Article

Bacteroidales-Specific Antimicrobial Genes Can Influence the Selection of the Dominant Fecal Strain of *Bacteroides vulgatus* and *Bacteroides uniformis* from the Gastrointestinal Tract Microbial Community

Hyunmin Koo [1],* and Casey D. Morrow [2],*

[1] Department of Genetics, Hugh Kaul Precision Medicine Institute, University of Alabama at Birmingham, Birmingham, AL 35294, USA
[2] Department of Cell, Developmental and Integrative Biology, Hugh Kaul Precision Medicine Institute, University of Alabama at Birmingham, Birmingham, AL 35294, USA
* Correspondence: khmkhm87@uab.edu (H.K.); caseym@uab.edu (C.D.M.)

Abstract: *Bacteroides vulgatus* and *Bacteroides uniformis* are known to be abundant in the human fecal microbial community. Although these strains typically remain stable over time in humans, disruption of this microbial community following antibiotics resulted in the transient change to new strains suggesting that a complex, dynamic strain community exists in humans. To further study the selection of dominant fecal microbial strains from the gastrointestinal tract (GIT) community, we analyzed three longitudinal metagenomic sequencing data sets using BLAST+ to identify genes encoding *Bacteroidales*-specific antimicrobial proteins (BSAP) that have known functions to restrict species-specific replication of *B. uniformis* (BSAP-2) or *B. vulgatus* (BSAP-3) and have been postulated to provide a competitive advantage in microbial communities. In the HMP (Human Microbiome Project) data set, we found fecal samples from individuals had *B. vulgatus* or *B. uniformis* with either complete or deleted BSAP genes that did not change over time. We also examined fecal samples from two separate longitudinal data sets of individuals who had been given either single or multiple antibiotics. The BSAP gene pattern from most individuals given either single or multiple antibiotics recovered to be the same as the pre-antibiotic strain. However, in a few individuals, we found incomplete BSAP-3 genes at early times during the recovery that were replaced by *B. vulgatus* with the complete BSAP-3 gene, consistent with the function of the BSAP to specifically restrict *Bacteroides* spp. The results of these studies provide insights into the fluxes that occur in the *Bacteroides* spp. GIT community following perturbation and the dynamics of the selection of a dominant fecal strain of *Bacteroides* spp.

Keywords: *Bacteroidales*-specific antimicrobial genes; antibiotics; microbial community; fecal microbial strain

1. Introduction

Numerous longitudinal studies have used metagenomic sequencing coupled with new informatics approaches to establish the dominant fecal microbial community at the strain level is unique to the individual and generally stable over time [1–4]. However, the stability of dominant fecal microbial strains can be influenced by gastrointestinal tract (GIT) perturbations, such as antibiotics, that result in the transient appearance of new strains, although this can vary between individuals [5–7]. Usually after a short time, the new fecal strains are replaced with the previous dominant strains supporting the existence of a dynamic GIT ecosystem with multiple microbial strains that compete for fecal dominance [7].

In a recent study, Roodgar et al. showed that microbial strains found in the feces were probably stably maintained in GIT reservoirs defined by metabolic or spatial niches [8].

How these specific strains are maintained in these niches is unknown. The *Bacteroides vulgatus* and *Bacteroides uniformis* are known to be abundant in the human fecal microbial community [1,9,10]. We have previously shown that dominant fecal *B. vulgatus* or *B. uniformis* strains are unique to the individual and related to each other as determined from longitudinal strain tracking analysis [2,4,7,11,12]. Microbes are known to encode a variety of proteins that have the potential to provide a competitive advantage in a complex ecosystem [13–16]. Previous studies have identified the *Bacteroidales*-specific antimicrobial proteins (BSAPs) that have known functions to restrict species-specific replication and colonization of *B. uniformis* (BSAP-2) and *B. vulgatus* (BSAP-3) [9,15]. BSAPs have been shown to encode membrane attack complex/perforin (MACPF) signatures in proteins that lyse cells by pore formation and target the predominant LPS specific for *B. uniformis* (BSAP-2) or *B. vulgatus* (BSAP-3) [17]. The presence of these microbial genes in an individual's *Bacteroides* spp. might then provide a competitive advantage to maintain that strain in the GIT ecosystem [14].

Based on the known functions of the BSAP genes, the objective of our study is to assess a possible role in the selection of the fecal dominant strain. We have determined the status of these genes in *B. vulgatus* or *B. uniformis* in fecal samples from studies that contained longitudinal metagenomic data sets from normal and antibiotic-treated individuals. The Integrative Genomics Viewer (IGV) has been used to visualize the genomic region (5′ and 3′ sequences) containing BSAP genes to determine the status of the BSAP genes in the individuals. Our study establishes the existence and long-term stability of the BSAP gene in individuals. Additionally, analysis of the data set from two studies using different antibiotics suggests a role for the BSAP genes in the selection of dominant fecal strains following the disruption of the GIT microbial ecosystem.

