



## OPEN ACCESS

## EDITED BY

Vasco Ariston De Carvalho Azevedo,  
Federal University of Minas Gerais, Brazil

## REVIEWED BY

Susu He,  
Dartmouth College, United States  
Silas Onyango Awuor,  
MOH-KISUMU County, Kenya

## \*CORRESPONDENCE

Patrick H. Degnan  
✉ patrick.degnan@ucr.edu

RECEIVED 19 June 2023

ACCEPTED 04 March 2024

PUBLISHED 26 March 2024

## CITATION

Ortañez J and Degnan PH (2024) Tracking and characterization of a novel conjugative transposon identified by shotgun transposon mutagenesis.

*Front. Microbiol.* 15:1241582.

doi: 10.3389/fmicb.2024.1241582

## COPYRIGHT

© 2024 Ortañez and Degnan. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Tracking and characterization of a novel conjugative transposon identified by shotgun transposon mutagenesis

Jericho Ortañez and Patrick H. Degnan\*

Department of Microbiology and Plant Pathology, University of California, Riverside, Riverside, CA, United States

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 *BoCTn* found in *Bacteroides ovatus* and a novel CTn, *PvCTn*, 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 *PvCTn in vitro*. Furthermore, our transcriptomics data revealed that BVU3433 plays a crucial role in the repression of *PvCTn* 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 *PvCTn*-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.

## KEYWORDS

*Phocaeicola vulgatus*, Bacteroidota, conjugative transposon, horizontal gene transfer, helix-turn-helix motif, ADP-ribosylglycohydrolase, Tn mutagenesis mobilization method

## 1 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 (Wexler, 2007; Smillie et al., 2010; Human Microbiome Project Consortium (HMP), 2012; Coyne et al., 2014; 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 (Akhter et al., 2012; Ozer et al., 2014; Roux et al., 2015; Johansson et al., 2020). 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) (Schlesinger et al., 2007; Waters and Salyers, 2013; Wood and Gardner, 2015). 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<sup>R</sup>) 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 ~0.6 kb *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 (Lampe et al., 1996; Kiljunen et al., 2017). 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 *PvCTn*. Further, this work provides insight into the regulatory mechanisms of *PvCTn*, most notably the helix-turn-helix motif encoding gene *BVU3433*. We also provide computational evidence for the prevalence of *PvCTn*-like elements among a panel of human gut-associated bacterial genomes and metagenomes from globally distributed patient samples. The observed mobilization of *PvCTn* and the widespread presence of genetically diverse *PvCTn*-like elements in Bacteroidota reinforce the importance of MGEs shaping the human microbiome.

## 2 Materials and methods

### 2.1 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<sub>2</sub>, 20% CO<sub>2</sub>, and 7% H<sub>2</sub>. *Escherichia coli* with the appropriate plasmid(s) (pSAM, pNBU2, pExchange, pLGB13) were grown on LB. All strains and plasmids are listed in Supplementary Table S1. 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 (*PvCTn*) 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 10 mL subcultures of LB at the following dilutions: 1:200, 1:500, and 1:750 dilutions. *Bacteroides 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 37°C. Bacteroidota were subcultured in an anaerobic chamber, stoppered, removed from the chamber, and incubated statically at 37°C. Growth was monitored and cells were pelleted (4,000 ×g for 5 min) when they reached an OD<sub>600</sub> of ~0.4. The supernatants were removed, and cells were resuspended with 1 mL TYG medium and pelleted again. After removing the supernatant, 1 mL 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 37°C for 24 h. 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:

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

### 2.2 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<sub>600</sub> of ~0.4, the cells were plated on TYG agar where 880 µM H<sub>2</sub>O<sub>2</sub> 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.

## 2.3 Transposon mutagenesis based identification of MGEs

### 2.3.1 Generating transposon (Tn) mutant libraries

*Bacteroides thetaiotaomicron*, *P. vulgatus*, and *B. ovatus* were mutagenized through conjugation with *E. coli* S17-1 containing the sequencing-adapted *mariner* transposon plasmid (pSAM) containing an erythromycin resistance cassette (Goodman et al., 2009). Conjugations were carried out as above, however, 100  $\mu$ L aliquots of the entire conjugation masses were plated on 50 BHI-HB plates supplemented with erythromycin and incubated anaerobically at 37°C for 48 h. Colonies were collected and pooled in TYG + 20% glycerol and stored at -80°C in 500  $\mu$ L aliquots.

### 2.3.2 Screening Tn libraries for MGEs

Aliquots of Tn mutant library donors (Erm<sup>R</sup>) were then conjugated with Tet<sup>R</sup> recipient strains of *B. thetaiotaomicron* VPI-5482, *B. thetaiotaomicron* 3731, *Parabacteroides merdae* ATCC 43184, *B. thetaiotaomicron* VPI-5482  $\Delta$ CPS, 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 h at 37°C, 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 37°C for 24 h.

### 2.3.3 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.

## 2.4 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 Supplementary Table S1. 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 and Comstock, 2019). We generated a ~2 kb 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 (Supplementary Table S1). 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*  $\Delta$ *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 20  $\mu$ g/mL FUdR or 100 ng/ $\mu$ L aTC were isolated and screened by PCR desired mutations.

## 2.5 Measuring growth kinetics of *Phocaeicola vulgatus*

Wild-type and mutant *P. vulgatus* were grown and washed in TYG media. The cells were normalized and diluted to OD<sub>600</sub> of 0.002 in TYG and dispensed in triplicate into a 96-well plate. Cell growth was measured every 30 min, over 36 h 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<sub>600</sub> ( $n = 3$ ).

## 2.6 RNA isolation and RNA-Seq

Samples for RNA-Seq of *P. vulgatus*, *P. vulgatus*  $\Delta$ *BVU3433*, and *B. thetaiotaomicron* PvCTn:*tetQ*:: $\Delta$ *BVU3433* were grown overnight in 5 mL TYG medium. Overnight cells were used to inoculate cultures at 10 mL TYG at a final dilution of 1:25 in biological triplicate. Cell growth was monitored and harvested at an OD<sub>600</sub> 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 h 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<sup>1</sup> 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 PvCTn when appropriate.

