Quorum Sensing System in *Bacteroides thetaiotaomicron* Strain Identified by Genome Sequence Analysis

Zhi Cheng Wu, Hong Xin Feng, Lin Wu,* Meng Zhang, and Wei Lan Zhou

**ABSTRACT:** This study is a bioinformatics assay on the microbial genome of *Bacteroides thetaiotaomicron*. The study focuses on the problem of quorum sensing as a result of adverse factors such as chemotherapy and antibiotic therapy. In patients with severe intestinal diseases, two strains of microorganisms were identified that were distinguished as new. Strains were investigated by conducting genome sequencing. The current concepts concerned with the quorum sensing system regulation by stationary-phase sigma factor and their coregulation of target genes in *B. thetaiotaomicron* were considered. The study suggested using bioinformatics data for the diagnosis of gastrointestinal disorders. In the course of the study, 402 genes having a greater than twofold change were identified. Biological pathways (KEGG pathways) were classified into the following categories: cellular processes, environmental information processing, genetic information processing, human disease, metabolism, and organismic systems. Among notable changes in the biofilm population observed in parallel to the planktonic *B. thetaiotaomicron* was the expression of genes in the polysaccharide utilization loci that were involved in the synthesis of O-glycans.

**INTRODUCTION**

Microbocenoses can be found in the environment, inside and outside of living organisms. Recently, scientists have established a connection between the intestinal microbiome of a healthy person and specific diseases.¹ Genetic studies play a leading role in establishing a correlation between intestinal microbiota composition, certain mutations, and diseases.² The latest research is aimed at developing sensory systems for the microbiota composition, certain mutations, and diseases.² The study suggested using bioinformatics data for the diagnosis of gastrointestinal disorders. In the course of the study, 402 genes having a greater than twofold change were identified with the 95% confidence level. The shortest and longest coding genes were predicted; the noncoding genes were detected. Biological pathways (KEGG pathways) were classified into the following categories: cellular processes, environmental information processing, genetic information processing, human disease, metabolism, and organismic systems. Among notable changes in the biofilm population observed in parallel to the planktonic *B. thetaiotaomicron* was the expression of genes in the polysaccharide utilization loci that were involved in the synthesis of O-glycans.

Acyl homoserine lactones (N-AHLs). Some authors discuss the possibilities of using QS for therapeutic purposes.⁶ In the fight for resources, bacteria secrete QS inhibitors, which are small extracellular molecules (peptides, fatty acids, ketones, adrenaline, norepinephrine, etc.), blocking the action of autoinducers and inhibiting the action of QS signaling molecules.⁷ The widely studied QSI are antagonistic peptides from natural resources, in particular polyphenols that can be found in tea, honey, garlic, cloves, marine organisms, and fungi. The synthetic QS-inhibiting substances include the preparations of 5-fluorouracil, azithromycin, poly(ethylene glycol), furanone derivatives, thiazolidinone derivatives, etc.⁸ Bacteria constantly modulate gene expression in response to changing environmental conditions. Many ways of regulating gene transcription respond to internal signaling molecules, metabolites, and environmental signals. For example, extracytoplasmic function (ECF) sigma (σ) factors are usually regulated by a transmembrane anti-σ factor, which, in turn, is regulated by environmental changes: the presence of a membrane or cell wall of improperly folded secretory proteins, iron chelate complexes, antibiotics. In the presence of an inducing stimulus,
the σ-factor is released and directs RNA polymerase to transcription of the corresponding target genes. It has been established that multiple ECF σ-factors belong to conservative functional groups, although many of these groups have yet to be investigated experimentally. The disclosure of the functions of the multiple ECF σ-factors present in many genomes poses a huge challenge for future research.7,8

B. thetaiotaomicron is a member of the obligate intestinal microflora that metabolizes glycans and breaks down indigestible cell wall polysaccharides, generating additional energy.9,10 Bacteria of the genus Bacteroides are sources of propionate and acetate.11,12 B. thetaiotaomicron and Bifidobacterium spp. that have the ability to utilize fructan molecules of different chain lengths in the intestinal tract demonstrate complex interactions.13 These interactions involve hydrolase family 78 α-1-rhamnosidase, an enzyme originating from B. thetaiotaomicron VPI 5482.

If the host’s diet is low in polysaccharides, then B. thetaiotaomicron can start consuming glycans, mucin and mucin-O-glycans. B. thetaiotaomicron has 12 clusters in the genome that enhance the transcription of genes for the utilization of 117 polysaccharides and mucopolysaccharides. B. thetaiotaomicron is a member of the normal intestinal microbiota. In various pathological conditions, the gene count may rise.14,15 Species such as Faecalibacterium prausnitzii, also known as butyrate producers that are rarely found in inflammation, are linked to high gene count (HGC), while the proinflammatory species (Bacteroides and Ruminococcus gravis) are frequent in low gene count (LGC).16,17 As it can be seen, such a division is closely related to the presence/absence of inflammatory bowel disease in LGC and HGC individuals. The results show that the gene count in a genome can eventually become a diagnostic tool for detecting intestinal inflammation and metabolic syndrome.