2. Materials and Methods

2.1. Publicly Available Data Sets Were Used in This Study

In this study, we used 3 publicly available metagenomic data sets for healthy individuals (1) from the NIH Human Microbiome Project (HMP, https://portal.hmpdacc.org/, accessed on 13 June 2012) [17], (2) pre and post treated with a single antibiotic (cefprozil) from Raymond et al. [18] (accession number: PRJEB8094), and (3) pre and post treated with 3 antibiotics (meropenem, gentamicin, and vancomycin) from Palleja et al. (accession number: ERP022986) [19]. The sequence data that we used in this study was fully anonymized before being made publicly available. For all three studies, longitudinal fecal samples were all taken from normal individuals with no known diseases. Further information on the characteristics of the individuals can be found in the primary references. For the HMP data set, 30 individual samples that were previously used to establish our window-based single-nucleotide variant (SNV) similarity (WSS) analysis were selected to run the analysis [2,17]. For Raymond et al., fecal samples were collected from 18 individuals at 3 different time points: pre-treatment (Day 0), end of antibiotic (cefprozil) treatment (Day 7), and 3 months post-treatment (Day 90), and we selected all individuals' samples to run the analysis. We have also additionally selected 6 individual fecal samples that did not receive antibiotic treatment as controls. For the Palleja data set, fecal samples from 12 individuals treated with 3 antibiotics (meropenem, gentamicin, and vancomycin) were collected at 5 different time points: pre-treatment (Day 0), immediately after antibiotics treatment (Day 4), and 3 post-treatment time points (Day 8, 42, and 180). From this data set, we have selected all 12 individual samples at 5 different time points (pre and post treatment (Day 4, 8, 42, and 180) to run the analysis.

2.2. Analysis of Bacteroides BSAP Genes

Before running the analysis, quality control steps included removing any human reference genome (hg19) using bowtie2 (version 2.3.4.3) with default parameters [20], and filtering low-quality reads (sliding window of 50 bases having a QScore < 20) using Trimmomatic (version 0.36) [21]. Each fecal metagenomics sample was used to align with

Bacteroides vulgatus CL09T03C04 for BSAP-3 and *Bacteroides uniformis* CL03T00C23 for BSAP-2 using the Burrows–Wheeler aligner program (BWA; version 0.7.13) BWA tool [22]. Aligned reads from each reference genome were then sorted and indexed using SAMtools (version 0.1.19) [23]. The resultant bam file was converted to FASTQ format using BEDTools (version 2.26.0) [24]. Each converted FASTQ file was then assembled using MEGAHIT, and the resultant contig file was selected for BLASTX search against the BSAP-2 and BSAP-3 sequence reads using BLAST+ [15,25]. To visualize aligned Bacteriocin genes (BSAP-2 and BSAP-3) for paired samples for each individual, the VCF file generated for *B. vulgatus* and *B. uniformis* of each sample was uploaded to the Integrative Genomics Viewer (IGV) and aligned to their reference genome [16]. For the BSAP-3, the region of nucleotides located between nucleotides 2,178,016 to 2,183,973 of *B. vulgatus* was selected to display 5′ and 3′ sequences that also include BSAP-3 genes. For the BSAP-2, nucleotide regions of *B. uniformis* were selected to show 5′ and 3′ sequences that also include BSAP-2 genes (location: 1,383,872–1,385,398). To further characterize the BSAP negative phenotype, nucleotide sequences of the selected 5′ and 3′ regions were extracted using IGV and then uploaded to Jalview for visualization [16,26].

2.3. Analysis of *Bacteroides* spp. *Strain Relatedness Using WSS*

To determine the relatedness of strains, individual paired samples were additionally used for the Window-based single-nucleotide variant (SNV) similarity (WSS) analysis which was previously developed based on the Human Microbiome Project (HMP) data set [2]. The resultant WSS score was used to compare against the cut-off value that was previously established in our previous study (related strain pair: WSS score > cut-off; unrelated strain pair: WSS score < cut-off) (Supplementary Material) [2,27,28].

3. Results

Previous studies have reported the identification of the BSAP genes that are specific for *B. vulgatus* (BSAP-3) or *B. uniformis* (BSAP-2) [13,29]. For BSAP-3, protein signatures identifying MACPF that had activity to restrict growth in a species-specific manner of *B. vulgatus* were identified in gene cluster 16 from B. vulgatus/B. dorei. Similarly, in *B. uniformis*, MACPF was also identified as a protein encoded in gene cluster 15 that could restrict species-specific replication of *B. uniformis,* which was named BSAP-2. To determine the presence of the gene encoding BSAP-3 or BSAP-2, the individual samples from three longitudinal data sets used in this study were used for the gene analysis of clusters 15 and 16 using BLAST+ [15] and IGV tools [16].

We first analyzed the human microbiome (HMP) data set that is composed of a longitudinal study of paired samples taken at separate times (up to 1 year apart) [1,17]. We selected 30 paired samples that we have previously shown to have related strains of *B. vulgatus* and *B. uniformis* over periods between 3 months and 1 year [2]. We screened for the complete genes encoding BSAP-3 for *B. vulgatus* and BSAP-2 for *B. uniformis* in each sample pair (Supplementary Table S1). We identified 11 sample pairs out of 30 for the complete gene encoding BSAP-3 and 7 of 28 samples for the gene encoding BSAP-2 (2 sample pairs, S64 and S70, had WSS scores below the cutoff value for *B. uniformis*) (Figure 1). For 30 BSAP-3 and 28 BSAP-2 sample pairs, we found the same BSAP (either positive or negative) for time points taken at times that were different by 3–6 months.

