Metagenomic Analysis of Bile Microbial Diversity in Primary Sclerosing Cholangitis

Final Report

Patrick SC Leung, Ph.D.

University of California, Davis.

Lay Summary

Microorganisms are ubiquitous in the environment. However, the association and specific contribution of microorganisms in human health has been largely unexplored until recent years. The human microbiome interacts with each other and with our body in dynamic equilibrium. These interactions also influence the immune system, resulting in immunological and physiological responses including autoimmunity. In this study, we have focused to examine the microbial diversity in the bile of patients with primary sclerosing cholangitis (PSC), an autoimmune bilary tract disease. We have applied the cutting edge technology in molecular biology and bioinformatics to examine the bacterial population in these bile samples. Our data, for the first time, reveal previously hidden microbiological diversity in human bile. Knowledge on the composition and relative microbial composition in the bile will be helpful to further understand interplay of microbial world and bile duct physiology in biliary tract disease.

Project Summary

While circumstantial evidence and clinical observation supports the thesis of a microbial contribution in primary sclerosing cholangitis (PSC), the identification of putative organisms responsible for biliary pathology has been challenging and were primarily based on non-standardized PCR, conventional microbiological cultures, and biochemical techniques. In this project, we have applied the state of the technology in molecular biology, bioinformatics and the use of clinical samples to determine the microbial composition in bile samples of patients PSC by analysis of bacterial 16S rRNA genes. Briefly, bile samples were collected for DNA isolation, amplification of microbial 16S rRNA gene sequence, nucleotide sequence determination and bioinformatics analysis of 16S rRNA gene sequence. 16S rRNA samples were from 9 subjects were amplified by PCR using a panel of microbial specific primers 16S-rRNA amplification and DNA sequencing was successful in 8/9 bile samples. These sequences were subsequent analyzed using QIIME 1.8.0 and operational taxonomic units (OUT) picking approach. Proteobacteria account for over 90% of the bacteria in 8/8 bile samples, followed by Bacteroidetes and Actinobacteria. The composition and diversity of bacterial population in 6/8 of the bile samples are very similar. Interestingly, a single bacterium *Pseudomonas* OTU (ID: 141105), is dominant in these 6 bile samples, accounting for almost half of the bacterial community. However, 2/8 bile samples are distinct in their microbial population, which with Haemophilus and unclassified Rhodobacteraceae 16S rRNA were present. The data is significant because of its unbiased sampling, rigorous analysis of microbial members of the sampled communities and has the potential to reveal previously hidden microbiological diversity in bile. Further work in this direction will help to understand the microbial involvement in the biliary tract in PSC.

MATERIAL AND METHODS

Subjects and Samples

Upon informed consent, bile samples were collected from patients undergoing endoscopic retrograde cholangiopancreatography for at hospital of University of California at Davis medical center (Sacramento, CA). Patient demographics and diagnoses were recorded and bile samples were collected during the procedure. Bile was aspirated after selective cannulation of the common bile duct, and stored at -80 ℃. This protocol was approved by the IRB.

DNA extraction from bile

Bile samples were thawed at room temperature; 1 mL of each specimen was centrifuged at 14 000 g for 30 min, and the supernatant was discarded. This has the effect of concentrating any bacterial cells in the pellet, and of removing some of the inhibitors present in bile.

The samples were run through the PowerSoil DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA) as per the manufacturer's instructions. Briefly, the pellet was suspended in beads and bead solution with 60 µl lysis solution C1, then inverted several times and vortexed for 5 minutes. The lysized extract was centrifuged at 10,000 x g for 30 seconds at room temperature. 500 µl supernatant was transferred to a new tube. 250 µl solution C2 was added to the tube, the mixed solution was vortexed for 5 seconds, incubated at 4 ℃ for 5 minutes, then centrifuged at 10,000 x g for 30 seconds at room temperature. 600 µl supernatant was transferred to a new tube. 200 µl solution C3 was added to the tube, the mixed solution was vortexed for 5 seconds, incubated at 4 ℃ for 5 minutes, then centrifuged at at 10,000 x g for 1 minute at room temperature. 750 µl supernatant was transferred to a new tube. 1200 µl solution C4 was added to the tube, 675 µl mixed solution was loaded to spin filter, centrifuged at 10,000 x g for 1 minute at room temperature. The flow through was discarded, the remaining supernatant was loaded and

centrifuged. DNA is selectively bound to the silica membrane in the Spin Filter device in the high salt solution. Contaminants pass through the filter membrane, leaving only DNA bound to the membrane. The spin filter column was washed by 500 µl solution C5, Solution C5 is an ethanol based wash solution used to further clean the DNA that is bound to the silica filter membrane in the Spin Filter. 30 µl of Solution C6 was added to the center of the white filter membrane, centrifuged at 10,000 x g for 30 seconds at room temperature to elute DNA. DNA concentration was measured by NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific, Wilmington, DE). The DNA thus extracted was stored at -20 ℃ until PCR testing.