The increased cellular density of a microbial community or biofilm and the consequent accumulation of signaling or QS molecules help the metabolic activity of microbial cells, including their tolerance to antimicrobials and synthesis of virulence factors, adapt better by synchronizing their expression of different genes at a particular cell density, according to each step of the infectious process.8 The prognosis of these chronic infections, compared with a worsened by a particular property. The study aimed to develop an algorithm for genomic data processing in applied medical and scientific research. This is a first attempt to conduct a bioinformatics analysis of rRNA sequencing data derived from the reference strain B. thetaiotaomicron VPI 5482 a strain grown in a biofilm.

### MATERIALS AND METHODS

B. thetaiotaomicron reference strains were obtained from the China General Microbiological Culture Collection Center (CGMCC).

The following two bacterial strains associated with the genus Bacteroides were isolated from the blood culture samples of the patients after surgery interventions:

- Sample t6228 was isolated from the whole blood of patients with suspected descending colon cancer in gastrointestinal tumor surgery. The instrument is unable to detect the species category of the bacteria, so molecular biology methods need to be further determined.
- Sample t6000 was isolated from the whole blood of patients with a patient undergoing long-term treatment due to complications of ulcerative colitis.

Bacterial strains were grown anaerobically for 12 h at 37 °C in the tryptone–yeast extract–glucose (TYG) medium.

**Isolation of B. thetaiotaomicron DNA from a Biofilm Sample.** First, zirconia/silica beads (300 mg, 0.1 mm in diameter; 100 mg, 0.5 mm in diameter; BioSpec Products) were added to a sample of cultured B. thetaiotaomicron strains (150 mg). Then, the sample was added to 1200 μL of warm lysis buffer (50 mM Tris–HCl, pH 8.0, 500 mM NaCl, 50 mM ethylenediaminetetraacetic acid (EDTA), 4% sodium dodecyl sulfate (SDS)), vortexed, and homogenized in a MiniBead-Beater (BioSpec Products) for 3 min. The resultant lysate was incubated for 15 min at 70 °C. Thereafter, the samples were centrifuged for 20 min at 22 000 rpm. The supernatant was collected in 2 mL tubes and kept on ice (4 °C). The homogenization process was repeated with 1200 μL of lysis buffer readded to the precipitate. The supernatants were then mixed together, added 2.0 volume of 96% ethanol and 0.1 volume of 3 M sodium acetate, and incubated for at least an hour at −20 °C. Once incubated, the samples were centrifuged for 20 min at 14.000 rpm. The resultant precipitate was washed twice with 1000 μL of 80% 34 ethanol, dried in air and dissolved in deionized water. The precipitate was resuspended in 400 μL of lysis buffer. After additional centrifugation for 15 min at 22 000 rpm, the supernatant was collected in a 2 mL tube and incubated for 1 h at 37 °C with 1 μL of RNase A solution (5 mg/mL) preadded. The quality of the isolated DNA was evaluated by 1% agarose gel electrophoresis on 3 μL of purified DNA. The DNA concentration in the solution was determined with a Qubit fluorimeter (Invitrogen) using the Quanti-Tet dsDNA Broad-Range and High-Sensitivity Assay Kits (Invitrogen), according to the manufacturer’s instructions.
Sequencing Methods. The next-generation sequencing library preparation was constructed in accordance with the manufacturer’s protocol. For each sample, 100 ng of genomic DNA was randomly fragmented to <500 bp by sonication (Covaris S220). The fragments were treated with the End Prep enzyme mixture to repair ends, S’ phosphorylation, and dA-tails in a single reaction, followed by T−A ligation to add adapters to both fragment ends. Size selection of adapter-ligated DNA was performed, and then, fragments of ~470 bp (with an approximate insert size of 350 bp) were recovered. Each sample was then amplified by PCR for eight cycles using the P5 and P7 primers, with both primers carrying sequences that can anneal with the flow cell to perform bridge PCR and the P7 primer carrying a six-base index allowing for multiplexing. The PCR products were purified and validated via an Agilent 2100 (Agilent Technologies, Palo Alto, CA) and quantified using a Qubit 3.0 fluorometer (Invitrogen, Carlsbad, CA). Then, libraries with different indices were multiplexed and loaded onto an Illumina HiSeq, according to the manufacturer’s instructions (Illumina, San Diego, CA). Sequencing was performed using a 2 × 150 paired-end (PE) configuration. Image analysis and basic calling were performed using the HiSeq Control Software (HCS) + OLB + GAPipeline-1.6 (Illumina) on a HiSeq 2500 instrument.25

Reads were filtered with CutAdapt v1.9.1 to remove low-quality (QC) reads, assembled using Velvet 1.2.10, scaffolded via SSPACE v3.0, and the gaps were filled by GapFiller v1.10.26−28 Coding genes in the bacterial strains were identified using Prodigal software.29,30 Transfer RNAs (tRNAs) were detected in the genome by the program tRNAscan-SE using default settings.31 rRNAs were identified using RNAmmer 1.2 Server.32 Repeat Model software (0.8.08) was employed to predict repeat sequences in the genome. The analysis process was divided into two phases: identification of repeated sequences with RECON 1.08 and RepeatScout 1.0.5 and optimization of results with RepeatModeler and analysis of repeats using RepeatMasker 4.0.5.

Table 1 displays tools employed for the bioinformatics analysis of bacterial genome data.