## 2.7 cDNA preparation, qPCR

*Phocaeicola vulgatus*  $\Delta$ *tdk* strains were grown overnight in 5 mL TYG medium. Overnight cells were used to inoculate cultures at 10 mL TYG at a final dilution of 1:25 in biological triplicate. Cell growth was monitored and harvested at an OD<sub>600</sub> of ~0.4. When appropriate, cells were perturbed by 880  $\mu$ M H<sub>2</sub>O<sub>2</sub> and immediately harvested after 30 min of incubation at 37°C. 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 1 N NaOH for 30 min at 65°C and neutralized with 1 N 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<sub>q</sub> method (Bookout et al., 2006) were used to process and calculate differences in *BVU3433* and *16S rRNA* (Supplementary Table S2) expression. *BVU3366* and *BVU3378* were used as candidate genes to confirm RNA-Seq expression profiles.

## 2.8 Identification of PvCTn family among gut Bacteroidota

The boundaries of PvCTn 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; Supplementary Table S3). This included functionally characterizing the genomic region using HMMER v3<sup>2</sup> with trusted cutoffs to search the PFAM v35 (El-Gebali et al., 2019) and TIGRFAM v15 (Haft et al., 2003) databases (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  $\geq 0.3$ ; Supplementary Table S3). The process to identify the boundaries of PvCTn 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  $\leq 0.0001$ ), filtered for  $\geq 20\%$  percent length aligned (PLA)<sup>3</sup> and then clustered using MCL (Inflation = 20) (Enright, 2002). Clusters were visualized in Cytoscape (Shannon et al., 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).

## 2.9 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 PvCTn 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').

## 2.10 Detection of PvCTn in human gut metagenomes

To identify the frequency of PvCTn-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 (Supplementary Table S3) 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 PvCTn-like elements were retrieved. Finally, the terminal 500 bp of PvCTn-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., 2017) from healthy patients and cohorts with colorectal cancer were retrieved from the NCBI SRA database (Supplementary Table S6). 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 and Salzberg, 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

1 <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

2 [hmmerr.org](http://hmmerr.org)

3 <https://doi.org/10.5281/zenodo.5138177>



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.

## 3 Results

### 3.1 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 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 1B). Some of these MGEs have been demonstrated to be functional, allowing them to act as positive controls (Reyes et al., 2013; Campbell et al., 2020; Frye et al., 2021).

The first step in the untargeted TMM (Figure 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<sup>R</sup> mutant library pools were then used as donors and mated with five Tet<sup>R</sup> 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<sup>R</sup> and Tet<sup>R</sup> colonies represent putative MGE transfer events (Figure 1C).

The success of the TMM 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 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 BoCTn (16/16), a CTn that encodes a vitamin B<sub>12</sub> 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 of the isolates (16/16) harbored a single newly acquired and uncharacterized *P. vulgatus* CTn (PvCTn).