DNA amplification

Ampicon libraries was prepared on the DNA samples using Phire Hot Start II (Thermal Scientific, Pittsburgh, PA) and specifically designed fusion primers. For each sample, a specific forward fusion primer will be designed, which composes of an adaptor required for pyrosequencing, a sample-specific MID tag and a template-specific primer for amplification of the 16SrRNA hypervariable region. The template-specific primers to be used are U789F (reversed) and U519F (forward). The primers are chosen based on their high coverage of Archaea and Eubacteria. U519F (position 519-537) covers 96.7% of Archaea and 98.5% Eubateria, and U789F (position 789-807) covers 97.7% of Archaea and 94.8% of Eubacteria. In addition, we also chose other primer pairs which covers the 16SrRNA V1-2 region for short amplicons (27F and 338R) and primer sets 27F and 534R which covers the 16SrRNA V1-3 region, primer set (357F and 926R) which covers the 16SrRNA V3-5 region for longer amplicons, which could serve as alternatives in the study. Importantly, known positive and negative sample will be used throughout as quality controls.

Template DNA was subjected to PCR using the 16SrRNA-specific primers, which amplify a 376-base pair segment within the conserved 16SrRNA gene. The PCR mix (50 µl) contained 0.1 mm of each of the four deoxynucleotide triphosphates (dNTPs), 2 µm of each of the two primers, 2 U taq polymerase and 10X reaction buffer (Thermal Scientific, Pittsburgh, PA)) and 5 µl of template DNA. The PCR was conducted in a Gene Amp PCR system 9700 (Applied Biosystems, Foster City, CA) under the following conditions: 98 ℃ for 30 seconds; 35 cycles at 98 ℃ for 10 seconds, 58 ℃ for 30 seconds and 72 ℃ for 30 seconds; 72 ℃ 3 min. The PCR products were separated in a 2% agarose gel stained with ethidium bromide. The expected PCR products were visualized with a gel documentation system FluorChem 8900 (Alpha Innotech, San Leandro, CA). The PCR samples were purified by QIAquick PCR Purification Kit (Qiagen Sciences Inc, Valencia, CA) and sequenced by Illumina sequencing platforms MiSeqs (The DNA Technologies and Expression Analysis Cores, Genome Center, Davis, CA).

Bioinformatics analysis

Following quality filtering, the sequences were de-multiplexed and trimmed before performing sequence alignments, identification of operational taxonomic units (OTU), clustering, and phylogenetic analysis using QIIME open-source software (http://qiime.org). QIIME 1.8.0 was used for the data analyses. The closed-reference operation taxonomic units (OTU) picking approach was adopted here: reads were first clustered into OTUs at a 97% similarity threshold and then searched against the Greengenes sequence collection (release 13_8). Reads that did not hit the dataset were discarded from the downstream analyses. All the datasets were rarified to 200,000 reads before alpha diversity analysis. Beta diversity among samples was compared using the principal coordinates analysis with weighted UniFrac distances as input.

RESULTS and DISCUSSION.

1. Proteobacteria is the dominant phylum.

Using a taxonomical approach, we first determined the major phylum in the samples. Proteobacteria accounts for over 90% of the bacteria in 8/8 bile samples, followed by Bacteroidetes and Actinobacteria. Of note, Actinobacteria is more prominent in 1/8 (sample B7) than Bacteroidetes. Firmicutes 16S rRNA was detected in 7/8 of the samples. (Figure 1).

2. Relative abundance of dominant taxomical bacterial class, order, family and genus.

Next, we examined the bacterial diversity in the samples according to their class, order, family and genus (Figures 2-5). Gammaproteobacteria and Alphaproteobacteria are the more dominant class in all bile samples. Pseudomonadales is predominant class in all but two (samples B5 and B8) of the bile samples. The relative abundance of bacterial orders Pasteurellales and Enterobacteriales in samples B5 and B8 is higher than the other 6 bile samples. The distribution of bacterial family follows the pattern in the class with highest abundance being Pseudomonadaceae in 6/8 bile samples; Pasteurellaceae and Enterobacteriaceae are predominant in samples B5 and B8. The genus *Pseudomonas* is present in all bile samples and accounts for over 40% -60% of the bacterial genera in 6/8 bile samples. However, we also note that Haemophilus and others unclassified Rhodobacteraceae are in high abundance in samples B5 and B8.