To further characterize the BSAP genes that were incomplete (i.e., negative), we used the IGV to display the 5′ and 3′ sequences bracketing the BSAP deletion [16] (Supplementary Figure S1A,B). We found that 4 of the 19 sample pairs for the deleted BSAP-3 gene were identical (S28, S31, S48, and S66) and 10 of the 19 sample pairs (S16, S17, S21, S23, S49, S60, S64, S65, S69, and S70) had a few nucleotide differences bracketing the deletion from the comparison of the two (a and b) time points (Figures 2A,B and S1A). Similarly, for BSAP-2 from the HMP, we found 2 of the 20 (S21 and S66) with deleted BSAP-2 had identical 5′ and 3′ sequences and 18 of the 20 sample pairs (S5, S10, S14, S16, S17, S19, S20, S23, S24, S28, S29, S31, S48, S51, S57, S60, S61, and S65) had only a few nucleotide differences in 5′

and 3′ sequences bracketing the deletion from the comparison of the a and b time points (Figures 3A,B and S1B).

HMP	*B. vulgatus* CL09T03C04		*B. uniformis* CL03T00C23	
Samples	BSAP-3	No.	BSAP-2	No.
A a	positive	11	positive	7
A b	positive		positive	
B a	negative	19	negative	21
B b	negative		negative	

Figure 1. BSAP summary for HMP data set. The number of individual pairs that (A) both pairs were BSAP-2 and BSAP-3 complete (denoted as positive) and (B) both pairs were BSAP-2 and BSAP-3 incomplete (denoted as negative).

Figure 2. BSAP-3 gene alignment for HMP data set. Individual pairs that both pairs had BSAP-3 negative were selected to extract aligned sequence reads that included the BSAP-3 gene along with 5′ and 3′ of the BSAP-3 gene (20 base pairs for each 5′ and 3′ end). (**A**) Both individual pairs had exacted 5′ and 3′ sequence reads; (**B**) minor differences (less than 5 nucleotides) were observed between individual pairs; Major differences (more than 5 nucleotides) were detected between individual pairs.

From a further inspection of the BSAP negative phenotype sample sets with IGV, we found a subset of individuals for BSAP-3 (S5, S14, S18, S62, and S63) and one individual (S49) for BSAP-2 that lacked a complete BSAP gene but did have a gene sequence fragment in the BSAP gene region. The BSAP-3 in individual S14 was particularly interesting since the *B. vulgatus* in the two samples were related as determined by WSS, with the first sample containing a deleted BSAP-3 while the second, later sample had an incomplete BSAP-3 gene (Supplementary Figure S1A). We also found that the pattern of the incomplete BSAP genes in these individuals was unique to that sample and thus differed between the two sample times. Thus, these results establish for both *B. vulgatus* and *B. unformis* that the BSAP patterns were generally stable over an extended time (3–6 months) although we did find some individuals with incomplete BSAP genes that had a few nucleotide differences between the two time points.
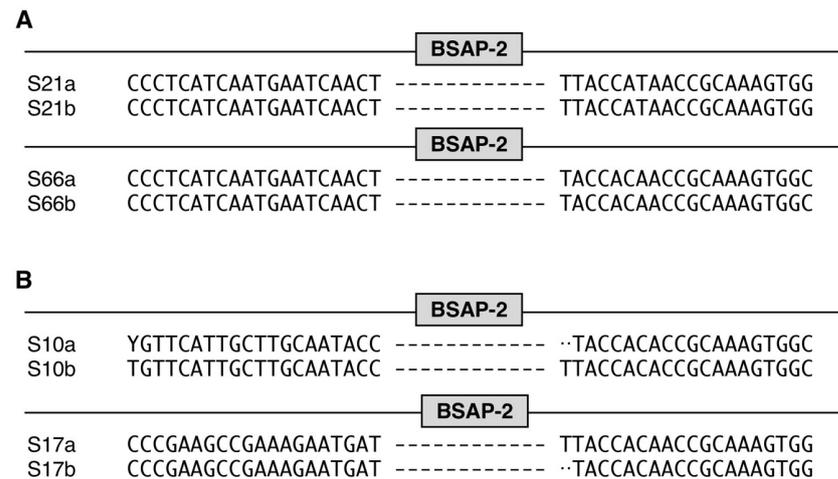
A

```
                                    ┌─────────┐
                                    │ BSAP-2  │
─────────────────────────────────────┤         ├─────────────────────────────
S21a    CCCTCATCAATGAATCAACT  ----------- TTACCATAACCGCAAAGTGG
S21b    CCCTCATCAATGAATCAACT  ----------- TTACCATAACCGCAAAGTGG
                                    ┌─────────┐
                                    │ BSAP-2  │
─────────────────────────────────────┤         ├─────────────────────────────
S66a    CCCTCATCAATGAATCAACT  ----------- TACCACAACCGCAAAGTGGC
S66b    CCCTCATCAATGAATCAACT  ----------- TACCACAACCGCAAAGTGGC
```