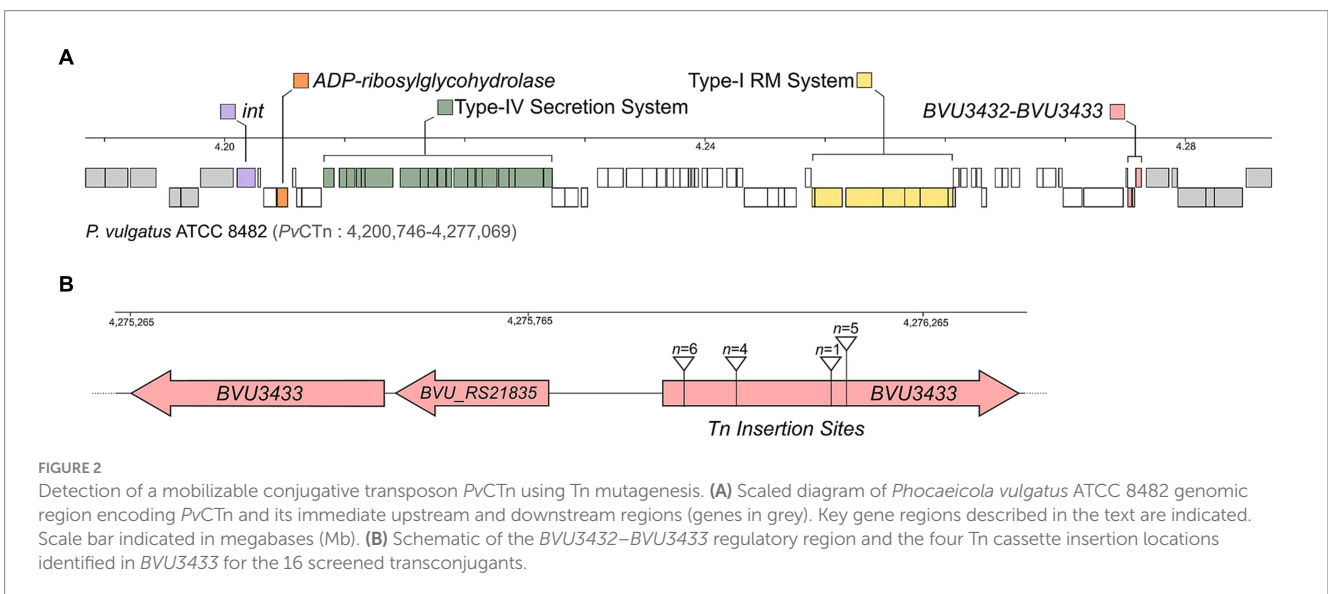
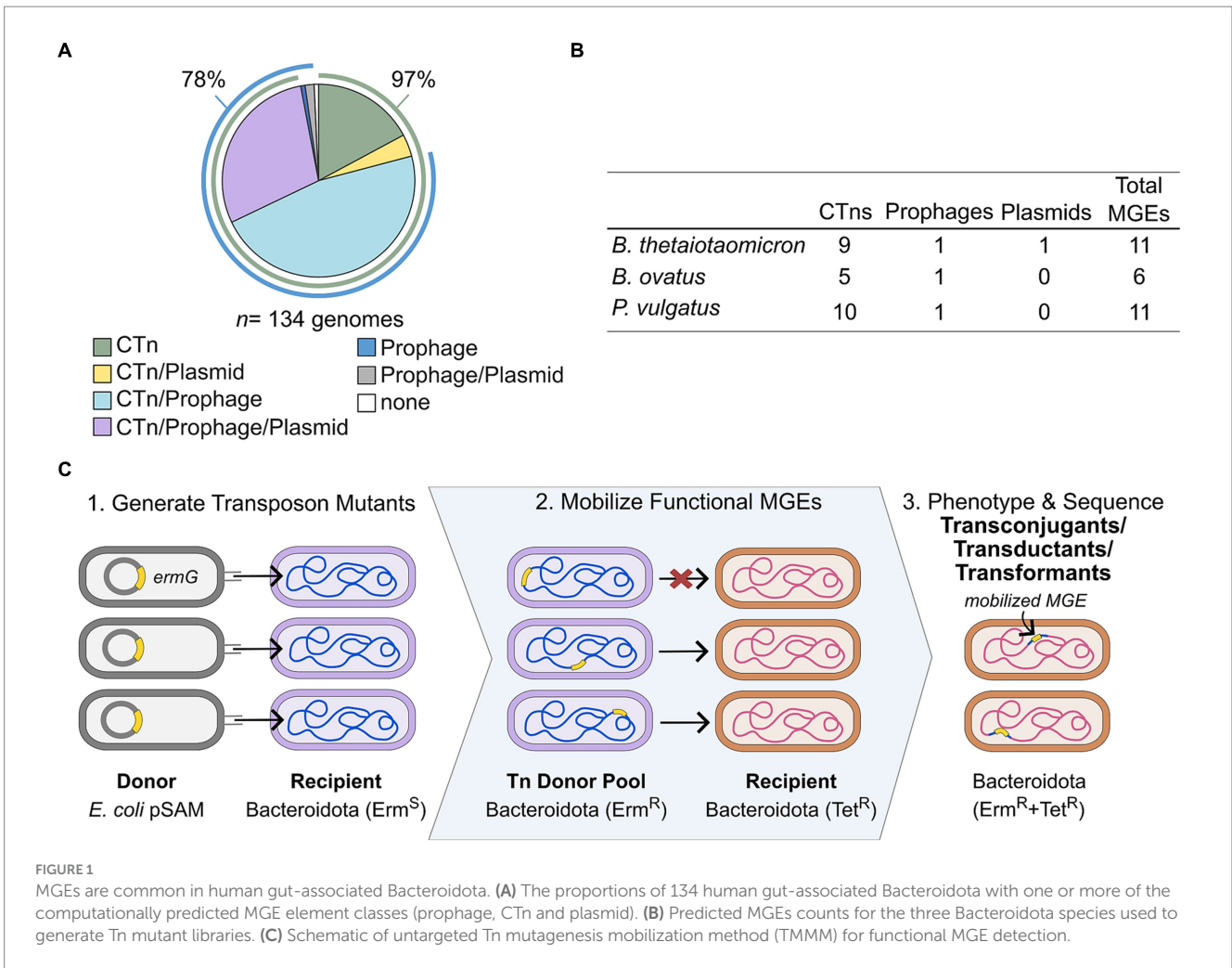
### 3.2 Fortuitous detection of a putative repressor of PvCTn activity

PvCTn is ~75 kb with 77 total genes including T4SS machinery, a putative ADP-ribosylglycohydrolase, and a Type I restriction-modification system (Figure 2A; Supplementary Table S2). Examination of the Tn insertion locations for the *B. thetaiotaomicron* VPI-5482 ΔCPS PvCTn 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) (Declerck et al., 2007; Kelley et al., 2015). 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 2B). Together, this suggests that the inactivation of BVU3433 through pSAM mutagenesis de-represses PvCTn conjugation genes, thereby increasing the frequency of PvCTn mobilization.

To directly test the function of BVU3433, we hypothesized that deleting BVU3433 from *P. vulgatus* should increase the conjugation efficiency of PvCTn. We first generated a clean deletion of BVU3433, followed by marking PvCTn in the wildtype and deletion background with an antibiotic resistance cassette using allelic exchange resulting in *P. vulgatus* ΔBVU3433 PvCTn:*tetQ* and *P. vulgatus* Δ*tdk* PvCTn:*tetQ*. We then carried out conjugations with these new donor strains using *B. thetaiotaomicron* VPI-5482 pNBU2\_ermG and *B. thetaiotaomicron* VPI-5482 ΔCPS pNBU2\_ermG as the recipient(s). The conjugation efficiencies of PvCTn ranged from  $1.71 \times 10^{-8}$  to  $2.11 \times 10^{-7}$  CFUs/mL for *P. vulgatus* PvCTn:*tetQ* and  $8.77 \times 10^{-7}$  –  $1.28 \times 10^{-6}$  CFUs/mL for *P. vulgatus* PvCTn:*tetQ* ΔBVU3433 depending on the recipient used (Figure 3A). Even though *P. vulgatus* ΔBVU3433 has a longer lag phase compared to WT (Supplementary Figure S1A), 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 3B). Regardless, the *P. vulgatus* ΔBVU3433 results recapitulate the phenotype detected with the initial Tn mutants.

