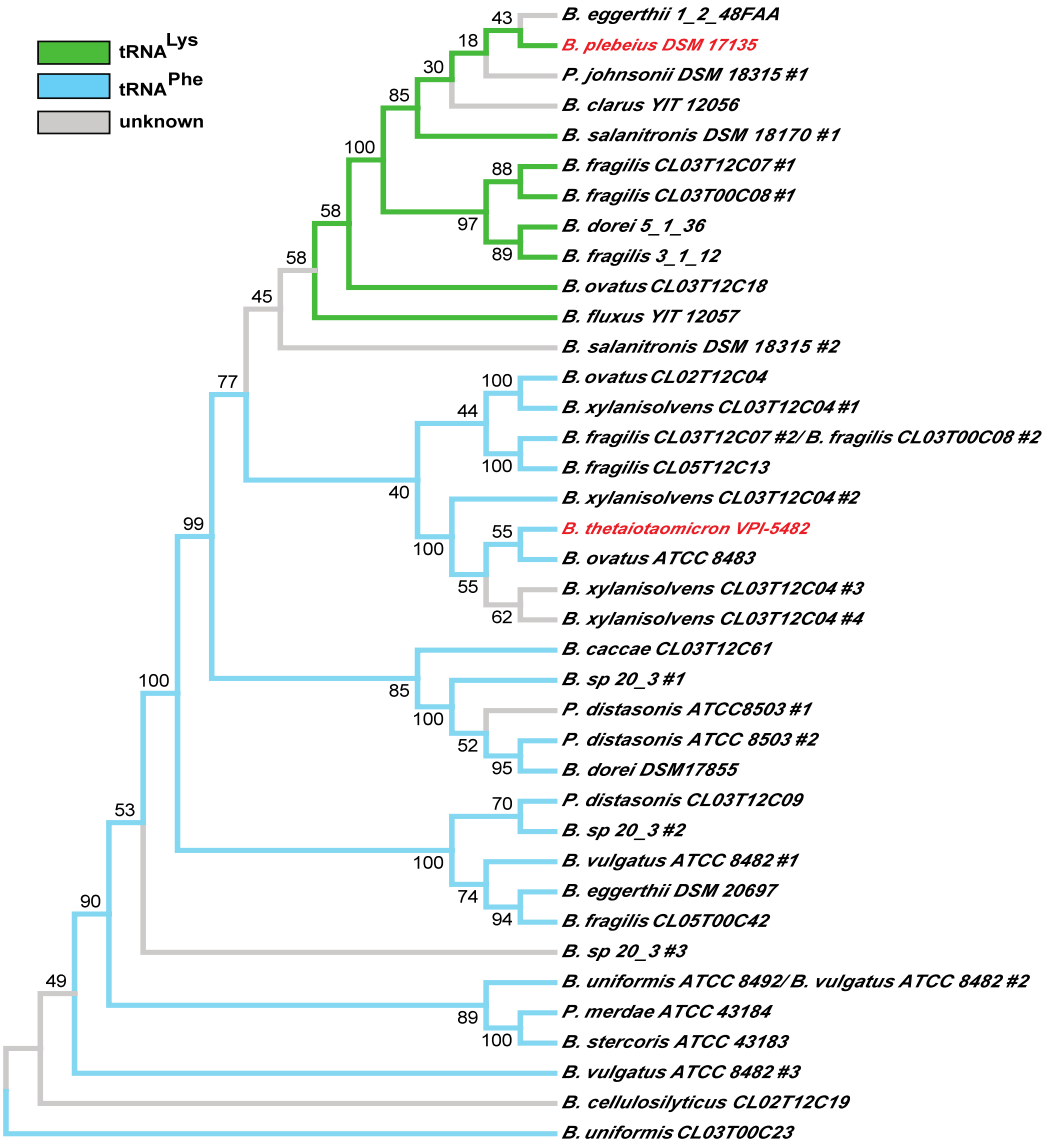


**Fig. 10** This evolutionary history was inferred using the Maximum Parsimony method. 2.4 kb of DNA from the *tra* operon was used in this analysis for each strain. The bootstrap consensus tree is inferred from 1000 replicates. Species with more than one related sequence are numbered accordingly. *B. theta* and *B. plebeius* are highlighted in red. ICEs with $tRNA^{Lys}$ insertion sites form a monophyletic group and are labeled in green. A similar grouping of related elements with $tRNA^{Phe}$ insertion sites are labeled blue.