B

```
                                    ┌─────────┐
                                    │ BSAP-2  │
─────────────────────────────────────┤         ├─────────────────────────────
S10a    YGTTCATTGCTTGCAATACC  ----------- ··TACCACACCGCAAAGTGGC
S10b    TGTTCATTGCTTGCAATACC  ----------- TTACCACACCGCAAAGTGGC
                                    ┌─────────┐
                                    │ BSAP-2  │
─────────────────────────────────────┤         ├─────────────────────────────
S17a    CCCGAAGCCGAAAGAATGAT  ----------- TTACCACAACCGCAAAGTGG
S17b    CCCGAAGCCGAAAGAATGAT  ----------- ··TACCACAACCGCAAAGTGG
```

Figure 3. BSAP-2 gene alignment for HMP data set. Individual pairs that both pairs had deleted BSAP-2 genes were selected to extract aligned sequence reads that included the BSAP-2 gene along with 5' and 3' of the BSAP-2 gene (20 base pairs for each 5' and 3' end). (**A**) Both individual pairs had exacted 5' and 3' sequence reads; (**B**) minor differences (less than 5 nucleotides) were observed between individual pairs; Major differences (more than 5 nucleotides) were detected between individual pairs. "Y" and "S" are degenerate nucleotide code: "Y" indicates C/T and "S" indicates C/G.

We next wanted to determine the BSAP pattern following perturbation and recovery of the gut ecosystem. For this analysis, we analyzed a data set from a previous study that analyzed gut microbes following a single antibiotic treatment [7,18]. The Raymond et al., data set contained the collection of fecal samples from six individuals at day 0, day 7 and day 90 that were used as untreated controls (Supplemental Table S2). Analysis of the BSAP genes from these samples found that three of the six had complete BSAP-3 genes (i.e., positive), while two of the six were positive for BSAP-2 (Supplementary Table S2 and Supplementary Figure S2A,B). The analysis of the samples from individuals with no antibiotics in the Raymond et al. study gave similar results as that found for the analysis of the HMP data set (Supplementary Table S2 and Supplementary Figure S2A,B).

We next selected 17 individual sample sets from the Raymond et al. study that were given cephalexin. The samples were collected one day after treatment and 82 days after treatment (a total of 90 days after the pre-sample). We had previously analyzed this data set using our strain tracking analysis to show recovery of the pre-antibiotic strain at the day 90 samples for most individuals. Analysis of BSAP genes from the pre-sample demonstrated that 3 out of 17 of the BSAP-3 and 4 out of 17 of the BSAP-2 were complete (Figure 4). In most of the sample sets the BSAP phenotype did not change following antibiotics. We did find an incomplete BSAP-3 gene in one individual (P15) at day 0 and following the antibiotics at day 7, but by day 90 we found a complete BSAP gene (Figure 4A, Supplementary Figure S3A, and Supplementary Table S2). In another individual (P17), we found a new strain on day 7 that had an incomplete BSAP-3 gene, but on day 90 the strain was related to the pre-sample and also had a complete BSAP-3 gene. Finally, one individual of note, P4, had an incomplete BSAP-3 gene at days 0 and 7; however, the same *B. vulgatus* strain at day 90 had a complete BSAP-3 gene. This individual was the only example where we found a BSAP phenotype change from negative to positive without a strain change (Supplementary Figure S3A). The presence of the complete BSAP-3 gene in the dominant fecal strain of these two individuals that had incomplete genes at early times during the recovery is consistent with the function of the BSAP to specifically restrict *Bacteroides* spp. that are without complete BSAP genes.

A Sample	Day 0, 7 vs. 90	BSAP-3 Day 0, 7, 90
P5	Related, Related	**pos, pos, pos**
P10	Related, Related	**pos, pos, pos**
P14	Related, Related	**incomplete, pos, pos**
P4	Related, Related	neg, neg, **pos**
P9	Related, Related	neg, neg, neg
P13	Related, Related	neg, neg, neg
P19	Related, Related	neg, neg, neg
P20	Related, Related	neg, neg, neg
P21	Related, Related	neg, neg, neg
P15	Related, Related	**incomplete, incomplete, pos**
P17	Related, Not related	**pos, incomplete, pos**
P2	Related, Not related	neg, neg, neg
P3	Related, Not related	neg, neg, neg
P18	Related, Not related	neg, neg, neg
P11	Related, Not related	**incomplete, incomplete, incomplete**
P1	Not related, Related	neg, **incomplete, incomplete**
P22	No Score	*No score*