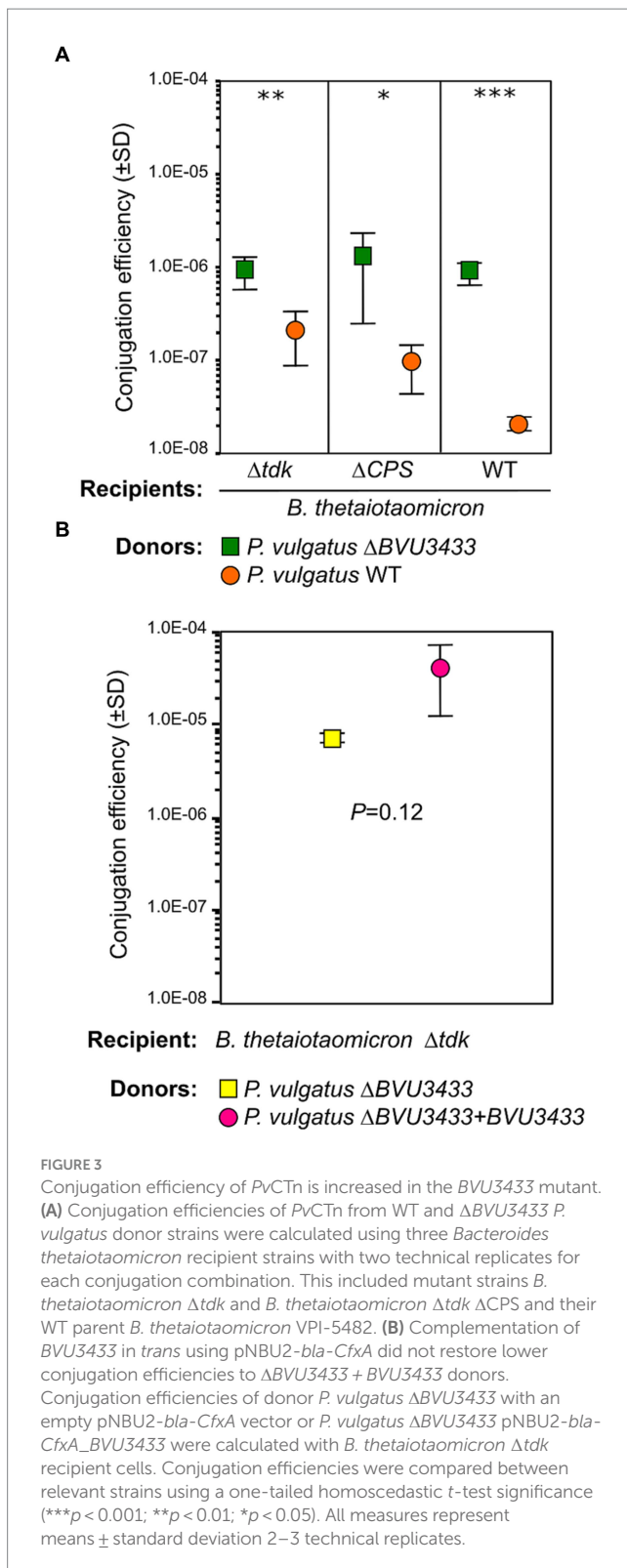
### 3.3 BVU3433 alters the PvCTn transcriptome

After observing the consistent increase in PvCTn conjugation efficiency of the BVU3433 mutant we investigated differences in PvCTn WT and ΔBVU3433 transcription through RNA-Seq. We hypothesized that the increase in the mutant PvCTn conjugation efficiency is due to BVU3433 no longer repressing the putative conjugation genes in PvCTn. Therefore, we expected the conjugation genes in PvCTn 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 PvCTn (Figure 4). In fact, 60% of PvCTn 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 PvCTn 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 (Supplementary Figure S2).

In addition to genes directly linked to the conjugation apparatus, we identified nine other PvcTn-encoded, putative regulatory genes that were differentially regulated by the absence of BVU3433 (Figure 4; Supplementary Table S4). 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 PvcTn 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 PvcTn 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 PvcTn:*tetQ*::ΔBVU3433 from *P. vulgatus* integrated into its genome. Remarkably, we determined that the expression profile of PvcTn:*tetQ*::ΔBVU3433 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 PvcTn 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* PvcTn:*tetQ*::ΔBVU3433 (Figure 4). The lack of conjugation gene expression needed to produce the T4SS structures encoded in operons 4 and 5 suggests that PvcTn:*tetQ*::ΔBVU3433 in a *B. thetaiotaomicron* recipient likely has reduced conjugation efficiency. Therefore, despite lacking BVU3433, the significant de-repression of genes observed in the mutant PvcTn 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 1B).

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 PvcTn 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 PvcTn in *P. vulgatus* (Figure 4A).

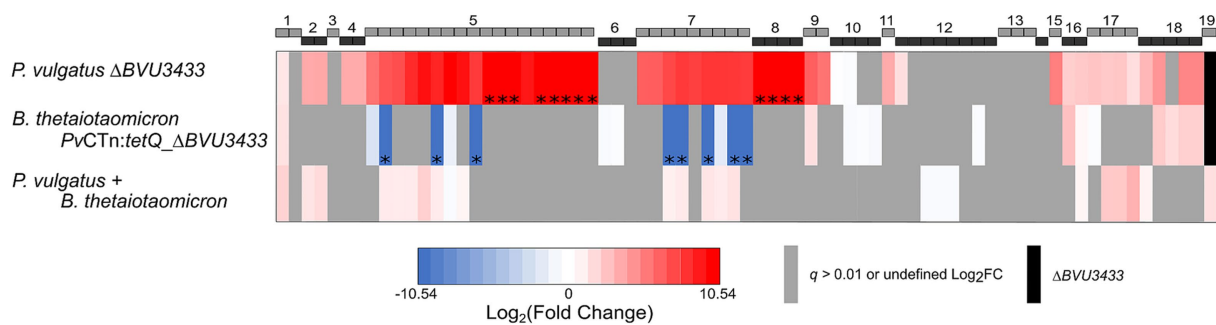


FIGURE 4

Loss of BVU3433 leads to the upregulation of PvCTn genes. Heatmap represents the log<sub>2</sub> fold change (log<sub>2</sub>FC) of RNA-Seq gene expression values for PvCTn genes from (i) *P. vulgatus*  $\Delta$ BVU3433, (ii) *B. thetaiotaomicron* PvCTn:tetQ\_ $\Delta$ BVU3433, and (iii) *P. vulgatus* WT x *B. thetaiotaomicron* when compared to *P. vulgatus* WT (only genes with a  $q < 0.01$  are shown, genes with  $q > 0.01$  and/or undefined log<sub>2</sub>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 Supplementary Table S2.