Sequencing results are influenced by many factors such as sequencers, reagents, and samples. Typically, the first few sequence bases have a high probability of being incorrect due to the high rate of error generation in the base calling process. In high-throughput sequencing, the probability of error will grow with the elongation of the sequence. In the first six positions, which equal the sequence length of a reverse transcription primer (RT-primer), the sequence error rate is also high. Therefore, the high frequency of erroneous base calls at the first six-base positions is arguably due to the incomplete binding of random primers and RNA templates. To detect whether there is a high frequency of erroneous base calls in the base sequence, the sequence error rate distribution was employed. For instance, the frequency of erroneous base calls in the middle of the sequence was significantly higher as compared to other positions. If the base mass value of Illumina HiSeqTM/MiSeq is expressed in Qphred, then the following relationship exists

\[ O_{\text{phred}} = -10 \log (e) \]  

(1)

where \( e \) is a sequence error rate.

Overall, the sequence error rate for each base position should be less than 0.5%.

Read Processing. The preprocessing of raw reads involved quality filtering. Sequencing reads with the mean quality value (QV) score of 30 were considered acceptable, while those with QV < 30 were labeled as low-quality reads and were excluded from the downstream analysis. The high-quality reads were mapped without insertion or deletion (gaps) using the Bowtie2 mapping program. The maximum number of mismatches was 3, meaning that a random genome was selected for a read that could be aligned to more than one position with an equal probability. Unmapped reads were retained for further analysis.

Determination of Taxonomic Composition. The taxonomic composition of samples was determined by mapping nucleotide sequences to reference genomes from a nonredundant gene catalog using Bowtie98 software. Data
bases addressed here involved the Human Microbiome Project (HMP) data (http://www.hmpdacc.org), the National Center for Biotechnology Information (NCBI) data (http://www.ncbi.nlm.nih.gov), and other open sources. Genomes were aligned against each other by means of Cmscan 1.1.2, and those with 80% unique units were included in the catalog. The remaining genomes were clustered at an 80% sequence similarity threshold. Representative sequences (one per cluster) were manually selected and added to the catalog. In total, the catalog included 353 genomes, fully assembled, with contigs. After aligning the set of reads to a reference genome, BAM files were created to interpret BEDtools coverage data, i.e., coverage depth and coverage width values. The coverage width served as a threshold for the detection of references in the genome; reads must cover at least 1% of the genome so that the match could be found. The coverage data was normalized for each genome based on the total length of mapped reads and on the length of the genome under consideration.

The taxonomic classification of each species was conducted using the RDP classifier.

Algorithm for Processing Data Streams. First, a DNA fragment was extracted from the whole genome using the de novo sequencing method. Subsequently, a sample was detected and a library of samples was created. A long DNA sequence was cut into fragments 500 bp in length. The resultant sticky end was repaired with the blunt end and then A base was added to the 3’ end so that the DNA unit could be ligated to a linker and to a T base at the 3’ end. The ligation reaction product can be restored by the use of electrophoresis. Then, polymerase chain reaction (PCR) amplification was carried out. DNA fragments with adapters at both ends were added, and finally, the library was applied in cluster generation and sequencing. Data of 16S ribosomal RNA gene and partial sequencing were processed with the NCBI database.

Bioinformatics Analysis. The bioinformatics analysis of samples was conducted. Following the assembly results and gene predictions, rRNAs, tRNAs, and other ncRNAs were found. A predictive functional analysis was performed using the following databases: Nr, KEGG, and gene ontology (GO). According to a genomic sequence obtained for a reference strain, repeated sequence analysis was performed with strains grown on a biofilm.

Statistical Analysis and Visualization. The distance between sequence samples was measured by means of the following metrics: the Jensen–Shannon divergence and the modified UniFrac model. The Jensen–Shannon distance was calculated in R using a specially written script.

The 16S rRNA gene sequences were aligned with the Cmscan 1.1.2 multiple alignment program. Each pair of samples submitted to UniFrac analysis was assigned a weight, and genomes with zero representation across samples were removed. Data were then submitted to conversion into a biom format with convert_biom.py of QIME104, and weighted UniFrac distances were computed by applying beta_diversity.py to samples from the biom table. The samples were classified into clusters based on their enterotypes. The matrix of Jensen–Shannon distances between samples was submitted as the input data. To establish over-representation or under-representation of functional genes between two sets of samples, the one-sided Mann–Whitney test was used. Multiple comparisons were corrected using the Benjamini–Hochberg method. The difference was considered significant with the adjusted p-value not greater than 0.05.

To identify metabolic pathways, the enrichment of which varied significantly between sample sets with the KEGG database, the relative representation of enzyme genes (KEGG Orthology or KO database) entering the path was compared via the one-sided Mann–Whitney test. After the p-values were corrected for multiple comparisons, two lists of KO groups were composed, one with increased and reduced gene representation in the first group as relating to the second group. These lists were analyzed in R by using the Piano package. Statistical analysis of gene sets was performed with the Reporter Features algorithm, a parametric analogue of the GSEA algorithm. Multiple comparisons were corrected using the Benjamini–Hochberg method. The difference was considered significant with the adjusted p-value not greater than 0.05.

Limitations. The research results correspond to the medical history of an individual patient.