B Sample	Day 0, 7 vs. 90	BSAP-2 Day 0, 7, 90
P14	Related, Related	**pos, pos, pos**
P9	Related, Related	neg, neg, neg
P10	Related, Related	neg, neg, neg
P13	Related, Related	neg, neg, neg
P15	Related, Related	neg, neg, neg
P18	Related, Related	neg, neg, neg
P20	Related, Related	neg, neg, neg
P3	Related, Not related	**pos, pos, pos**
P2	Related, Not related	neg, neg, neg
P11	Related, Not related	neg, neg, neg
P17	Related, Not related	neg, neg, neg
P21	Related, Not related	neg, neg, neg
P22	Related, No score	**pos**, *no score*, **pos**
P1	Not related, Not related	**pos, pos, pos**
P19	Not related, Not related	neg, neg, neg
P4	Not related, Related	neg, neg, neg
P5	No score	*No score*

Figure 4. BSAP summary table for Raymond et al. data set. Gene analysis that includes WSS and BLAST was conducted for the Raymond et al. data set. For each sample, strain relatedness was determined by WSS analysis, and the presence/absence of (**A**) BSAP-3 and (**B**) BSAP-2 genes were determined by BLAST analysis. The "pos" indicates the complete BSAP-3/BSAP-2 gene was observed; "incomplete" indicates a sequence fragment of BSAP-3/BSAP-2 gene was observed; "neg" indicates BSAP-3/BSAP-2 gene was not observed; "No score" indicates that no score was observed from WSS analysis.

We next analyzed the BSAP sequence changes using IGV [16]. For the BSAP-3 negative samples, we found eight samples in which the strains that were related at Day 0 and Day 90 had identical or few nucleotide differences in the 5′ and 3′ regions bracketing the BSAP-3 deletion (P2, P3, P9, P13, P18, P19, P20, and P21) supporting the recovery of the pre-antibiotic strain (Figures 4A and S3A). For the samples without BSAP-2 genes, 10 samples that had no BSAP-2 genes for the Day 0 and Day 90 samples all had identical or few nucleotide differences in the 5′ and 3′ regions bracketing the deleted BSAP gene (P2, P9, P10, P11, P13, P15, P17, P18, P20, and P21) (Figures 4B and S3B). Collectively, these results are consistent with the resiliency of the GIT strain reservoir to recover following perturbation.

To investigate the impact of a greater perturbation on the strain reservoir, we characterized the strain stability of individuals from the Palleja et al. data set that had been treated with a suppressive antibiotics cocktail consisting of three antibiotics (meropenem, gentamicin, and vancomycin) [7,19]. In this longitudinal study, we found that in some instances the treatment resulted in a strain change from the pre-antibiotic strain. Overall, 2 of 10 individuals showed complete BSAP-3 genes and 3 of 10 showed complete BSAP-2 genes (Figure 5, Supplementary Figure S4, and Supplementary Table S3). We found several examples where there was a change in an individual from an incomplete BSAP gene to a complete BSAP gene. Individual P1 had a strain change early after antibiotics that were BSAP-3 negative, but the same samples at later times (42 and 180 days) were the same strain as the pre-strain and both had a complete BSAP-3 gene (Figure 5A). Another individual (P4) that had an incomplete BSAP gene on day 0, had a strain change, and the new strains at day 42 and 180 both had complete BSAP-3 genes (Figure 5A). Similarly, in this individual, for *B. uniformis* there was also a new strain at days 42 and 180 that both had complete BSAP-2 genes (Figure 5A,B). Finally, for P11, there was no BSAP-2 change in the strain of *B. uniformis*, but they had a complete BSAP-2 gene (Figure 5B).

A

Sample	Day 0, 7, 42 vs.180	BSAP-3 Day 0, 8, 42,180
P1	R, NR, R	**pos**, neg, **pos, pos**
P4	NR, NS, R	**incomplete**, *no score*, **pos, pos**
P6	NR, NR, R	**incomplete**, neg, **incomplete**, neg
P8	R, NR, R	neg, neg, neg, **incomplete**
P3	NR, R, R	neg, neg, neg, neg
P5	NR, NS, R	neg, *no score*, neg, neg
P9	NR, NS, R	neg, *no score*, neg, neg
P11	R, NS, NR	neg, *no score*, neg, neg
P12	R, NR, R	neg, neg, neg, neg
P10	NS, NS ,NS	*No score*

B

Sample	Day 0, 7, 42 vs.180	BSAP-2 Day 0, 8, 42,180
P4	NR, NS, R	**pos**, *no score*, **pos, pos**
P10	NR ,NS, R	**pos**, *no score*, **pos, pos**
P11	R, NS, R	**incomplete**, *no score*, **pos, pos**
P6	NS, NS, NR	*no score, no score*, neg, **incomplete**
P8	NR, NS, NR	neg, *no score*, neg, neg
P5	NR, NS, R	neg, *no score*, neg, neg
P9	R, NS, R	neg, *no score*, neg, neg
P3	R, NR, NR	neg, neg, neg, neg
P12	NR, NR, R	neg, neg, neg, neg
P1	NS, NS, NS	*No Score*

Figure 5. BSAP summary table for Palleja et al. data set. Gene analysis that includes WSS and BLAST was conducted for the Palleja et al. data set. For each sample, strain relatedness was determined by WSS analysis and the presence/absence of the (**A**) BSAP-3 and (**B**) BSAP-2 genes were determined by BLAST analysis. "R" indicates related strain, "NR" indicates not related strain observed, "NS" indicates no WSS score observed from WSS analysis. "pos" indicates BSAP-3/BSAP-2 gene was observed; "incomplete" indicates sequence fragment of BSAP-3/BSAP-2 gene was observed; "neg" indicates BSAP-3/BSAP-2 gene was not observed; "No score" indicates that no score was observed from WSS analysis.