| Species | Insertion sequence 5'→ 3' | tRNA |
|---|---|---|
| *B. plebeius DSM 17135* | ccccaggcggatcactca | Lysine |
| *B. salanitronis DSM 18315 #1* | ccccaagcggatcac | Lysine |
| *B. fragilis CL03T12C07 #1* | ccaggcggatcact | Lysine |
| *B. fragilis CL03T00C08 #1* | ccccaggcggatcact | Lysine |
| *B. dorei 5_1_36* | ccccaggcggatcact | Lysine |
| *B. ovatus CL03T12C18* | ccaggcggatcact | Lysine |
| *B. fluxus YIT 12057* | ccccaagcggatcactca | Lysine |
| *B. ovatus CL02T12C04* | gttcgattcctggtggcaccac | Phenylalanine |
| *B. xylanisolvens CL03T12C04 #1* | gttcgattcctggtggcaccac | Phenylalanine |
| *B. fragilis CL03T12C07 #2* | ggttcgattcctggtggcaccac | Phenylalanine |
| *B. fragilis CL03T00C08 #2* | ggttcgattcctggtggcaccac | Phenylalanine |
| *B. fragilis CL05T12C13* | gttcgattcctggtggcaccac | Phenylalanine |
| *B. xylanisolvens CL03T12C04 #2* | gttcgattcctggtggcaccac | Phenylalanine |
| *B. thetaiotaomicron VPI-5482* | gttcgattcctggtggcaccac | Phenylalanine |
| *B. ovatus ATCC 8483* | ggttcgattcctggtggcaccac | Phenylalanine |
| *B. caccae CL03T12C61* | gttcgattcctggtggcaccac | Phenylalanine |
| *P. distasonis ATCC8503 #2* | ggttcgattcctggtggcaccac | Phenylalanine |
| *B. dorei DSM17855* | ggttcgattcctggtggcaccac | Phenylalanine |
| *P. distasonis CL03T12C09* | ggttcgattcctggtggcaccac | Phenylalanine |
| *B. sp 20_3 #2* | ggttcgattcctggtggcaccac | Phenylalanine |
| *B. vulgatus ATCC 8482 #1* | ggttcgattcctggtggcaccac | Phenylalanine |
| *B. eggerthii DSM 20697* | ggttcgattcctggtggcaccac | Phenylalanine |
| *B fragilis CL05T00C42* | ggttcgattcctggtggcaccac | Phenylalanine |
| *B. uniformis ATCC 8492* | ggttcgattcctggtggcaccac | Phenylalanine |
| *B. vulgatus ATCC 8482 #2* | ggttcgattcctggtggcaccac | Phenylalanine |
| *P. merdae ATCC 43184* | ggttcgattcctggtggcaccac | Phenylalanine |
| *B. stercoris ATCC 43183* | ggttcgattcctggtggcaccac | Phenylalanine |
| *B. vulgatus ATCC 8482 #3* | ggttcgattcctggtggcaccac | Phenylalanine |
| *B. uniformis CL03T00C23* | ggttcgattcctggtggcaccac | Phenylalanine |

**Table 1.** The direct repeats that mark the insertion sites for the ICEs contained on the branch with *B. theta* and *B. plebeius* elements are shown. The minimum conserved insertion sequence for the $tRNA^{Lys}$ insertion site is highlighted in green while the minimum conserved insertion sequence at a $tRNA^{Phe}$ is highlighted in blue.

Similarly, I found the direct repeat sequences that define the "left" and "right" ends of putative transposons for twenty out of the twenty-five ICEs remaining on this branch. These sequences match the direct repeats for the *B. theta* ICE (Table 1). These elements also contain an integrase tyrosine recombinase that is directly adjacent to or near a $tRNA^{Phe}$, indicating that the

remaining ICEs on this branch insert into the genome at a conserved target site. A total of 22 base pairs are conserved at all $tRNA^{Phe}$ insertion sites.

**Genetic analysis of the ICEs that are related to the *B. theta* and *B. plebeius* conjugative elements**

To determine what other gene clusters are contained on these ICEs, I explored the genetic makeup of the elements on the branch with the *B. theta* and *B. plebeius* elements. Using the Integrated Microbial Genomes (IMG) system (29), the BioCyc database (41), and the Broad Institute of MIT and Harvard database (35) (36), I was able to thoroughly examine the ICEs on this branch, as well as their associated cargo genes (FIG. 11).

I expected to find many PULs, which turned out to be true. Interestingly, the ICEs on this branch also harbor a diverse range of other cargo genes. *B. plebeius* was the only species with a $tRNA^{Lys}$ insertion site that contained a PUL. The other elements with $tRNA^{Lys}$ insertion sites appear to predominantly carry drug or antibiotic resistance genes indicated in dark green (FIG. 11). In general, each ICE's novel cargo was contained at a similar location and this cargo was positioned adjacent to genes that are homologous to RteA and RteB, suggesting that they all share a related mobilization mechanism with other *Bacteroides* ICEs. The ICEs with the $tRNA^{Lys}$ insertion sites that may harbor drug resistance capabilities all contain arsenite reductase genes, as well as other *ars* genes, suggesting that these elements harbor resistance to arsenic. In fact, arsenite reductase enzymes have been shown to help confer arsenic resistance to *Herminiimonas arsenicoxydans* (42). It should also be noted that arsenic was first used as an antibiotic in the early 1900s to treat syphilis and trypanosomiasis, so it is possible that these elements carry one of the first gene clusters for antibiotic resistance (43). Alternatively, arsenic is present in our

environment as a natural contaminant that is found in soil, air, water, and even food, so it is also possible that these bacteria gained their arsenic resistance more recently (43). It is known that members of the Bacteroidetes phylum do in fact exchange antibiotic resistance genes via HGT—including on elements with similar genetic organization to the $tRNA^{Phe}$ and $tRNA^{Lys}$ elements in this study (*e.g.*, cTnDOT and cTnERL) —so it is possible that the strains containing arsenite reductase genes may be able to transfer these clusters via HGT using their ICEs (44).

In addition to harboring genes for drug and antibiotic resistance, there were eight ICEs with insertions sites at a $tRNA^{Phe}$ that appear to carry capsule synthesis genes; these are shown in yellow in FIG. 11. Many Bacteroidetes are known to secrete dense polysaccharide capsules on their cell surface. In fact, the *B. theta* type strain alone contains eight capsule loci within its genome. These capsules may serve an evolutionary function that helps the bacteria to evade the immune system and *Bacteroides* capsules have also been shown to be co-regulated with PUL expression (45). Therefore, it may be advantageous for bacteria to exchange these capsule polysaccharide synthesis genes via HGT in the same way that they exchange catabolic genes.

There were also five additional PULs that I found on ICEs that contain a $tRNA^{Phe}$ insertion sequence, indicating that this family of ICEs is capable of transferring carbohydrate degradation traits. These PULs are shown in dark blue (FIG. 11). Discovering what polysaccharides these PULs are targeting, if these PULs are being expressed by the bacteria in vivo, and whether or not these PULs are mobile will be important future steps in understanding the evolutionary mechanisms of these conjugative elements and their associated polysaccharide utilization cargo.