### 3.4 H<sub>2</sub>O<sub>2</sub> decreases BVU3433 expression and increases conjugation efficiency

Due to the observed increase in conjugation and de-repression of PvCTn genes in *P. vulgatus*  $\Delta$ 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<sub>2</sub>O<sub>2</sub>, salinity, antibiotics, taurocholic acid, heat, and UV. We discovered a significant ( $p = 0.015$ ) ~4.4-fold decrease in BVU3433 expression of H<sub>2</sub>O<sub>2</sub>-exposed *P. vulgatus* when compared to the control group (Figure 5A). Considering this decrease in BVU3433 expression, we tested how H<sub>2</sub>O<sub>2</sub>-exposure affects conjugation efficiency. We observed an insignificant ( $p = 0.42$ ) 1.73-fold increase in the conjugation efficiency of the H<sub>2</sub>O<sub>2</sub>-exposed WT *P. vulgatus* (Figure 5B) and the *P. vulgatus*  $\Delta$ BVU3433 strain followed the same trend with an insignificant ( $p = 0.36$ ) 1.45-fold increase in conjugation efficiency for the H<sub>2</sub>O<sub>2</sub>-exposed mutant (Figure 5B). Together, the qPCR and conjugation efficiency data suggest that subinhibitory H<sub>2</sub>O<sub>2</sub> stress can increase conjugation efficiency. However, since the conjugation efficiency of both WT and  $\Delta$ BVU3433 *P. vulgatus* increased with H<sub>2</sub>O<sub>2</sub> stress, it is possible that H<sub>2</sub>O<sub>2</sub> 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.

### 3.5 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 (Supplementary Table S5). 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 6A;

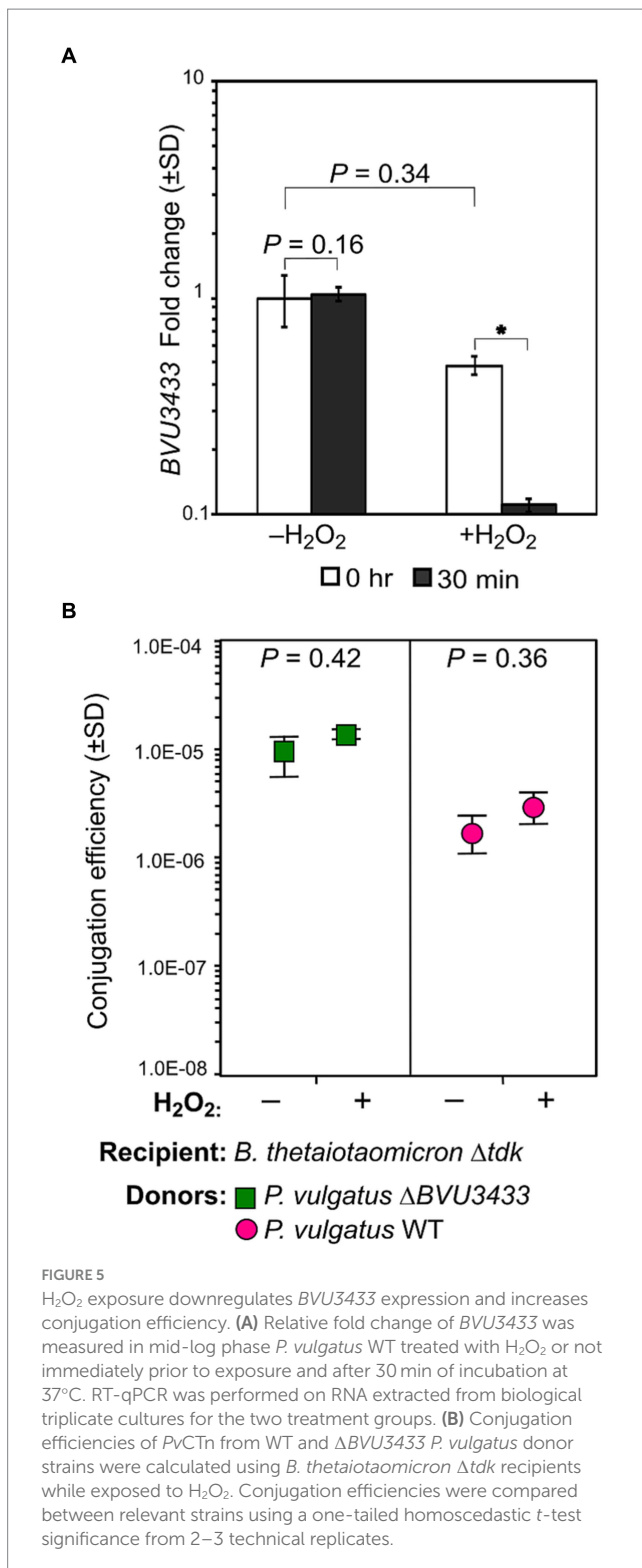
Supplementary Figure S3A). Like PvCTn, most encode restriction-modification systems, while others encode putative metal (tellurite) resistance genes (PF02342, PF05099, PF15616), UV protection genes (PF00817, PF11700) and/or protein phosphatases (PF13672).

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 *rhuA-rmuA-ppdK*. Specifically, we identified the *attB* as the 20 bp motif 5'-GYS GCN CAR GCK GCH RTH GA-3' within the 3' end of *ppdK* itself (Figure 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* PvCTn:tetQ: $\Delta$ BVU3433 transconjugant detected transcriptional readthrough across both the novel *attL* and *attR* sites (Supplementary Figure S3B). These data confirm the preference for this integration location by PvCTn. The conservation of the *attB* is somewhat surprising due to the sequence diversity observed among the predicted integrases (mean = 78.3%  $\pm$  9.9% amino acid identity; Supplementary Figure S3C). 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 (Supplementary Figure S3C).

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 7A). The divergent transcriptional organization of these HTH domain proteins is akin to that of *cl* and *cro* in phage Lambda 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 PvCTn. 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 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 7A; Supplementary Figure S3D). The analysis also identified 2 conserved motifs with imperfect dyad

symmetry that we observed in all five upstream regions (Figures 7B,C; Supplementary Figure S3D). It is possible that one or both of these motifs act as binding sites for *BVU3433* or another *PvCTn* encoded regulator (e.g., *BVU3432*) and be responsible for activation or repression of the conjugative apparatus.