RESULTS

The bioinformatics analysis was performed on samples t6000 and t6228. The raw sequencing data was submitted to base calling with the subsequent conversion of base call files with Bcl2fastq version 2.17.1.14. During sequencing, the preliminary quality analysis was conducted through Illumina GAPipeline v1.9 software. Reads were mapped in the FASTA format as follows:

strain t6000

ATGGGCCTGTGCTCGGCTTACACATGCAAGTCGAGGGCCGACATTTCAGTTTGCTTGCAAACTGGAGATGGCGACCGGCGCACGGGTGAGTAGGGCGACCCGGCAGCCGGTGAAGTACACGTTACACGCCGAACCTGCCGTAACCTCGGGGATCGCTTTCGAGAAAGATTAAAC

strain t6228

TGCTCGGGCTTACACATGCAAGTCGAGGGCCGAGATTTCAGTTTGCTTGCAAACTGGAGATGGCGACCGGCGACCGGTGAAGCCGCATCCTCCAACCTGCCGTAACCTCGGGGATCGCTTTCGAGAAAGATTAAAC

For pairwise sequencing data, there were two fq files, one for Read 1 and another for Read 2. Table 2 shows raw data from each sequencing sample.

The relative representation of single genes was evaluated in terms of KO groups. The KEGG pathways of an investigated strain were divided into six categories: cellular processes, environmental information processing, genetic information processing, human diseases, metabolism, and organismic
systems. The number of genes involved in each metabolic pathway is statistically displayed in Figures 1 and 2.

After genome sequence alignment, the orthologous protein sequences from B. thetaiotaomicron strains were identified with COG. The results are presented in Tables 3 and 4 and Figures 3 and 4.

The sequence samples were mapped according to the gene ontology (GO) database and the GO annotations were retrieved alongside the relative representation values. The results of the GO analysis are presented in Tables 5 and 6.

Table 7 reports on the results of the analysis of repetitive units found in the reference genome (t6228) and the genome under study (t600).

This table also shows the results of the analysis of long sequences found in the genome under study.

We concluded that this bacterium can monitor the indicator of the disease, so to detect it by PCR analysis, we designed a series of primers (Table 8).

We also found a sigma factor regulatory protein, FecR/PupR family [B. thetaiotaomicron]

MAQINFNSIYTAAYRAFLFTLSYVHNDLVAEDIVSEAIHLWSEKREIPSVEAILTYIRSKSLNYL
KHIQAGENVFQTLDKGQRELEIRISTLEACDPKEIILSEELRAKHALLESMPEKTRTAFIRDRDLGKSHKEIAELGIVSGVYHSRAVKIRDLKNDYAPFLLFFI

**Figure 1.** KEGG pathway analysis for the t6000 sample associated with B. thetaiotaomicron strain.

**Table 1.** The statistical analysis of the sample t6000.

| Pathway category | Number of genes |
|------------------|-----------------|
| Environmental Processes | 10 |
| Cellular Processes | 15 |
| Metabolism | 20 |
| Genetic Information Processing | 10 |
| Information Processing | 5 |
| Human Diseases | 5 |
| Metabolism | 10 |
| Organism Systems | 5 |