Finally, there were sixteen ICEs on the tree that carry cargo with unknown functions. For most of these ICEs, there was simply not enough data published for me to determine what their

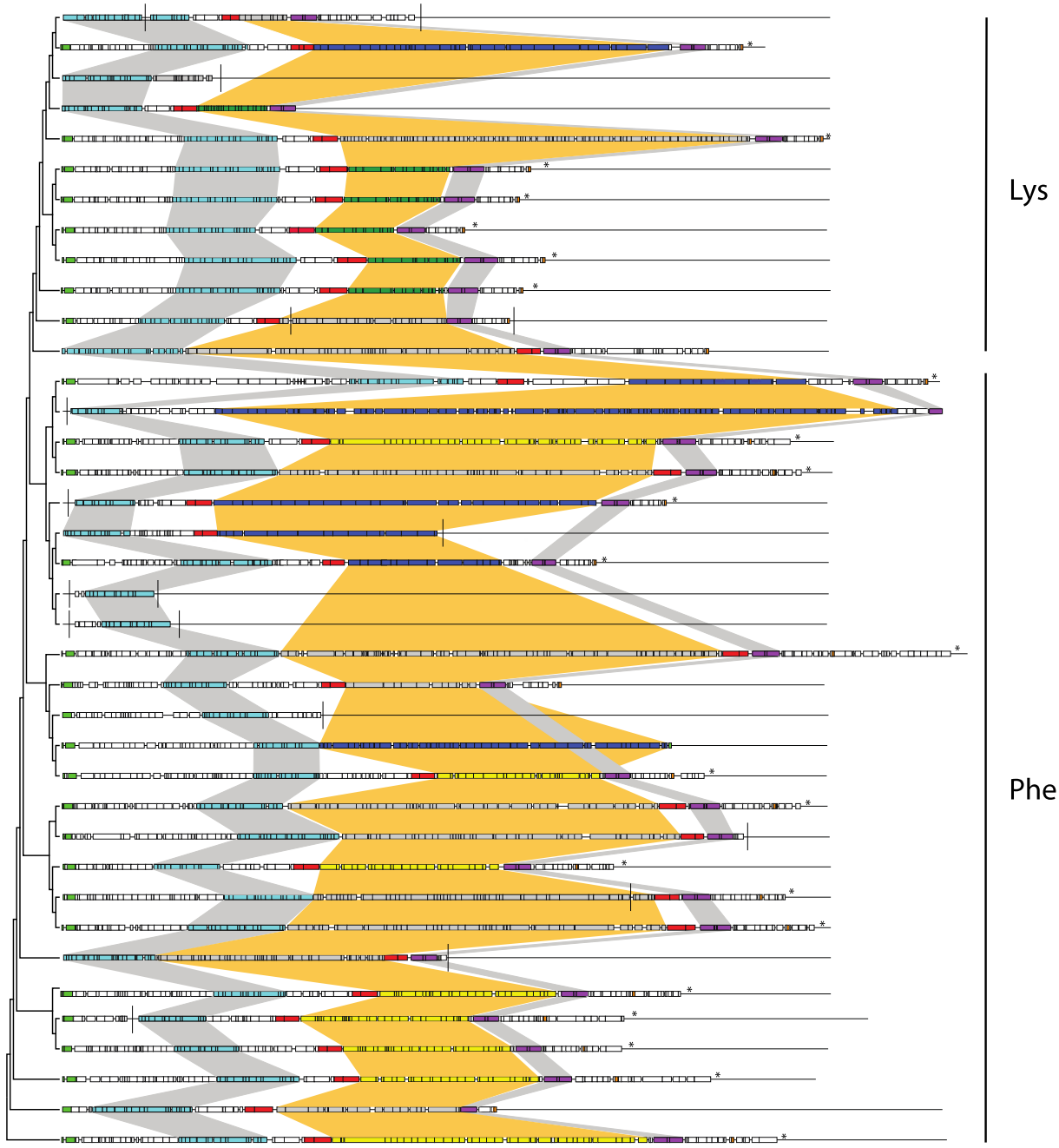


**Fig. 11** The genetic analysis of the branch with the *B. theta* and *B. plebeius* ICEs is shown. The grey regions indicate genetic homology. The orange regions denote the unique cargo genes associated with the ICEs. Dark blue cargo denotes polysaccharide utilization loci. Yellow cargo indicates capsule synthesis genes. Dark green cargo denotes drug or antimicrobial resistance genes. Bright blue genes are the transfer operon and bright green genes are integrase tyrosine recombinases that denote the "left" end of ICEs. The purple genes are topoisomerases found on the "right" end of the ICEs. Red genes are homologues of RteA/RteB. The small orange genes on the far right are excisionases. The ICEs that insert at a $tRNA^{Lys}$ are on the top of the tree and the ones that insert at a $tRNA^{Phe}$ are on the bottom of the tree. Asteriks at the end of the ICEs indicate the sequences where direct repeats were found. Vertical bars demonstrate that there was no more genetic information available in the databases for that area of the genome.

functions were. It is possible that these ICEs carry genes for antibiotic resistance, capsule synthesis or polysaccharide utilization. It is also possible that these ICEs harbor genes with novel functions that will become clear as more data continues to become available. As new data continues to be published, it will be interesting to explore what other cargo genes these ICEs are harboring. Moreover, casual perusal of the remaining majority of ICEs on other branches of my protein-based tree, suggest that other elements also contain carbohydrate utilization traits. However, the organization and target site preference for these remaining elements are different and each family will need to be explored more rigorously.