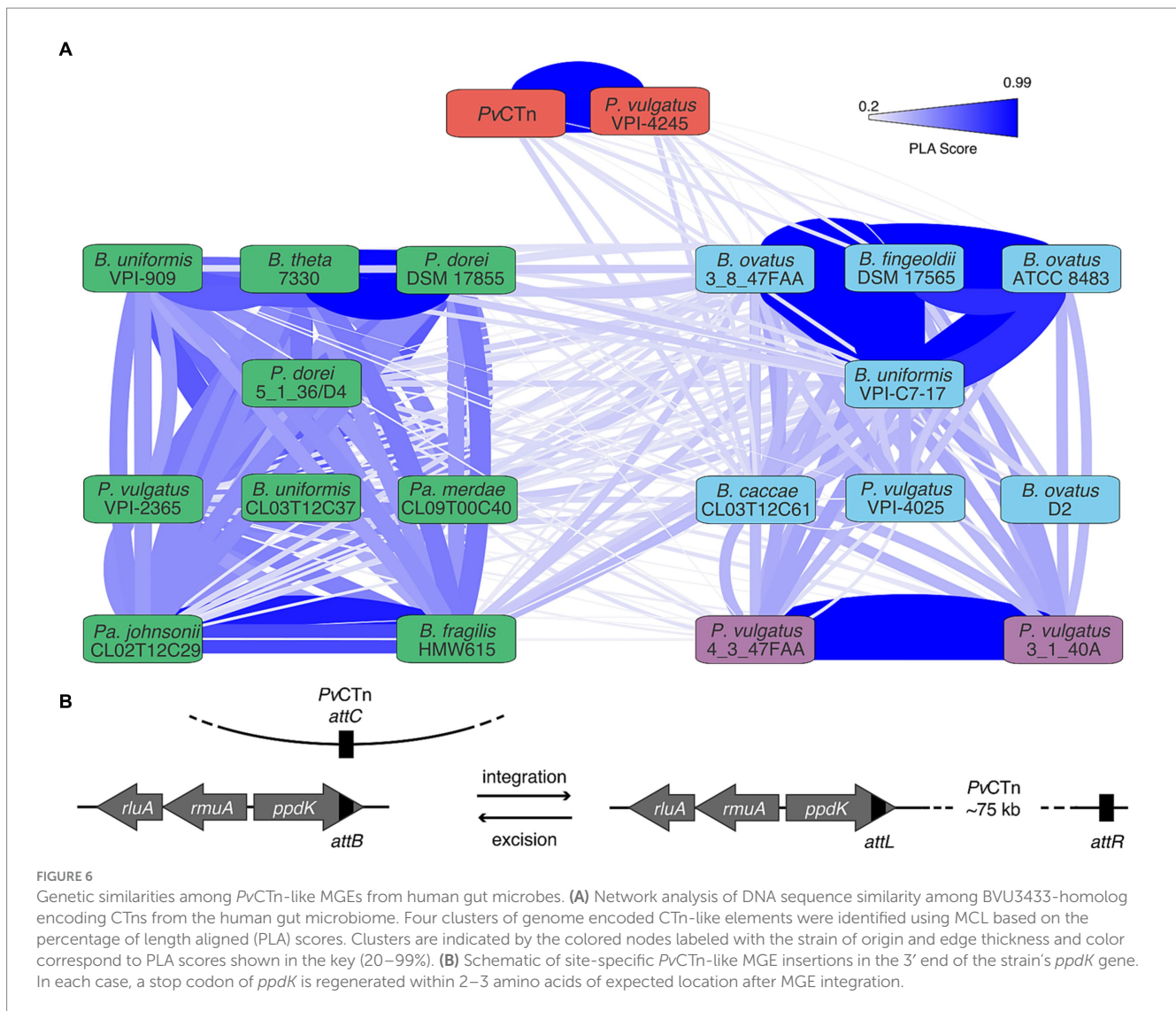
### 3.6 Detection of *PvCTn* in human gut metagenomes

Using a marker gene approach, evidence for *PvCTn*-like elements was detected among Bacteroidota species from both healthy patients and those with colorectal cancer (CRC) in 3 geographically distinct studies (Austria, United States, China) (Figure 8A; Supplementary Figures S4A,B). We found sequence evidence for *PvCTn*-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 *PvCTn*-like elements (e.g., ~7–12% *attL* and ~2–7% *attR*; Figure 8A; Supplementary Figures S4A,B). 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 (Supplementary Figure S4C). Further, despite the previously described differences in community composition including a greater proportion of Bacteroidota cells (Figure 8B) and immunological status between the healthy patients and those with colorectal cancer (Feng et al., 2015; Vogtmann et al., 2016; Yu et al., 2017) we saw no difference in the frequency of *PvCTn*-like elements after normalizing for Bacteroidota abundance (Figure 8A).

Analysis of simulated datasets suggests reliable detection of *PvCTn* when strains with integrated elements account for ≥1% of the population, however, when using the described marker genes and *PvCTn* encoding strains approach 0.1% detection becomes unpredictable (~50:50; Supplementary Figure S4C). 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 (Supplementary Figure S4C). 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 *PvCTn*-like MGE integration events).

## 4 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 TMM as a novel approach for tracking functional MGEs in Bacteroidota (Figure 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<sub>NDM</sub>*) 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.