**DISCUSSION**

This study examined the genetic correspondence of samples t6000 and t6228 samples associated with B. thetaiotaomicron strain obtained by the blood culture method used to reference strain B. thetaiotaomicron VPI 5482. Many authors argue that the induction of the QS system enhances the pathogenicity of bacteria. We found bacteria that are intestinal symbionts in the general bloodstream of patients suffering from severe pathologies of the large intestine. The t6000 sample was found to have a biofilm-forming ability as compared to the reference strain B. thetaiotaomicron VPI 5482. The bioinformatics data showed that sample t6000 contained more virulent- and stress response-associated genes as targets of the QS system when compared to t6228. These genes were responsible for adaptation, antibiotic resistance, and reduction reactions in metabolism and for the processing of information about the environment. They also regulated cellular processes (Figure 2). By contrast, the sample t6228 contained more genes responsible for the hydrolytic activity, transport and metabolism of carbohydrates and amino acids, and defense (Tables 3–6). However, sample t6000 had a gene regulating the process of alginate biosynthesis, which was not found in samples from t6228. Sample t6000 obviously exhibits the biofilm-forming function due to the synthesis of algic acid and is a potentially pathogenic strain. The literature also indicates that the C-terminal truncation of BT3147 promotes the formation of a B. thetaiotaomicron biofilm. This is consistent with the results of studies based on the genetic modifications of B. thetaiotaomicron, which were created with the use of transposons. In the studied biofilm-forming sample t6000, BT3147 was lacking the last nine C-terminal amino acids (BT3147Δ9). Since the downstream genes BT3146 and BT3145 were not involved in the phenotype under consideration, the authors constructed a chromosomal BT3147Δ9 de novo by removing the reference strain region that spanned the last nine codons of BT3147. The downstream genes BT3146 and BT3145 were replaced by a stop codon. The expression of BT3148-BT3147 inhibited the formation of the VPI 5482 biofilm, while the chromosomal-based expression of BT3148-BT3147Δ9 led to increased biofilm formation as compared to VPI 5482. It was established that the removal of gene-encoded BT3148 from B. thetaiotaomicron VPI 5482 would lead to a significant decrease in biofilm formation. The shortening of the C-terminus of BT3147 was found to promote biofilm formation in B. thetaiotaomicron. This widespread but highly variable ability to form biofilms is observed in many aerobic bacteria, including Escherichia coli, Staphylococcus epidermidis, and Bacteroides fragilis, which also show a variable biofilm formation capacity caused by sigma factor RpoQ. Biofilm formation by pathogenic bacteria is a major cause of chronic and recurring infections. The pathogenicity and abscess formation of B. fragilis are associated with the biofilm and adhesion to the peritoneal epithelium. Considering that B. thetaiotaomicron usually behaves like a nonpathogenic intestinal bacterium, it is also associated with infections of the abdominal cavity and deep wounds and accounts for up to 14% of all bacteria found at these sites. Since five of 14 best biofilm-forming strains that have been previously identified (36) were isolated from different sites of infection (bone biopsy, blood, abscesses), the B. thetaiotaomicron mutation may be associated with opportunistic extracellular infections. Further studies may provide insights into this aspect of B. thetaiotaomicron biology. As it was demonstrated, the reference strain B. thetaiotaomicron VPI 5482 forms insufficient biofilms in vitro, which is consistent with other investigations where a VPI 5482 biofilm was received only 8 days after the incubation in a chemostat. Significant biofilm growth in VPI 5482 was not observed even
after 48 h, and this may suggest repression of the VPI 5482 biofilm formation in laboratory conditions. For instance, the uropathogenic *E. coli* are known for causing infections by adhering to the bladder epithelium with fimbriae, but they exhibit a weak capacity to form biofilms in vitro. Similarly, *B. fragilis* tends to display unsatisfactory biofilm formation in standard conditions but generates good biofilms on mucin-coated plates that mimic the intestinal mucosa, its natural habitat. The present study detected Mfa1-like proteins, BT3148 and BT3147, which are potentially involved in the formation of biofilms in *B. thetaiotaomicron* VPI 5482. The studied strain had properties that allowed evaluating it in terms of...
of the following functional categories: cellular processes, environmental information processing, genetic information processing, human diseases, metabolism, and organismic systems. Its important trait is the ability to indicate gastrointestinal diseases that involve about a hundred genes. The literary sources provide a wide spectrum of data regarding the microbiome of the human gastrointestinal tract.

There have been reports on the variation in metagenome composition of microorganisms linked to a range of factors. In ulcerative colitis, an elevated level of *Bacteroides* and *Prevotella* spp. was observed. The mechanisms by which *Bacteroides* can contribute to a chronic inflammatory process are largely unknown. Among possible ones, the production of mucin-degrading sulfatases. Elevated levels of bacterial mucin-desulfating sulfatases have been reported in patients with active ulcerative colitis. In addition, the existence of enzymes that partially desulfate mucins has been demonstrated in relation to *B. thetaiotaomicron*, *B. fragilis*, and for *Prevotella* spp., suggesting that members of the Bacteroidetes may contribute to chronic inflammation due to the impaired barrier function in the epithelial cells. However, it turned out that *B. thetaiotaomicron* is linked to a rather healthy GI microbiome.
Table 6. GO Function Annotation Results of the t6228 Sample Associated with *B. thetaiotaomicron* Strain

| gene_ID | GO_ID | GO_terms | GO_aspect |
|---------|-------|----------|-----------|
| 1_2,1_7;1_81;1_83;1_363;1_418;1_450;1_485;2_242;2_253;2_271;2_273;3_127;3_255;3_350;4_90;4_230;4_252;4_260;5_82;5_134;7_90;7_165;7_237;8_120;8_154;13_62;15_18;15_43;17_81;17_82;17_116;17_133;17_143;19_1;20_46;25_21; | GO:0004553 | hydrolase activity, hydrolyzing O-glycosyl compounds | molecular_function |
| 1_2,1_7;1_12;1_81;1_83;1_126;1_249;1_306;1_363;1_418;1_450;1_485;2_113;2_120;2_121;2_133;2_242;2_253;2_271;2_272;2_273;2_339;3_63;3_98;3_119;3_123;3_127;3_269;3_302;3_330;3_350;3_371;3_435;4_90;4_186;4_230;4_252;4_260;4_285;5_54;5_82;5_134;6_13;6_67;6_90;6_118;6_128;7_84;7_90;7_237;7_254;7_318;7_374;8_35;8_42;8_51;8_53;8_73;8_79;8_86;8_92;8_120;8_154;10_6;11_57;11_107;12_7;12_20;12_38;12_56;12_67;13_62;14_120;15_18;15_42;15_60;15_79;17_27;17_42;17_43;17_71;17_72;17_78;17_81;17_82;17_95;17_116;17_133;17_143;19_1;20_46;20_68;25_21; | | |
| 1_356;1_440;3_79;3_160;3_240;5_213;6_91;7_107;8_125;8_149;10_87;10_119;18_10; | GO:000595 | carbohydrate metabolic process | biological_process |