Future directions

This work utilized the depth of genomic information that has recently become available to analyze the evolution, insertion sequences and genetic makeup of PULs on ICEs as well as other conjugative elements that harbor diverse genetic cargo. ICEs appear to be one important mechanism by which PULs have entered the genomes of the gut microbiota. To better understand this mechanism of PUL evolution, investigating the five novel PULs on the branch with the *B. theta* and *B. plebeius* elements will be important. Discovering what polysaccharides these PULs are targeting and whether or not these elements are mobile will allow us to better appreciate one method by which PULs are transferred from one bacterium to another. Furthermore, as more data continues to become available, it is imperative to study the sixteen ICEs with cargo that have unknown functions on this branch. This data will elucidate the diversity of cargo genes contained on ICEs and may help us discover new PULs that could be on mobile elements.

In addition to investigating the elements on this one branch, exploring the rest of the amino-acid based tree will provide even more information about ICEs. In particular, I am curious if related ICEs always group by insertion sequence (casual perusal suggests this is the case). Additionally, I want to know what other cargo genes are contained on the elements on other branches. Do ICEs exclusively carry PULs, capsule synthesis genes and genes that confer antibiotic resistance? Or do they harbor a diverse and vast array of genetic clusters?

Finally, studying ICEs with PULs that appear to be recently acquired is an important future step in understanding the extent of bacterial evolution via conjugative transposition. After the construction of both MP trees for this study, new sequence data became available for a few *B. xylanisolvens* strains isolated from the GI tracts of pigs. These pigs were recently fed a diet that has both yeast and mannan. Interestingly, the sequences for these bacterial strains contain an element that appears to be identical to the α-mannan ICE in *B. theta.* Furthermore, *B. xylanisolvens* strains from other sources do not have this element. Therefore, it is possible that these pig strains recently acquired this element via HGT. If this were true, we would expect that this element could possibly excise from the genome when the strains are grown on mannan. In the near future, I plan to test whether or not these ICEs excise from the genome in the presence of mannan. If these elements are capable of excision, transferring these ICEs from one bacterium to another in vivo will provide evidence that Bacteroidetes polysaccharide utilization traits can be moved between species and thereby transfer new polysaccharide utilization traits into related bacteria via HGT utilizing ICEs. It will also provide a tractable model to study the function of these elements using molecular genetics and other approaches.

Methods

**Bacterial growth**

*B. plebeius* DSM 17135 and *B. thetaiotamicron* VPI-3731 were grown anaerobically (Coy Lab Products, Grass Lake, MI) in a chopped meat broth medium (Table 2). First, 2X filter sterilized minimal media was prepared as described by Martens et al. (17), however, 0.5% of chopped meat extract, Trace Mineral Solution (Table 3), Balch's vitamin mixture (46), Amino Acid solution (Table 4) and Purine and Pyrimidine solution (Table 5) was added to the medium. The 2X media was then added to a 2X porphyran solution. The concentration of the 2X porphyran solution was 10mg/ml in pure dd$H_2O$. The growth profiles of *B. plebeius* were assessed in 96-well plates (Costar) and optical density (OD) was measured utilizing a robotic plate handler (Biotek, Winooski, VT). As a control, we also measured the OD of wells that lacked porphyran.

| **Compound** | **Amount** |
|---|---|
| Beef extract | 2.5 g |
| Pancreatic digest of casein | 7.5 g |
| Yeast Extract | 1.25 g |
| $K_2HPO_4$ | 1.25 g |
| L-Cysteine | 250 mg |
| Fructose | 250 mg |
| Galactose | 250 mg |
| Glucose | 250 mg |
| Mannose | 250 mg |
| N-Acetyl Glucosamine | 250 mg |
| Xylose | 250 mg |
| Vitamin $K_3$ | 250 μg |
| Vitamin $B_{12}$ | 2.5 μg |
| Hematin/ L-Histidine Stock | 7.6 μM/ 800 μM final |
| Balch's Vitamin Mixture | 2.5 ml |
| Trace Mineral Solution | 2.5 ml |
| Purine and Pyrimidine Solution | 2.5 ml |
| Amino Acid Solution | 2.5 ml |

**Table 2. Chopped Meat Broth Contents.** The broth was dissolved in dd$H_2O$ and brought up to 250 ml. The pH was brought to 7.2. The medium was filter sterilized and stored at 37°C in the anaerobic chamber.

| Compound | Amount |
|---|---|
| EDTA | 0.5 g |
| $MgSO_4*7H_2O$ | 3 g |
| $MnSO_4*H_2O$ | 0.5 g |
| NaCl | 1 g |
| $FeSO_4*7H_2O$ | 0.1 g |
| $CaCl_2$ | 0.1 g |
| $ZnSO_4*7H_2O$ | 0.1 g |
| $CuSO_4*5H_2O$ | 0.01 g |
| $H_3BO_3$ | 0.01 g |
| $Na_2MoO_4*2H_2O$ | 0.01 g |
| $NiCl_2*6H_2O$ | 0.02 g |

**Table 3. Trace Mineral Solution.** The contents were dissolved in $ddH_2O$ to 1 L and brought to a pH of 7.0. The solution was stored at room temperature after being filter sterilized.

| Compound |
|---|
| Alanine |
| Arginine |
| Asparagine |
| Aspartic Acid |
| Cysteine |
| Glutamic Acid |
| Glutamine |
| Glycine |
| Histidine |
| Isoleucine |
| Leucine |
| Lysine |
| Methionine |
| Phenylalanine |
| Proline |
| Serine |
| Threonine |
| Tryptophan |
| Tyrosine |
| Valine |

**Table 4. Amino Acid Solution.** 62.5 mg of each amino acid was dissolved in $ddH_2O$ and brought to 1 L and a pH of 7.0. The solution was stored at room temperature after being filter sterilized.

| Compound |
|---|
| Adenine |
| Guanine |
| Thymine |
| Cytosine |
| Uracil |

**Table 5. Purine and Pyrimidine Solution.** 200 mg of each nucleotide was added to $ddH_2O$ that was brought to 1 L. The pH was brought to 7.0. The solution was stored at room temperature after being filter sterilized.