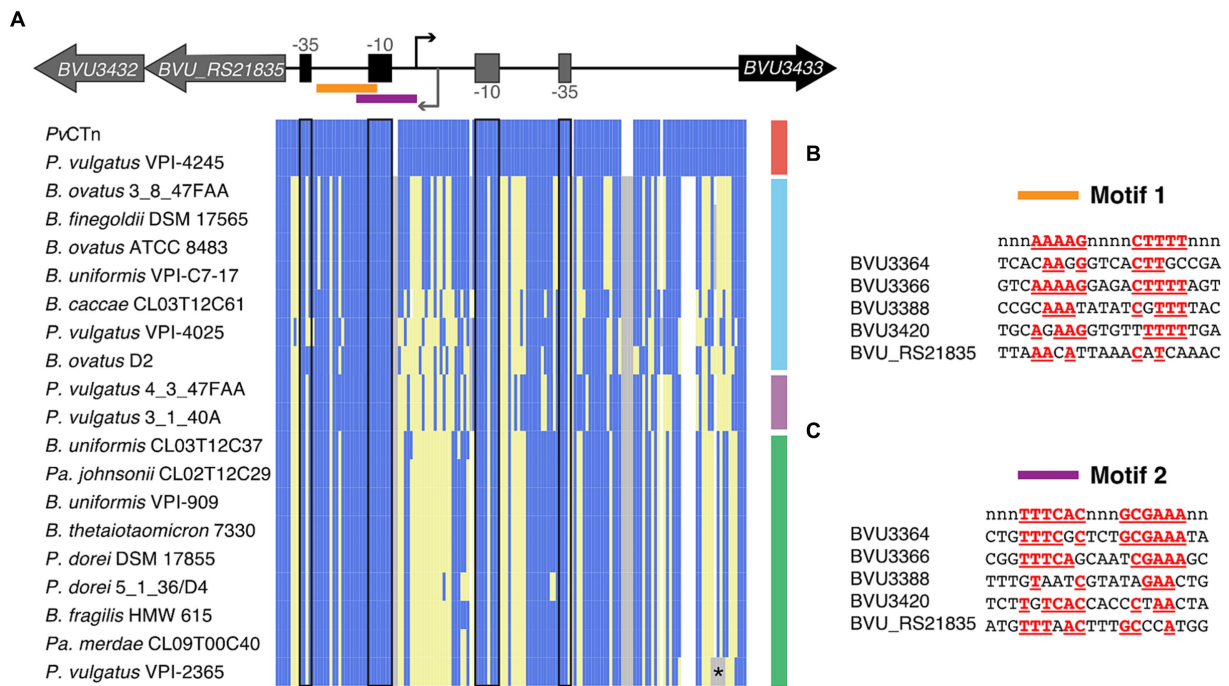
#### 4.1 Constraints of existing approaches for identifying functional MGEs

Computational methods are effective at identifying potential MGEs in bacterial genomes (Akhter et al., 2012; Ozer et al., 2014; Roux et al., 2015; Johansson et al., 2020). However, predictions alone cannot confirm the ability of an MGE to mobilize, which often

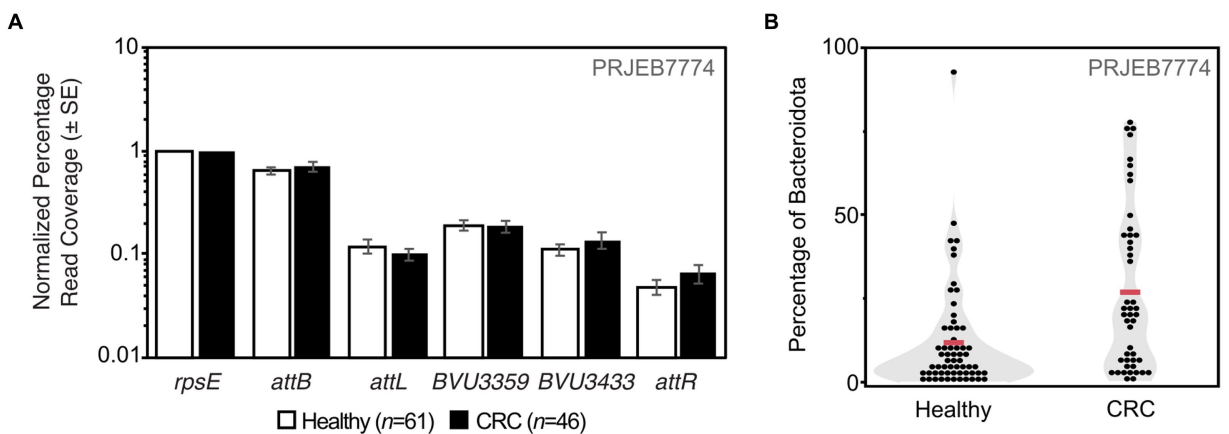
requires targeted mutations to confirm or the reliance on capturing chance transfer events in genomic data (Coyné et al., 2014). Our untargeted MGE capture method bypasses the need for directed mutations (Saak et al., 2020). Although a previous study developed transposon-aided capture (TRACA), which allows for the preferential capture of circular extrachromosomal MGEs (i.e., plasmids) (Jones and Marchesi, 2006), it is unlikely to capture elements like PvCTn. On the other hand, entrapment vectors like pBACpAK (Tansirichaiya et al., 2022), if adapted for Bacteroidota hosts, may be able to capture PvCTn. 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.

#### 4.2 Tn mutagenesis mobilization method provides insight into CTn regulation

In this study, we successfully identified the mobilization of *BoCTn* and *PvCTn* using this Tn mutagenesis method. Curiously, all the *PvCTn* transconjugants harbored pSAM-mediated disruptions of



**FIGURE 7**  
 PvCTn-like MGEs encode conserved regulatory regions. **(A)** Sequence alignment of BVU\_RS21835–BVU3433 intergenic region from all 20 PvCTn-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 PvCTn, 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,C)** Alignments of two conserved motifs with dyad symmetry identified by MEME and found in the upstream regions of five PvCTn operons that are differentially expressed in the ΔBVU3433 strain and broadly conserved in PvCTn-like MGEs. Majority rule consensus sequence is shown on the top line in red and matching positions are shown for each upstream region.