In summary, all bile samples, except B5 and B8, shared similar bacterial composition and diversity. Apart from B5 and B8, the remaining bile samples were dominated by a single *Pseudomonas* OTU (ID: 141105), which composed about half of the bacterial community in each sample. However, the exact identity of the OTU could not be revealed due to the short read length (~250bp). Furthermore, Hemophilus and Rhodobacteraceae are present in 3/8 bile

samples (B2, B5 and B8) with a higher abundance in B5 and B8. The similarity and differences of the bacterial composition in the bile could reflect ecological environments that contribute to clinical diversification of bile duct related diseases. The frequent association of autoimmune-mediated inflammation of the liver in PSC with chronic inflammatory bowel disease (IBD) such as ulcerative colitis and Crohn's colitis suggest that translocation of bacteria/bacterial antigens into the portal circulation can trigger immunological responses in the biliary tract and that cross-recognition between microbial antigens and host components of the immune system, along with stimulation via pattern recognition receptors, might lead to chronic hepatic inflammatory disorders. Future study with larger numbers of patients, including patients at different stages of disease activity and including patients with PSC with chronic IBD are warranted to understand the microbial involvement in the biliary tract in PSC.

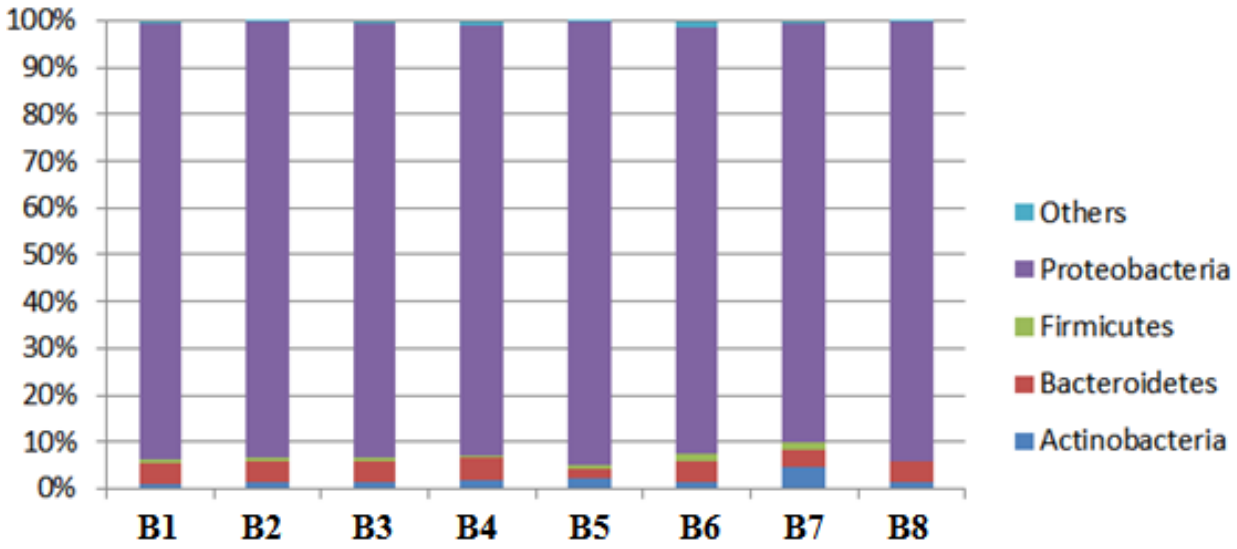

Figure 1. Relative abundance of dominant bacterial phyla in each bile sample. Only phyla >1% in any one group was included.