## Quantitative PCR (qPCR) of *B. plebeius* PUL genes

*B. plebeius* was grown on galactose or porphyran as its sole carbon source in triplicate. Bacterial cells were harvested during the exponential growth phase ($A_{600}$ values of 0.57, 0.56, and 0.57 for porphyran and 0.41, 0.42 and 0.43 for galactose). RNA was protected using RNAProtect (Qiagen) and then extracted with the RNeasy kit (Qiagen). Finally, cDNA was synthesized with the SuperScript III Reverse Transcriptase kit (Invitrogen). A Mastercycler Realplex (Eppendorf) was used with a SYBR green-based master mix (Kapa Biosystems, Woburn, MA) and the primers listed in Table 6 to measure the transcript of each PUL gene. The transcript abundance was normalized using 16S rRNA transcript primers.

| Primer Name | Sequence (5'→ 3') |
|---|---|
| BP1667F | AGGTCGCTTTCGTGAGGATTTG |
| BP1667R | CCGCACATTACCAGGCCAGTCATA |
| BP1668F | AAGACAAATCGGACCATAATACAATAACTA |
| BP1668R | AGAGCGGCCACTACTTCACG |
| BP1669F | TACTGGCCAATCTACTTACAAACGCTATC |
| BP1669R | AGTCCCAAACCAAAACCTTCAGA |
| BP1670F | ATTGTGGCGCGATGATTTTC |
| BP1670R | CCCGTGCCACCAGTGTAGTTGT |
| BP1671F | GAGCGTAAGGCCGATTCATTGG |
| BP1671R | AGCATTGCACTCGGGTCTCTTCTTACTA |
| BP1672F | ATGGAAGGCGGTGATTGAAAGAGA |
| BP1672R | GCCGCACATATCCAGGAGACCACAT |
| BP1673F | CTCCGGATATACAGACGTGCTTGAT |
| BP1673R | GATGCCCGCTATGTACTTCTTGTC |
| BP1674F | AACAAGGCGGGAGGTATTTC |
| BP1674R | GCCCGGTGCGCTCAACAGATTA |

| | |
|---|---|
| BP1675F | GCTTGCCGGCATGAGGTGATAGGT |
| BP1675R | CGGAAGGTCGCGTGTGGCATACT |
| BP1676F | TCAATGCCTGCGAACACAACGAGAC |
| BP1676R | CGCAACCTATCACGGCCACAGTATCA |
| BP1677F | CTTGGCGGCAGATGGATAGGT |
| BP1677R | GCCGAGGTCGCGTAGAGCAAGTG |
| BP1678F | TGCTTGCCCGCCATAAAAAT |
| BP1678R | GTGCGCGGATAGTCGGAACC |
| BP1679F | GACCATGCAGCTCCGGACACAGTT |
| BP1679R | GCGTTTATCACCCCTACACCCTTTTCA |
| BP1680F | TCCGGAATTGATTGAAGCCTACAG |
| BP1680R | ATTTTGCCACATATTCCTCCCATTC |
| BP1682F | GTGGCCGACGGACTGATTCTCT |
| BP1682R | AGCTCCTTATAACTCCCCCTGGTGT |
| BP1683F | TCGAGTATGCGGAAAAACCTGACA |
| BP1683R | GAATTTTTGCCCTTTTGCGTATGAGC |
| BP1684F | ATATGGCGAAAGGGGAACTACAGA |
| BP1684R | CGTCAGGCAGGCTCAACCAT |
| BP1685F | GTGTGGCCGGTTCTTGA |
| BP1685R | AGTCCATTCTGCTCCGTTGTTGA |
| BP1686F | TGATGGAACACCGGCTGAGGAG |
| BP1686R | TTGTAACCGGTATAAGGGAGATT |
| BP1689F | CGGCCGCCTGCAACTTTCAA |
| BP1689R | ATATTTATCGCCCGGCTTTCCATCA |
| BP1692F | ATCCTTTGCTCCGTTCATATCTTTACA |
| BP1692R | TATCCTCTTCGTGACATCAATCTCCA |
| BP1693F | GGTGTTCCTCCCGGTATTTATTGTC |
| BP1693R | TTGCCTACCTCATGCGTCTTGTTAT |
| BP1694F | CCGCTCTTTCTGGGGACCGTTCTCTTAT |
| BP1694R | AGCCCATGCACCCGCCTTCACT |
| BP1695F | GGTTTTGTTCGGAGCATGTGATAAT |
| BP1695R | AAACTTGGAAACCTCGCCTGTCT |
| BP1696F | TCCGCGTGTATGATCCTAACCTC |
| BP1696R | ACGCCAGTATCCGCACATTTG |
| BP1697F | CTTTGCGTGGCTTGTTTTATTTTTATCT |
| BP1697R | ATCGGTTTTCTCCCATTTGTCATTC |
| BP1698F | GTCAATAATAACGCGCCAATACAGAA |
| BP1698R | ACCCCCAAGCCGGAGATAACT |
| BP1699F | GCATGCCATCGAGTCAAAACCAC |
| BP1699R | AGCCCTGAGCCTTCTGTACCTAAATA |
| BP1700F | GGTGCGGCAATCTGTCGTGAGG |
| BP1700R | CGGTTCCGGCTGTTGTAGG |
| BP1701F | AATGGGGCTCCGAAGGCACAA |
| BP1701R | CTTTACGCAACGCAGATACGATTTCAG |
| BP1702F | TCTGCGTTTCGGACTTTCTTCTCACA |

| | |
|---|---|
| BP1702R | ACCTTCGGGCTCATGCAGTCTTTC |
| BP1703F | CACTGCCCGAAGGATATACTCACAC |
| BP1703R | CGACATCCAGAAACCGAAAACC |
| BP1704F | TGATGCGGCGATATTTACTTACGA |
| BP1704R | CAGGCTTGTTTAGATTGGGGTTTC |
| BP1705F | TTGACAGATGGCGGAAAGAAGGT |
| BP1705R | CATAAACGGATAGCTGGTGAAAGTGG |
| BP1706F | TACCGGCGGGCTGAAAGAATACTG |
| BP1706R | TAATGGAGCGGTCGGCACTGATAAC |
| *B. plebeius* 16SF | GGTAGTCCACACGGTAAACGATGGA |
| *B. plebeius* 16SR | CCCGTCAATTCCTTTGAGTTTC |