From IGV analysis for BSAP-3 in B. vulgatus, we found that only 1 of 10 samples (P1) had a complete BSAP-3 gene at days 0, 42, and 180 (Figure 5A). We did find two individuals, P11 and P12, where the *B. vulgatus* strain at Day 0 was related to the strain at Day 180. In the case of P11, the 5′ and 3′ regions at Day 0 and Day 180 showed that fecal dominant strains were different, whereas the 5′ and 3′ sequences from P12 were the same at Day 0 and 180 (Figures 5A and S4A). For B uniformis, individuals P3 and P9 had related strains at Day 0 and 180 and both the 5′ and 3′ sequences were the same (Figures 5B and S4B).

4. Discussion

In this study, we have analyzed the presence of genes encoding BSAP proteins in fecal *B. vulgatus* or *B, uniformis* strains from individuals with normal and perturbed GIT ecosystems. Our analysis combines the identification of BSAP genes (BSAP-2 and BSAP-3) with BLAST and the use of IGV tools to provide a comprehensive analysis of the fecal dominant *Bacteroides* spp. strains. Our studies support that the fecal dominant *Bacteroides* spp. strain is most probably maintained in a microbial strain reservoir within the GIT and, in some individuals, the selection of the dominant strain could be influenced by the BSAP.

Previous studies have shown the *Bacteroidales*-specific antimicrobial genes function in a species-specific fashion to restrict replication and subsequent colonization [13,29]. The analysis of the status of the BSAP genes then provides an opportunity to investigate possible interactions between *Bacteroides* spp. that occur in the GIT. For example, the presence of the BSAP genes would provide a selective advantage over the *Bacteroidales* microbe without BSAP in the GIT. To further define the presence or absence of the BSAP gene in the population, we first focused on the well-characterized HMP data set that contained paired samples from individuals taken at two different times, 3 months to 1 year apart [1,17]. Using WSS analysis, we had previously determined that, for the samples collected, the strains of *B. vulgatus* and *B. uniformis* of an individual were related to each other over this period [2]. Consistent with other studies, we found most individuals were either BSAP positive or negative [13,29]. From our examination of the longitudinal samples, we found that the BSAP phenotypes were stable over time in the unperturbed, normal GIT microbial ecosystem.

In most individuals, the *Bacteroides* spp. strain and BSAP gene were stable following treatment with single or multiple antibiotics. We did find specific instances where the BSAP changed during recovery from the antibiotic. For example, in one individual, P17 of the Raymond et al. study, we found that the cephalexin treatment resulted in the appearance of a new *B. vulgatus* strain on Day 7 where an incomplete BSAP-3 gene was replaced by the original *B. vulgatus* strain on day 90 that had a complete BSAP-3 gene. Individual P4 of the same Raymond et al. study had *B. vulgatus,* a strain with no BSAP-3 gene, at days 0 and 7, but had the *B. vulgatus* strain with a complete BSAP-3 gene at day 90. Similarly, for individual P1 of the *Palleja* et al. samples, the BSAP-3 changed from complete to incomplete from day 0 to day 7 post antibiotics. The Day 7 *B. vulgatus* strain was replaced by the pre-antibiotic BSAP-3 positive *B. vulgatus* strain at days 42 and 180. Collectively, these results support that in some individuals the complete BSAP gene in the GIT strain reservoir influenced the selection of the dominant fecal strain.

We used the IGV tool to visualize the BSAP genes and surrounding 5′ and 3′ regions. For those samples that were BSAP positive, all had complete BSAP genes although we did find that in some instances regions 5′ and 3′ had discrete deletions. In contrast, in IGV analysis of the samples with incomplete BSAP-3 and BSAP-2 genes, we found that for most of the samples the deletions encompassed the BSAP genes and surrounding 5′ and 3′ regions. Surprisingly, we found different individuals that had shared sequence deletions (Figures 2 and 3). The sharing between different individuals suggests a common mechanism for the generation of deletions. One possibility would be that the deletions might be a result of a transposition event to delete the BSAP gene and surrounding DNA regions, although we did not find the consensus sequence for a known *Bacteroides* conjugative transposon (CTnDOT) [30,31]. Previous studies have reported that the *Bacteroides* genome region containing the BSAP gene was included in a region of integrative and conjugative elements which included numerous, yet undefined, putative transposases [32,33]. We also found that a few individuals with incomplete BSAP gene patterns were different between the sampling time points. Although the origin of these BSAP gene patterns is unknown, one possibility is that they occurred during a non-typical transposition process that left remnants of the BSAP gene. Environmental stress such as antibiotic treatment, nutritional changes, or even disease have been shown to enhance transposition [34–36]. Further understanding of the dynamics of the appearance of these deleted BSAP genes in an individual over time will be needed to determine if a correlation exists with microbial ecosystems that were or are currently under stress.