Table 7. Long Repeats in Genome

| chr1, the chromosome ID of the query repeat sequence | start1 query repeat sequence start coordinates | end1 query repeat sequence termination coordinates | chr2, the chromosome ID of the target repeat sequence | start2 target repeat sequence starting coordinates | end2 target repeat sequence termination coordinates | strand, the target repeat sequence is in the positive/negative chain of the chromosome | length, the length of the repeat sequence | identity, % repeats, the similarity between sequences |
|-------------------------------------------------|-----------------------------------------------|-----------------------------------------------|-------------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| scaffold12size552143                             | 552014                                        | 552139                                        | scaffold2size62496                               | 1                                             | 126                                           | −                                             | 126                                          | 100.00                                        |
| scaffold18size88844                              | 88702                                        | 88844                                        | scaffold4size340877                               | 340735                                        | 340877                                        | −                                             | 143                                          | 100.00                                        |
| scaffold24size45583                              | 1                                             | 129                                          | scaffold27size69222                                | 69017                                        | 69235                                        | −                                             | 129                                          | 100.00                                        |
| scaffold25size65248                              | 45119                                        | 45248                                        | scaffold3size494252                                | 494123                                        | 494252                                        | +                                             | 130                                          | 100.00                                        |
| scaffold26size36413                              | 36284                                        | 36412                                        | scaffold27size69222                                | 1                                             | 129                                           | −                                             | 129                                          | 100.00                                        |
Table 8. Finding Primers Specific to *B. thetaiotaomicron* (Using Primer3 and BLAST)

| Primer pair | Sequence (5′ → 3′) | Template strand | Length | Start | Stop | Tm   | GC% | Self complementarity | Self 3′ complementarity |
|-------------|--------------------|-----------------|--------|-------|------|------|-----|--------------------|-------------------------|
| 1           | forward primer     | TTACACATGCACTGAGGGG | plus   | 20    | 9    | 28   | 59.75 | 55.00               | 4.00                    | 0.00                    |
|             | reverse primer     | TACGTGTACTCATCACCAGTGC | minus  | 20    | 89   | 70   | 60.04 | 55.00               | 7.00                    | 3.00                    |
|             | Product length     | 81               |        |       |      |      |      |                    |                         |                         |
| 2           | forward primer     | GGGCGACATTTTCAGTTTCTG | plus   | 20    | 26   | 45   | 59.68 | 50.00               | 4.00                    | 1.00                    |
|             | reverse primer     | GCTATCCCCAGTTATCCGTC | minus  | 20    | 117  | 98   | 60.11 | 60.00               | 5.00                    | 3.00                    |
|             | Product length     | 92               |        |       |      |      |      |                    |                         |                         |
| 3           | forward primer     | GGGGCGACATTTTCAGTTTGC | plus   | 20    | 25   | 44   | 59.83 | 55.00               | 3.00                    | 0.00                    |
|             | reverse primer     | CGAGTTATCAGGCAGTTTGA | minus  | 20    | 109  | 90   | 59.83 | 55.00               | 3.00                    | 0.00                    |
|             | Product length     | 85               |        |       |      |      |      |                    |                         |                         |
| 4           | forward primer     | TCACGTATACATCACTGCTG | plus   | 20    | 4    | 23   | 60.74 | 55.00               | 4.00                    | 2.00                    |
|             | reverse primer     | ATGCAGCAGTTATCAGGC | minus  | 20    | 106  | 84   | 60.18 | 55.00               | 4.00                    | 2.00                    |
|             | Product length     | 100              |        |       |      |      |      |                    |                         |                         |
| 5           | forward primer     | TGCAAACTGGAGATGGCGAC | plus   | 20    | 46   | 65   | 60.96 | 55.00               | 4.00                    | 2.00                    |
|             | reverse primer     | TCAGAGCGATATCCCCGAGT | minus  | 20    | 124  | 105  | 60.40 | 55.00               | 5.00                    | 1.00                    |
|             | Product length     | 79               |        |       |      |      |      |                    |                         |                         |
| 6           | forward primer     | CGAGGCAGAGCAGTTTACGTT | plus   | 20    | 22   | 41   | 60.96 | 55.00               | 3.00                    | 0.00                    |
|             | reverse primer     | CAAGAGCGATATCCCCGAGT | minus  | 20    | 123  | 104  | 59.54 | 55.00               | 5.00                    | 1.00                    |
|             | Product length     | 102              |        |       |      |      |      |                    |                         |                         |
| 7           | forward primer     | AGGGGCACATTTTCAGTTT | plus   | 20    | 24   | 43   | 59.03 | 50.00               | 3.00                    | 1.00                    |
|             | reverse primer     | CGCAAGGTGATGATCGTTG | minus  | 20    | 101  | 82   | 60.67 | 55.00               | 4.00                    | 2.00                    |
|             | Product length     | 78               |        |       |      |      |      |                    |                         |                         |
| 8           | forward primer     | GTCTACATCAGTCGAGAGGG | plus   | 20    | 7    | 26   | 59.00 | 55.00               | 4.00                    | 2.00                    |
|             | reverse primer     | TCTCGAAAGCGATATCCCCGAG | minus  | 20    | 125  | 106  | 59.25 | 55.00               | 6.00                    | 2.00                    |
|             | Product length     | 119              |        |       |      |      |      |                    |                         |                         |
| 9           | forward primer     | GGCTTACATCGAGGTGCTG | plus   | 20    | 6    | 25   | 59.00 | 55.00               | 4.00                    | 2.00                    |
|             | reverse primer     | CGGATGTAGGGGAGGTCGGGG | minus  | 20    | 110  | 91   | 60.81 | 60.00               | 5.00                    | 0.00                    |
|             | Product length     | 105              |        |       |      |      |      |                    |                         |                         |
| 10          | forward primer     | CTCGAGTACATCGAGCTGC | plus   | 20    | 3    | 22   | 59.00 | 55.00               | 4.00                    | 2.00                    |
|             | reverse primer     | CCCGAGTTATCCGGACAGGGTGG | minus  | 20    | 111  | 92   | 60.81 | 60.00               | 5.00                    | 1.00                    |
because only about 100 genes were potentially responsible for the pathological processes that took place. This is consistent with other research results that showed that B. thetaiotaomicron could indirectly enhance the therapeutic effect of immunotherapy for malignant neoplasms by increasing the T-cell production.  