**Table 6. Oligonucleotide primers used for Quantitavie PCR (QPCR).** All primers were designed for use with *B. plebeius.*

### Standard curve to measure basal expression of *B. plebeius SusC/D pairs*

Standard curves were generated using serial dilutions of genomic DNA. Primers specific for each SusC and SusD gene were used to generate each curve (BP1697F/BP1697R, BP1698F/BP1698R, BP1704F/BP1704R, and BP1705F, and BP1705R). The standard curves were generated via qPCR reactions with a Mastercylcer Realplex (Eppendorf) and a SYBR green-based master mix (Kapa Biosystems, Woburn, MA). The standard curves were constructed by plotting $C_T$ vs. $\log_{10}^{copies}$ and the basal expression of each gene was calculated by comparing the standard curves to qPCR reactions done in triplicate with cDNA grown on galactose.

### Nested PCR to test for excision of ICEs

Nested PCR was performed to test for excision of the *B. plebeius* and *B. theta* ICEs. For each ICE, eight primers were designed that aligned on each end of the direct repeats that designated the “left” and “right” ends of the element. Two primers aligned in the genome on the “left side” of the ICE and pointed towards the “left” direct repeat (ICE 1 and ICE 2). Similarly, two primers aligned in the genome on the “right side” of the ICE and point towards the “right” direct repeat

(ICE 7 and ICE 8). Within the ICE, there were two primers on the "left side" of the ICE that pointed towards the "left" genomic flanks (ICE 3 and ICE 4) and there were two primers on the "right side" of the ICE that faced the "right" genomic flanks (ICE 5 and ICE 6). This schematic is illustrated in the supplementary material of a recent paper by Hehemann et al. and in FIG. 7 (27). See Table 7 for the primers used for both the *B. plebeius* and *B. theta* elements.

PCR reactions were performed using various primer combinations for two rounds of PCR. Approximately 250 ng of genomic DNA was utilized from cells grown in stationary phase in each condition (minimal media (MM), plus either glucose (glc), porphyran (por), α-mannan (mann) or tryptone-yeast extract medium (TYG)). For the second round of PCR, the DNA products from the first round were used after a PCR cleanup (Qiagen). The samples were eluted in the same volume of water as the original amplification reaction. Each PCR reaction contained: 4mm of each primer plus Platinum Taq polymerase (Invitrogen) according to manufacturer's suggested reaction conditions. Reactions were cycled 35X with a 1.5 minute extension time at 72 °C and an annealing temperature of 59 °C. If excision occurred, we saw products with the primer combinations ICE1/ICE8 or ICE2/ICE7, which would represent the genomic backbone, and we saw the product of the circular element with the primers ICE 3/ICE 6 or ICE4/ICE5. The PCR products that were hypothesized from an excision event were verified via sequencing.

| **Primer Name** | **Organism** | **Sequence (5'→ 3')** |
|---|---|---|
| *B. plebeius* ICE 1 | *B. plebeius* | AGGAGAACGTACCCGTCTGG |
| *B. plebeius* ICE 2 | *B. plebeius* | CATTCTGAAGCAGGCGCTGG |
| *B. plebeius* ICE 3 | *B. plebeius* | CCAATCGCGCATTCTTTTTGCG |
| *B. plebeius* ICE 4 | *B. plebeius* | TGCTGTAATAGGTTTACCGTCC |
| *B. plebeius* ICE 5 | *B. plebeius* | AGAAAAGCATAACAGTCATGAGC |
| *B. plebeius* ICE 6 | *B. plebeius* | AAGTATTCTGAAAGATAATACCCC |
| *B. plebeius* ICE 7 | *B. plebeius* | ATGCGAATTGGCTATGCCTGC |
| *B. plebeius* ICE 8 | *B. plebeius* | TCTGCCCGGAGGATGTAGCC |
| *B. theta* ICE 1 | *B. theta* | CAGTGTATAATAATTGGTATTGTGG |
| *B. theta* ICE 2 | *B. theta* | ATTGTTTGTAGTGCCCTGATCG |
| *B. theta* ICE 3 | *B. theta* | CCATTACATTCAAGCAGGTATGC |

| | | |
|---|---|---|
| *B. theta* ICE 4 | *B. theta* | CATTAGAATAGGTGCTTCACC |
| *B. theta* ICE 5 | *B. theta* | CGGACAGATACAGACCTATCC |
| *B. theta* ICE 6 | *B. theta* | GATTTAGTGCTGTCTATCAGGAC |
| *B. theta* ICE 7 | *B. theta* | GAAAAGCTGTTGCTTAACATGAG |
| *B. theta* ICE 8 | *B. theta* | CAGTCTATGCATGTCTTCTG |

**Table 7. Primers used for nested PCR.** The primers are numbered like they are shown in FIG. 7.

## Searching for TraJ protein sequences

To construct a phylogenetic tree, I first searched for protein sequences for the *traJ* gene. To find TraJ protein sequences, I performed a BLAST analysis with the *B. theta* TraJ sequence using the using the National Center for Biotechnology Information database (31). I used all sequences from the BLAST analysis that had an e-value greater than $10^{-10}$. I then BLASTed the sequence with the lowest e-value taken from my initial search. I performed a total of eight BLAST searches in this way to find the TraJ sequences. I also performed one similar BLAST search on the Broad Institute of MIT and Harvard database for the Bacteroidetes (35) and the Parabacteroidete*s* (36). Additionally, I included all sequences from a Hidden Markov Model-based TraJ classification scheme present in the pfam database (37). I retrieved a total of 501 sequences.