5. Conclusions

As discussed by Wexler and Goodman, the *Bacteroides* spp. are an ideal group in which to uncover fundamental principles for the persistence of microbes over time in the host [9]. The analysis of the BSAP gene following antibiotic treatment lends support for a possible role in the persistence and recovery of the dominant fecal strain of *Bacteroides* spp. Since the human microbiome is recognized as being diverse and generally individual specific, a strength of our study was the use of three longitudinal data sets that allowed the comparison of samples from the same individual pre and post treatment. Additional follow up studies with clinical implications would include an analysis of patients pre and post fecal transplant and those individuals that have had perturbation of their GIT ecosystem from oral drugs [37–39].

References

1. Schloissnig, S.; Arumugam, M.; Sunagawa, S.; Mitreva, M.; Tap, J.; Zhu, A.; Waller, A.; Mende, D.R.; Kultima, J.R.; Martin, J.; et al. Genomic variation landscape of the human gut microbiome. *Nature* **2013**, *493*, 45–50. [CrossRef] [PubMed]
2. Kumar, R.; Yi, N.; Zhi, D.; Eipers, P.; Goldsmith, K.T.; Dixon, P.; Crossman, D.K.; Crowley, M.R.; Lefkowitz, E.J.; Rodriguez, J.M.; et al. Identification of donor microbe species that colonize and persist long term in the recipient after fecal transplant for recurrent Clostridium difficile. *NPJ Biofilms Microbiomes* **2017**, *3*, 12. [CrossRef] [PubMed]
3. Truong, D.T.; Tett, A.; Pasolli, E.; Huttenhower, C.; Segata, N. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res.* **2017**, *27*, 626–638. [CrossRef]
4. Koo, H.; Hakim, J.A.; Crossman, D.K.; Lefkowitz, E.J.; Morrow, C.D. Sharing of gut microbial strains between selected individual sets of twins cohabitating for decades. *PLoS ONE* **2019**, *14*, e0226111. [CrossRef] [PubMed]
5. Dethlefsen, L.; Huse, S.; Sogin, M.L.; Relman, D.A. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol.* **2008**, *6*, e280. [CrossRef] [PubMed]
6. Rashid, M.-U.; Zaura, E.; Buijs, M.J.; Keijser, B.J.F.; Crielaard, W.; Nord, C.E.; Weintraub, A. Determining the long-term effect of antibiotic administration on the human normal intestinal microbiota using culture and pyrosequencing methods. *Clin. Infect. Dis.* **2015**, *60*, S77–S84. [CrossRef] [PubMed]
7. Koo, H.; Hakim, J.A.; Crossman, D.K.; Kumar, R.; Lefkowitz, E.J.; Morrow, C.D. Individualized recovery of gut microbial strains post antibiotics. *NPJ Biofilms Microbiomes* **2019**, *5*, 30. [CrossRef]
8. Roodgar, M.; Good, B.H.; Garud, N.R.; Martis, S.; Avula, M.; Zhou, W.; Lancaster, S.M.; Lee, H.; Babveyh, A.; Nesamoney, S.; et al. Longitudinal linked-read sequencing reveals ecological and evolutionary responses of a human gut microbiome during antibiotic treatment. *Genome Res.* **2021**, *31*, 1433–1446. [CrossRef]
9. Wexler, A.G.; Goodman, A.L. An insider's perspective: Bacteroides as a window into the microbiome. *Nat. Microbiol.* **2017**, *2*, 17026. [CrossRef]
10. Kraal, L.; Abubucker, S.; Kota, K.; Fischbach, M.A.; Mitreva, M. The prevalence of species and strains in the human microbiome: A resource for experimental efforts. *PLoS ONE* **2014**, *9*, e97279. [CrossRef]
11. Koo, H.; Crossman, D.K.; Morrow, C.D. Strain Tracking to Identify Individualized Patterns of Microbial Strain Stability in the Developing Infant Gut Ecosystem. *Front. Pediatr.* **2020**, *8*, 549844. [CrossRef] [PubMed]
12. Koo, H.; McFarland, B.C.; Hakim, J.A.; Crossman, D.K.; Crowley, M.R.; Rodriguez, J.M.; Benveniste, E.N.; Morrow, C.D. An individualized mosaic of maternal microbial strains is transmitted to the infant gut microbial community. *R. Soc. Open Sci.* **2020**, *7*, 192200. [CrossRef] [PubMed]
13. McEneany, V.L.; Coyne, M.J.; Chatzidaki-Livanis, M.; Comstock, L.E. Acquisition of MACPF domain-encoding genes is the main contributor to LPS glycan diversity in gut Bacteroides species. *ISME J.* **2018**, *12*, 2919–2928. [CrossRef] [PubMed]
14. Matano, L.M.; Coyne, M.J.; García-Bayona, L.; Comstock, L.E. Bacteroidetocins Target the Essential Outer Membrane Protein BamA of Bacteroidales Symbionts and Pathogens. *mBio* **2021**, *12*, e02221–e02285. [CrossRef] [PubMed]
15. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [CrossRef]
16. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [CrossRef]
17. Methé, B.A.; Nelson, K.E.; Pop, M.; Creasy, H.H.; Giglio, M.G.; Huttenhower, C.; Gevers, D.; Petrosino, J.F.; Abubucker, S.; Badger, J.H.; et al. A framework for human microbiome research. *Nature* **2012**, *486*, 215.