**FIGURE 8**  
 Metagenome detection of PvCTn-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).

BVU3433. Our subsequent analyses demonstrated a consistent increase in the conjugation efficiency of *P. vulgatus* ΔBVU3433 mutants compared to WT. While additional Tn insertions likely occurred in T4SS genes or activators of PvCTn 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 BoCTn which encodes a vitamin B<sub>12</sub> 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 multi-layer, tight regulation (Salysers et al., 1995; Johnson and Grossman, 2015). 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 and Salysers, 2013). Our results identify the critical role of BVU3433 in repressing many P<sub>v</sub>CTN genes and as a result significantly reducing mobilization. Yet our findings also indicate that BVU3433 is likely only the first component of the P<sub>v</sub>CTN regulatory cascade, as nine other P<sub>v</sub>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 P<sub>v</sub>CTN remain to be elucidated. Given the broad upregulation of P<sub>v</sub>CTN  $\Delta$ BVU3433 genes compared to WT P<sub>v</sub>CTN (Figure 4), we hypothesize that BVU3433 binds to one or more P<sub>v</sub>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 P<sub>v</sub>CTN, operon 2 of P<sub>v</sub>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 P<sub>v</sub>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 P<sub>v</sub>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 P<sub>v</sub>CTN's regulatory mechanism, which is likely to be complex and involve additional layers of transcriptional and translational regulation.

### 4.3 P<sub>v</sub>CTN-like elements are diverse and globally distributed

Our characterization of P<sub>v</sub>CTN and BVU3433 enabled our detection of a diverse group of P<sub>v</sub>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 P<sub>v</sub>CTN-like elements in human gut metagenomes (Feng et al., 2015; Vogtmann et al., 2016; Yu et al., 2017), and found they were widespread in ~90% of patient samples. Together these results suggest P<sub>v</sub>CTN-like elements contribute to genetic and functional diversity of human gut microbes.

## 5 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 P<sub>v</sub>CTN functionality and identify a conserved conjugation repressor protein. Further, P<sub>v</sub>CTN represents one member of a diverse family of elements that can be detected in patient samples from around the globe. Further, these P<sub>v</sub>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 P<sub>v</sub>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 P<sub>v</sub>CTN-like elements in human gut-associated bacteria.

## Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NCBI BioProject, accession PRJNA983822.

## Author contributions

JO: conceptualization, methodology, lab work, manuscript writing and reviewing, and data analysis. PD: conceptualization, methodology, developed software, manuscript writing and reviewing, and data analysis. All authors contributed to the article and approved the submitted version.

## Funding

This research was supported by initial complement funding to PD from UCR. JO was supported in part by a Higher Education Emergency Relief Fund (HEERF) Fellowship.

## Acknowledgments

We thank Ansel Hsiao and Jason Stajich for useful feedback on the research involved in this manuscript. We would also like to acknowledge Danielle Campbell for the PLA code and BvCTN clustering strategy.



## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2024.1241582/full#supplementary-material>

## 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 *bla*NDM. *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. 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
- 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–W49. doi: 10.1093/nar/gkv416
- 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.* 73, 15.8.1–15.8.28. doi: 10.1002/0471142727.mb1508s73
- 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
- Coyne, M. J., Zitomersky, N. L., McGuire, A. M., Earl, A. M., and Comstock, L. E. (2014). Evidence of extensive DNA transfer between Bacteroidales species within the human gut. *mBio* 5, e01305–e01314. doi: 10.1128/mBio.01305-14
- 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
- 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. *Proc. Natl. Acad. Sci. USA* 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., Barry, N. A., Mok, K. C., Taga, M. E., and Goodman, A. L. (2014). Human gut microbes use multiple transporters to distinguish Vitamin B<sub>12</sub> analogs and compete in the gut. *Cell Host Microbe.* 15, 47–57. doi: 10.1016/j.chom.2013.12.007
- Dempwolff, F., Sanchez, S., and Kearns, D. B. (2020). TnFLX: a third-generation mariner-based transposon system for *Bacillus subtilis*. *Appl. Environ. Microbiol.* 86:e02893-19. doi: 10.1128/AEM.02893-19
- 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. *Proc. Natl. Acad. Sci. USA* 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. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30:1575–1584. doi: 10.1093/nar/30.7.1575
- 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
- Frye, K. A., Piamthai, V., Hsiao, A., and Degnan, P. H. (2021). Mobilization of vitamin B<sub>12</sub> transporters alters competitive dynamics in a human gut microbe. *Cell Rep.* 37:110164. doi: 10.1016/j.celrep.2021.110164
- García-Bayona, L., and Comstock, L. E. (2019). Streamlined genetic manipulation of diverse *Bacteroides* and *Parabacteroides* isolates from the human gut microbiota. *mBio* 10, e01762–e01719. 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
- 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
- Human Microbiome Project Consortium (HMP) (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. doi: 10.1038/nature11234
- 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
- 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
- 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
- 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
- Lampe, D. J., Churchill, M. E., and Robertson, H. M. (1996). A purified mariner transposase is sufficient to mediate transposition in vitro. *EMBO J.* 15, 5470–5479. doi: 10.1002/j.1462-2075.1996.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
- 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

- McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumbly, 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
- 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
- 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
- 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
- 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. (2023). Toxin-linked mobile genetic elements in major enteric bacterial pathogens. *Gut Microbiome* 4:E5. doi: 10.1017/gmb.2023.2
- Partridge, S. R., Kwong, S. M., Neville, F., and Jensen, S. O. (2018). Mobile genetic elements associated with antimicrobial resistance. *Clin. Microbiol. Rev.* 31:e00088-17. doi: 10.1128/CMR.00088-17
- 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
- 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 USA* 110:20236–20241. doi: 10.1073/pnas.1319470110
- 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. (1995). 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
- 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
- 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
- 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., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., and de la Cruz, F. (2010). Mobility of plasmids. *Microbiol. Mol. Biol. Rev.* 74, 434–452. doi: 10.1128/MMBR.00020-10
- Tansirichaiya, S., Goodman, R. N., Guo, X., Bulgasim, I., Samuelsen, Ø., Al-Haroni, M., et al. (2022). Intracellular transposition and capture of Mobile genetic elements following intercellular conjugation of multidrug resistance conjugative plasmids from clinical Enterobacteriaceae isolates. *Microbiol. Spectr.* 10, e02140–e02121. doi: 10.1128/spectrum.02140-21
- 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
- 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
- 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–e00513. doi: 10.1128/mBio.00569-13
- 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
- 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
- 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