## Construction of radial TraJ tree

The 501 amino acid sequences were aligned in MUSCLE (38) using the default settings and the sequences were trimmed so that they were all the same length. A MP tree was constructed using the default settings in MEGA 5 (39). The bootstrap consensus tree was inferred from 1000 replicates and the branches corresponding to partitions reproduced in less than 50% of the bootstrap replicates were condensed.

**Construction of DNA MP tree using sequences from the *tra* operon**

To confirm the relationships on the branch with *B. theta* and *B. plebeius* elements, another phylogenetic tree was constructed with these species using their DNA sequences from many genes within the *tra* operon. Specifically, sequences were taken from the 5' end of the *traH* gene to the 3' end of the *traQ* gene. The sequences were aligned with MUSCLE (38) using the default settings and trimmed using the default setting of the program G-Blocks (40). Furthermore, the 3' ends of the sequences were trimmed manually so that there was exactly 2.4kb of each sequence. The aligned, trimmed file was then used to make a MP tree in MEGA 5 (39) with the default settings. The bootstrap consensus tree was inferred from 1000 replicates.

**Searching for direct repeat sequences that define the "left" and "right" ends of the ICEs**

To find the insertion sequences for the various ICEs, I performed a BLAST search with the organism of interest using the tRNA sequence that was near the integrase tyrosine recombinase plus an additional 100 bp off the 3' end. The BLAST search was done using the National Center for Biotechnology Information database (31). I then searched in the BLAST results for two 18-22 bp matching sequences that were located on the "left" and "right" ends of the respective ICE.

**Genetic comparisons of cargo genes on ICEs**

To make the genetic comparisons shown in FIG. 9, published data from the Integrated Microbial Genomes (IMG) database (29), the BioCyc database (41), and the Broad Institute of MIT and Harvard database were used (35) (36). Gene maps from these sites were used to align homologues, perform comparisons, and discover the cargo genes contained on each ICE.

Acknowledgments

This research project would not have been possible without the support of many people. I express my gratitude to my advisor, Dr. Eric Martens, who has offered invaluable guidance, support and assistance with this project. Additionally, I would like to thank Nicholas Pudlo and Elizabeth Cameron for their assistance in the lab and I would like to thank Dr. Matthew Chapman for his co-sponsorship of my project. I am also grateful for the help of Dr. Blaise Boles, who agreed to evaluate my work. Lastly, I would like to express my gratitude to Dr. Jan-Hendrik Hehemann for his collaboration and to Dr. Timothy McKay, the director of the Honors Program at the University of Michigan, for his advice and financial support throughout the summer of 2012 through the Honors Summer Fellowship.

References


1. Martens EC*, et al.* (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* in press.
2. Palmer C, Bik EM, Digiulio DB, Relman DA, & Brown PO (2007) Development of the Human Infant Intestinal Microbiota. (Translated from Eng) *PLoS Biol* 5(7):e177 (in Eng).
3. Koropatkin NM, Cameron EA, & Martens EC (2012) How glycan metabolism shapes the human gut microbiota. (Translated from eng) *Nat Rev Microbiol* 10(5):323-335 (in eng).
4. Round JL & Mazmanian SK (2009) The gut microbiota shapes intestinal immune responses during health and disease. (Translated from eng) *Nat Rev Immunol* 9(5):313-323 (in eng).
5. Hill MJ (1997) Intestinal flora and endogenous vitamin synthesis. (Translated from eng) *Eur J Cancer Prev* 6 Suppl 1:S43-45 (in eng).
6. Wardwell LH, Huttenhower C, & Garrett WS (2011) Current concepts of the intestinal microbiota and the pathogenesis of infection. (Translated from eng) *Curr Infect Dis Rep* 13(1):28-34 (in eng).
7. Flint HJ, Bayer EA, Rincon MT, Lamed R, & White BA (2008) Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. (Translated from eng) *Nat Rev Microbiol* 6(2):121-131 (in eng).
8. Packey CD & Sartor RB (2009) Commensal bacteria, traditional and opportunistic pathogens, dysbiosis and bacterial killing in inflammatory bowel diseases. (Translated from eng) *Curr Opin Infect Dis* 22(3):292-301 (in eng).
9. O'Keefe SJ*, et al.* (2009) Products of the colonic microbiota mediate the effects of diet on colon cancer risk. (Translated from eng) *J Nutr* 139(11):2044-2048 (in eng).
10. Ley RE, Turnbaugh PJ, Klein S, & Gordon JI (2006) Microbial ecology: human gut microbes associated with obesity. (Translated from eng) *Nature* 444(7122):1022-1023 (in eng).
11. Cani PD*, et al.* (2007) Selective increases of bifidobacteria in gut microflora improve high-fat-diet-induced diabetes in mice through a mechanism associated with endotoxaemia. (Translated from eng) *Diabetologia* 50(11):2374-2383 (in eng).
12. Rombeau JL & Kripke SA (1990) Metabolic and intestinal effects of short-chain fatty acids. (Translated from eng) *JPEN. Journal of parenteral and enteral nutrition* 14(5 Suppl):181S-185S (in eng).
13. Hamer HM*, et al.* (2008) Review article: the role of butyrate on colonic function. (Translated from eng) *Aliment Pharmacol Ther* 27(2):104-119 (in eng).
14. Salyers AA, Vercellotti JR, West SE, & Wilkins TD (1977) Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon. (Translated from eng) *Appl Environ Microbiol* 33(2):319-322 (in eng).
15. Cho KH & Salyers AA (2001) Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by Bacteroides thetaiotaomicron. (Translated from eng) *J Bacteriol* 183(24):7224-7230 (in eng).
16. D'Elia JN & Salyers AA (1996) Contribution of a neopullulanase, a pullulanase, and an alpha-glucosidase to growth of Bacteroides thetaiotaomicron on starch. (Translated from eng) *J Bacteriol* 178(24):7173-7179 (in eng).