The recent clinical study failed to detect B. thetaiotaomicron in the intestinal microflora in patients with ulcerative colitis. This also suggests the protective role of B. thetaiotaomicron and it may be recommended as a probiotic in cholelithiasis. However, randomized clinical trials are required to evaluate the safety and efficacy of its probiotic strains.

It has been found that the antitumor efficacy and immunostimulatory effect of the cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) blockade were linked to the activity of various species of Bacteroides such as B. thetaiotaomicron and B. fragilis.  

Genome-sequenced isolates allow drawing a conclusion about the functional capacity of genes in the reference genomes and identifying genetic variation of the studied strains in a fashion such that permits the understanding of all biological processes in which the target gene can be involved. This eliminates the need for ultradeep metagenomic sequencing and ensures that complete functional pathways are contained within an individual bacterium. In addition to increased accuracy, this method also has the capacity to improve sensitivity for functional analysis, allowing identification of functions that, although not prevalent, may represent fundamental differences between the study cohorts.

The study introduced an algorithm of the bioinformatics analysis for applied medical and scientific research and tested it on samples 6000 and 106228. The results of the bioinformatics assay on the intestinal symbiont B. thetaiotaomicron isolated from blood were presented. This was the first study to bioinformatically analyze the genome of two B. thetaiotaomicron strains based on the 16S rRNA gene sequencing data. It was found that the sample 10600 had more genes responsible for virulence as compared to the reference. Potentially, this strain represents microorganisms with acquired resistance to antibiotics. The possibility of using B. thetaiotaomicron as a marker of severe gastrointestinal diseases was examined. The results of the bioinformatics analysis may be useful when identifying gastrointestinal diseases, optimizing treatment for infections of the abdominal cavity and deep wounds, and when detecting resistance to antibiotics. Sigma factor (σ factor or specificity factor) is a protein required to initiate transcription in bacteria. This is a factor in initiating bacterial transcription, which provides specific binding of RNA polymerase (RNAP) to gene promoters. It is homologous to archeological transcription factor B and eukaryotic factor TFIIIB. We have suggested that under the action of chemotherapeutic drugs and antibiotics in bacteroids there is a violation of the initiation of transcription of the relevant genes that will vary depending on the gene and the environmental signals required to initiate the transcription of this gene. Thus, the determination of the corresponding sigma factor will make it possible to identify the external factor that led to the commission of the corresponding gene and hence to the diagnosis. Because of the genome sequencing of these bacteria, it was found that they cannot be presented to any of the known strains of B. thetaiotaomicron and it should be assumed that these bacteria are new species and require research that is more detailed.

**AUTHOR INFORMATION**

**Corresponding Author**

Lin Wu — School of Tropical and Laboratory Medicine and Key Laboratory of Tropical Translational Medicine, Ministry of Education, Hainan Medical University, Haikou, Hainan 571199, China; Department of Biotechnology and Biotechnics, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv 03056, Ukraine; orcid.org/0000-0003-0239-4456; Phone: +86089866711796; Email: linlin20wu@gmail.com

**Authors**

Zhi Cheng Wu — Department of Laboratory, First Affiliated Hospital of Hainan Medical College, Haikou, Hainan 570102, China  
Hong Xin Feng — School of Tropical and Laboratory Medicine, Hainan Medical University, Haikou, Hainan 571199, China  
Meng Zhang — Sanya People’s Hospital, Sanya 572000, China  
Wei Lan Zhou — Department of Laboratory, First Affiliated Hospital of Hainan Medical College, Haikou, Hainan 570102, China

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c03986

**Author Contributions**

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Z.C.W. and H.X.F. The first draft of the manuscript was written by M.Z., and all authors commented on previous versions of the manuscript. All authors have read and approved the final manuscript. Conceptualization, Z.C.W.; methodology, H.X.F.; formal analysis and investigation, W.L.Z.; writing original draft, M.Z.; review and editing, L.W.; and supervision, Z.C.W.

**Funding**

This research was supported and receive funding from National Natural Science Foundation of Hainan Province (417146 and 817324), the Colleges and Universities Scientific Research Projects of the Education Department of Hainan Province (20181181072).

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

This work was supported by the National Natural Science Foundation of Hainan Province (417146 and 817324) and the Colleges and Universities Scientific Research Projects of the Education Department of Hainan Province (20181181072).