18. Raymond, F.; Ouameur, A.A.; Déraspe, M.; Iqbal, N.; Gingras, H.; Dridi, B.; Leprohon, P.; Plante, P.-L.; Giroux, R.; Bérubé, È.; et al. The initial state of the human gut microbiome determines its reshaping by antibiotics. *ISME J.* **2016**, *10*, 707–720. [CrossRef] [PubMed]

19. Palleja, A.; Mikkelsen, K.H.; Forslund, S.K.; Kashani, A.; Allin, K.H.; Nielsen, T.; Hansen, T.H.; Liang, S.; Feng, Q.; Zhang, C.; et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nat. Microbiol.* **2018**, *3*, 1255–1265. [CrossRef] [PubMed]

20. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357. [CrossRef]

21. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]

22. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [CrossRef]

23. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]

24. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [CrossRef] [PubMed]

25. Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **2015**, *31*, 1674–1676. [CrossRef] [PubMed]

26. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.; Clamp, M.; Barton, G.J. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [CrossRef] [PubMed]

27. Kumar, R.; Grams, J.; Chu, D.I.; Crossman, D.K.; Stahl, R.; Eipers, P.; Goldsmith, K.; Crowley, M.; Lefkowitz, E.J.; Morrow, C.D. New microbe genomic variants in patients fecal community following surgical disruption of the upper human gastrointestinal tract. *Hum. Microbiome J.* **2018**, *10*, 37–42. [CrossRef]

28. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.1–11.10.33. [CrossRef] [PubMed]

29. Roelofs, K.G.; Coyne, M.J.; Gentyala, R.R.; Chatzidaki-Livanis, M.; Comstock, L.E. Bacteroidales secreted antimicrobial proteins target surface molecules necessary for gut colonization and mediate competition in vivo. *mBio* **2016**, *7*, e01055-16. [CrossRef]

30. Salyers, A.A.; Shoemaker, N.B.; Stevens, A.M.; Li, L.-Y. Conjugative transposons: An unusual and diverse set of integrated gene transfer elements. *Microbiol. Rev.* **1995**, *59*, 579–590. [CrossRef]

31. Cheng, Q.; Paszkiet, B.J.; Shoemaker, N.B.; Gardner, J.F.; Salyers, A.A. Integration and excision of a Bacteroides conjugative transposon, CTnDOT. *J. Bacteriol.* **2000**, *182*, 4035–4043. [CrossRef] [PubMed]

32. Wozniak, R.A.; Waldor, M.K. Integrative and conjugative elements: Mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.* **2010**, *8*, 552–563. [CrossRef] [PubMed]

33. Coyne, M.J.; Zitomersky, N.L.; McGuire, A.M.; Earl, A.M.; Comstock, L.E. Evidence of extensive DNA transfer between bacteroidales species within the human gut. *mBio* **2014**, *5*, e01305–e01314. [CrossRef] [PubMed]

34. Capy, P.; Gasperi, G.; Biémont, C.; Bazin, C. Stress and transposable elements: Co-evolution or useful parasites? *Heredity* **2000**, *85*, 101–106. [CrossRef] [PubMed]

35. Twiss, E.; Coros, A.M.; Tavakoli, N.P.; Derbyshire, K.M. Transposition is modulated by a diverse set of host factors in Escherichia coli and is stimulated by nutritional stress. *Mol. Microbiol.* **2005**, *57*, 1593–1607. [CrossRef]

36. Maharjan, R.P.; Ferenci, T. A shifting mutational landscape in 6 nutritional states: Stress-induced mutagenesis as a series of distinct stress input–mutation output relationships. *PLoS Biol.* **2017**, *15*, e2001477. [CrossRef] [PubMed]

37. Koo, H.; Morrow, C.D. Perturbation of the human gastrointestinal tract microbial ecosystem by oral drugs to treat chronic disease results in a spectrum of individual specific patterns of extinction and persistence of dominant microbial strains. *PLoS ONE* **2020**, *15*, e0242021. [CrossRef] [PubMed]

38. Koo, H.; Morrow, C.D. Time series strain tracking analysis post fecal transplantation identifies individual specific patterns of fecal dominant donor, recipient, and unrelated microbial strains. *PLoS ONE* **2022**, *17*, e0274633. [CrossRef]

39. Koo, H.; Morrow, C.D. Identification of donor *Bacteroides vulgatus* genes encoding proteins that correlate with early colonization following fecal transplant of patients with recurrent Clostridium difficile. *Sci. Rep.* **2023**, *13*, 14112. [CrossRef]