17. Martens EC, Chiang HC, & Gordon JI (2008) Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. (Translated from eng) *Cell Host Microbe* 4(5):447-457 (in eng).
18. Xu J*, et al.* (2003) A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. (Translated from eng) *Science* 299(5615):2074-2076 (in eng).
19. Sonnenburg ED*, et al.* (2010) Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. (Translated from eng) *Cell* 141(7):1241-1252 (in eng).
20. Hehemann JH*, et al.* (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. (Translated from eng) *Nature* 464(7290):908-912 (in eng).
21. Martens EC, Koropatkin NM, Smith TJ, & Gordon JI (2009) Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. (Translated from eng) *J Biol Chem* 284(37):24673-24677 (in eng).
22. Jonas DA*, et al.* (2001) Safety considerations of DNA in food. (Translated from eng) *Ann Nutr Metab* 45(6):235-254 (in eng).
23. Thomas CM & Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. (Translated from eng) *Nat Rev Microbiol* 3(9):711-721 (in eng).
24. Krupovic M, Prangishvili D, Hendrix RW, & Bamford DH (2011) Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. (Translated from eng) *Microbiol Mol Biol Rev* 75(4):610-635 (in eng).
25. Bouma JE & Lenski RE (1988) Evolution of a bacteria/plasmid association. (Translated from eng) *Nature* 335(6188):351-352 (in eng).
26. Wozniak RA & Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. (Translated from eng) *Nat Rev Microbiol* 8(8):552-563 (in eng).
27. Hehemann JH, Kelly AG, Pudlo NA, Martens EC, & Boraston AB (2012) Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. (Translated from eng) *P Natl Acad Sci USA* 109(48):19786-19791 (in eng).
28. Correc G, Hehemann JH, Czjzek M, & Helbert W (2011) Structural analysis of the degradation products of porphyran digested by Zobellia galactanivorans beta-porphyranase A. (Translated from English) *Carbohyd Polym* 83(1):277-283 (in English).
29. Markowitz VM*, et al.* (2012) IMG: the Integrated Microbial Genomes database and comparative analysis system. (Translated from eng) *Nucleic Acids Res* 40(Database issue):D115-122 (in eng).
30. Lowe EC, Basle A, Czjzek M, Firbank SJ, & Bolam DN (2012) A scissor blade-like closing mechanism implicated in transmembrane signaling in a Bacteroides hybrid two-component system. (Translated from English) *P Natl Acad Sci USA* 109(19):7298-7303 (in English).
31. Geer LY*, et al.* (2010) The NCBI BioSystems database. (Translated from eng) *Nucleic Acids Res* 38(Database issue):D492-496 (in eng).
32. Moon K, Shoemaker NB, Gardner JF, & Salyers AA (2005) Regulation of excision genes of the Bacteroides conjugative transposon CTnDOT. (Translated from English) *J Bacteriol* 187(16):5732-5741 (in English).

33. Moon K, Sonnenburg J, & Salyers AA (2007) Unexpected effect of a Bacteroides conjugative transposon, CTnDOT, on chromosomal gene expression in its bacterial host. (Translated from eng) *Mol Microbiol* 64(6):1562-1571 (in eng).
34. Keeton CM*, et al.* (2013) The excision proteins of CTnDOT positively regulate the transfer operon. (Translated from eng) *Plasmid* 69(2):172-179 (in eng).
35. Bacteroides group Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/)
36. Parabacteroides group Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/)
37. Punta M*, et al.* (2012) The Pfam protein families database. (Translated from eng) *Nucleic Acids Res* 40(Database issue):D290-301 (in eng).
38. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. (Translated from eng) *BMC Bioinformatics* 5:113 (in eng).
39. Tamura K*, et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. (Translated from eng) *Mol Biol Evol* 28(10):2731-2739 (in eng).
40. Talavera G & Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. (Translated from eng) *Syst Biol* 56(4):564-577 (in eng).
41. Caspi R*, et al.* (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. (Translated from eng) *Nucleic Acids Res* 40(Database issue):D742-753 (in eng).
42. Muller D*, et al.* (2007) A tale of two oxidation states: bacterial colonization of arsenic-rich environments. (Translated from eng) *PLoS Genet* 3(4):e53 (in eng).
43. Hughes MF, Beck BD, Chen Y, Lewis AS, & Thomas DJ (2011) Arsenic exposure and toxicology: a historical perspective. (Translated from eng) *Toxicol Sci* 123(2):305-332 (in eng).
44. Whittle G, Shoemaker NB, & Salyers AA (2002) The role of Bacteroides conjugative transposons in the dissemination of antibiotic resistance genes. (Translated from eng) *Cell Mol Life Sci* 59(12):2044-2054 (in eng).
45. Martens EC, Roth R, Heuser JE, & Gordon JI (2009) Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *Journal of Biological Chemistry* e-pub ahead of print M109.008094.
46. Balch WE, Fox GE, Magrum LJ, Woese CR, & Wolfe RS (1979) Methanogens: reevaluation of a unique biological group. (Translated from eng) *Microbiol Rev* 43(2):260-296 (in eng).