**REFERENCES**

(1) van der Meulen, T. A.; Harmsen, H. J.; Bootsma, H.; Spijkervet, F. K.; Kroese, F. G.; Vissink, A. The microbiome—systemic diseases connection. Oral Dis. 2016, 22, 719–734.  
(2) Lukáš, F.; Gorenc, G.; Kopecny, J. Detection of possible AI-2-mediated quorum sensing system in commensal intestinal bacteria. Folia Microbiol. 2008, 53, 221–224.  
(3) Pedrolli, D. B.; Ribeiro, N. V.; Suziato, P. N.; de Jesus, V. N.; Cozetto, D. A.; Tuma, R. B.; Gracindo, A.; Cesar, M. B.; Freire, P. J. C.; da Costa, A. F. M.; Lins, M. R. C. R.; Correa, G. G.; Cerri, M. O. Engineering microbial living therapeutics: the synthetic biology toolbox. Trends Biotechnol. 2019, 37, 100–115.  
(4) Antunes, L. C. M.; Ferreira, L. Q.; Ferreira, E. O.; Miranda, K. R.; Avelar, K. E. S.; Domingues, R. M. C. P.; de Souza Ferreira, M. C.
Bacteroides species produce Vibrio harveyi autoinducer 2-related molecules. *Anaerobe* 2005, 11, 295–301.

(5) Brennan, J. J.; Gilmore, T. D. Evolutionary origins of toll-like receptor signaling. *Mol. Biol. Evol.* 2018, 35, 1576–1587.

(6) Abaturov, A. E.; Kryuchko, T. A. Pharmacological effect on biofilm dispersion. Derivatives of the diffusible signal factor family. *Child Health* 2019, 14, 386–392.

(7) Saakyan, S. V.; Myakoshina, E. B.; Krichievskaya, G. I.; Slepova, O. S.; Panteleeva, O. G.; Andryushin, A. E.; Khoroshilova, L. P.; Zakharova, G. P. Testing patients with ulcerative melanoma for herpesvirus infections. *Viropro Viruslogi* 2016, 61, 284–287.

(8) Nervo, V. V.; Saakyan, S. V.; Myakoshina, E. B.; Okhotinskaya, T. D.; Fadeeva, V. A. Role of optical coherence tomography angiography in diagnostics of early choroidal melanoma and circumscribed choroidal hemangioma. *Vestn. Oftalmol.* 2018, 134, 4–18.

(9) Jogà, W.; Maurice, C. F. Polysaccharide protection: How Bacteroides thetaiotaomicron survives an antibiotic attack. *Cell Metab.* 2019, 30, 619–621.

(10) Porter, N. T.; Luis, A. S.; Martens, E. C. Bacteroides thetaiotaomicron. *Trends Microbiol.* 2018, 26, 966–967.

(11) Liou, C. S.; Sink, S. J.; Diaz, C. A.; Klein, A. P.; Fischer, C. R.; Higginbottom, S. K.; Erez, A.; Donia, M. S.; Sonnenburg, J. L.; Sattely, E. S. A metabolic pathway for glucosinolate activation by the human gut symbiont Bacteroides thetaiotaomicron. *bioRxiv* 2019, 180, 717–728.

(12) Mihajlovic, J.; Bechon, N.; Ivanova, C.; Chain, F.; Almeida, A.; Langella, P.; Beloin, C.; Ghigo, G. M. A putative type V pilus contributes to Bacteroides thetaiotaomicron biofilm formation capacity. *J. Bacteriol.* 2019, 201, No. e00650.

(13) Falony, G.; Calmeyn, T.; Leroy, F.; De Vuyt, L. Cocculture fermentations of Bifidobacterium species and Bacteroides thetaiotaomicron reveal a mechanistic insight into the prebiotic effect of inulin-type fructans. *Appl. Environ. Microbiol.* 2009, 75, 2312–2319.

(14) Shelton, A. N.; Seth, E. C.; Mok, K. C.; Han, A. W.; Jackson, S.; Eberhardt, R. Y.; Eddy, S. R.; Floden, E. W.; Gardner, P. P.; Jones, T. A.; Tate, J.; Finn, R. D. Rfam 12.0: updates to the RNA families’ database. *Nucleic Acids Res.* 2015, 43, D130–D137.

(15) Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, E.; Lewis, S.; Matse, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G. Gene ontology: tool for the unification of biology. *Nat. Genet.* 2000, 25, 25–29.

(16) Lazar, V. Biofilm tolerance and new anti-biofilm strategies based on traditional antimicrobial natural products containing quorum sensing inhibitors. *Microb. Ecol. Health Dis.* 2013, 24, 9–10.

(17) Chang, P. V. Chemical mechanisms of colonization resistance by the gut microbial metabolome. *ACS Chem. Biol.* 2020, 15, 1119–1126.

(18) Roizch, N. S.; Blair, A. B.; Burkhart, R. A. Organoids: A Model For Precision Medicine. In *Precision Medicine for Investigators, Practitioners and Providers*; Faintuch, J.; Faintuch, S., Eds.; Academic Press: New York, 2020; pp 123–129.

(19) Sattely, E. S. A metabolic pathway for glucosinolate activation by the human gut symbiont Bacteroides thetaiotaomicron survives an antibiotic attack. *Cell Metab.* 2009, 201, No. e12856.
Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. Science 2015, 350, 1079–1084.