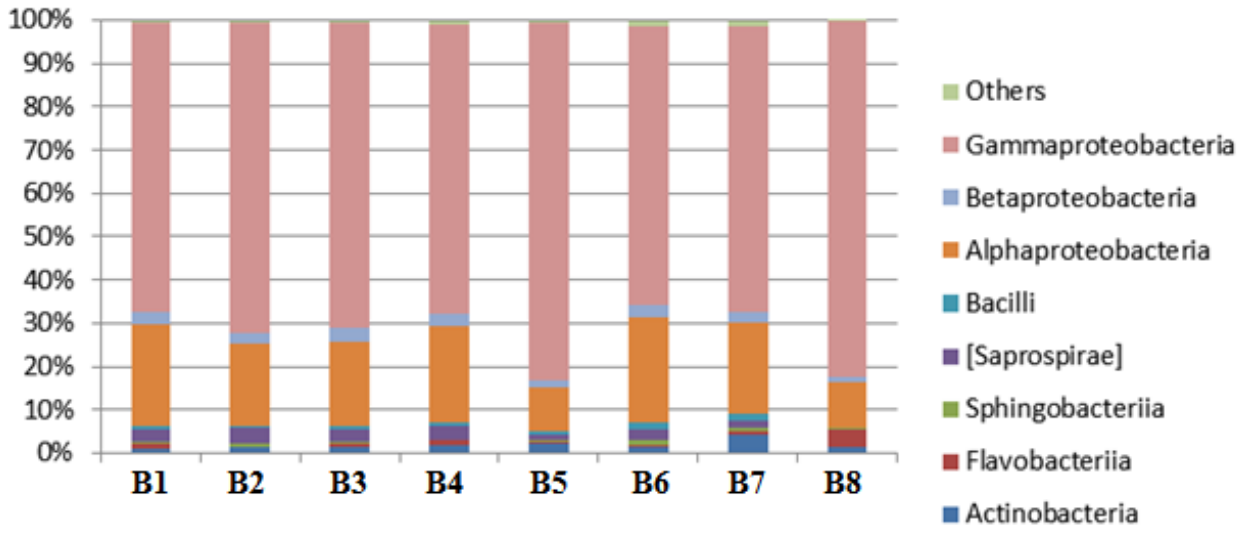

Figure 2. Relative abundance of dominant bacterial classes in each bile sample. Only classes >1% in any one group was included.

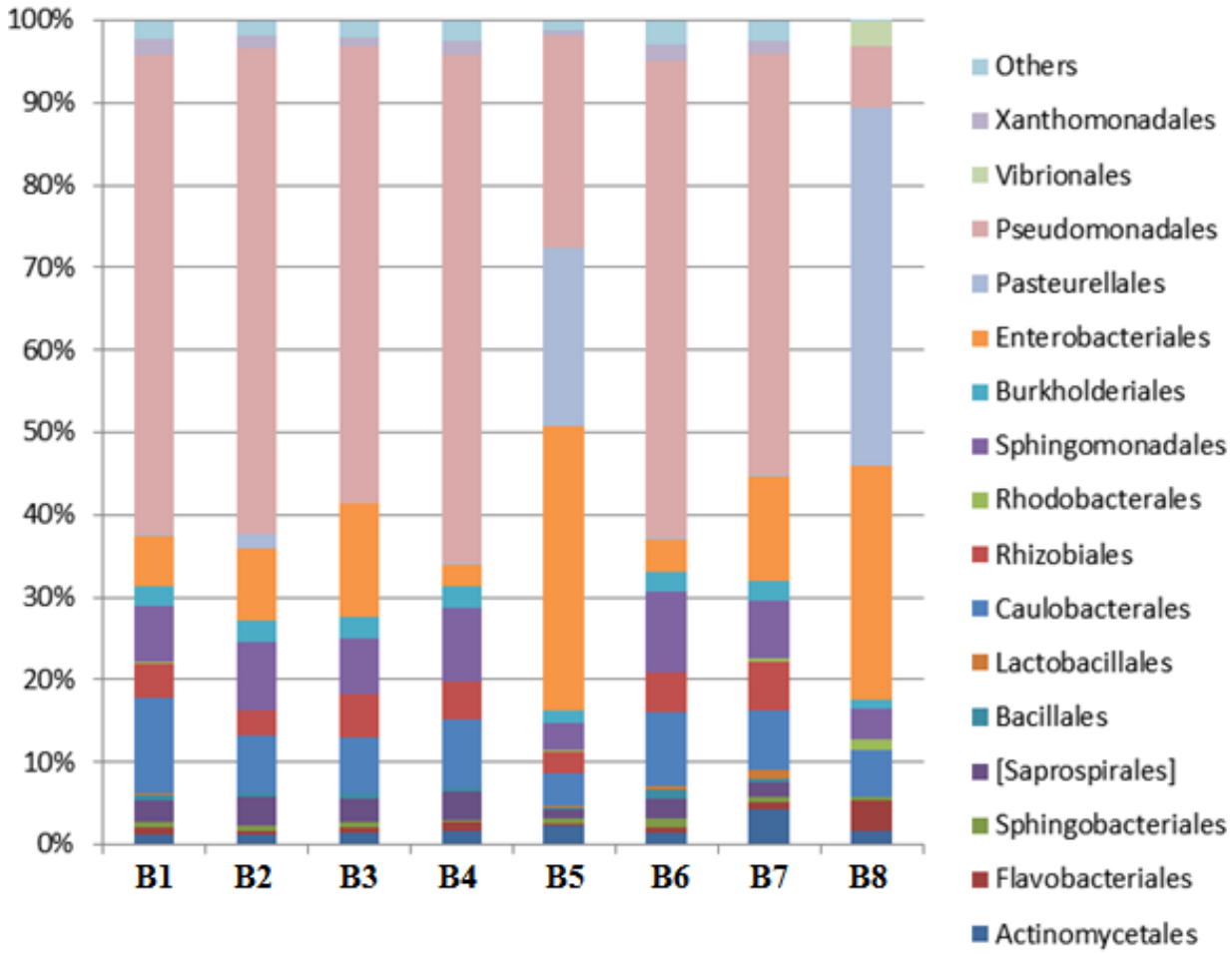

Figure 3. Relative abundance of dominant bacterial orders in each bile sample. Only orders >1% in any one group was included.

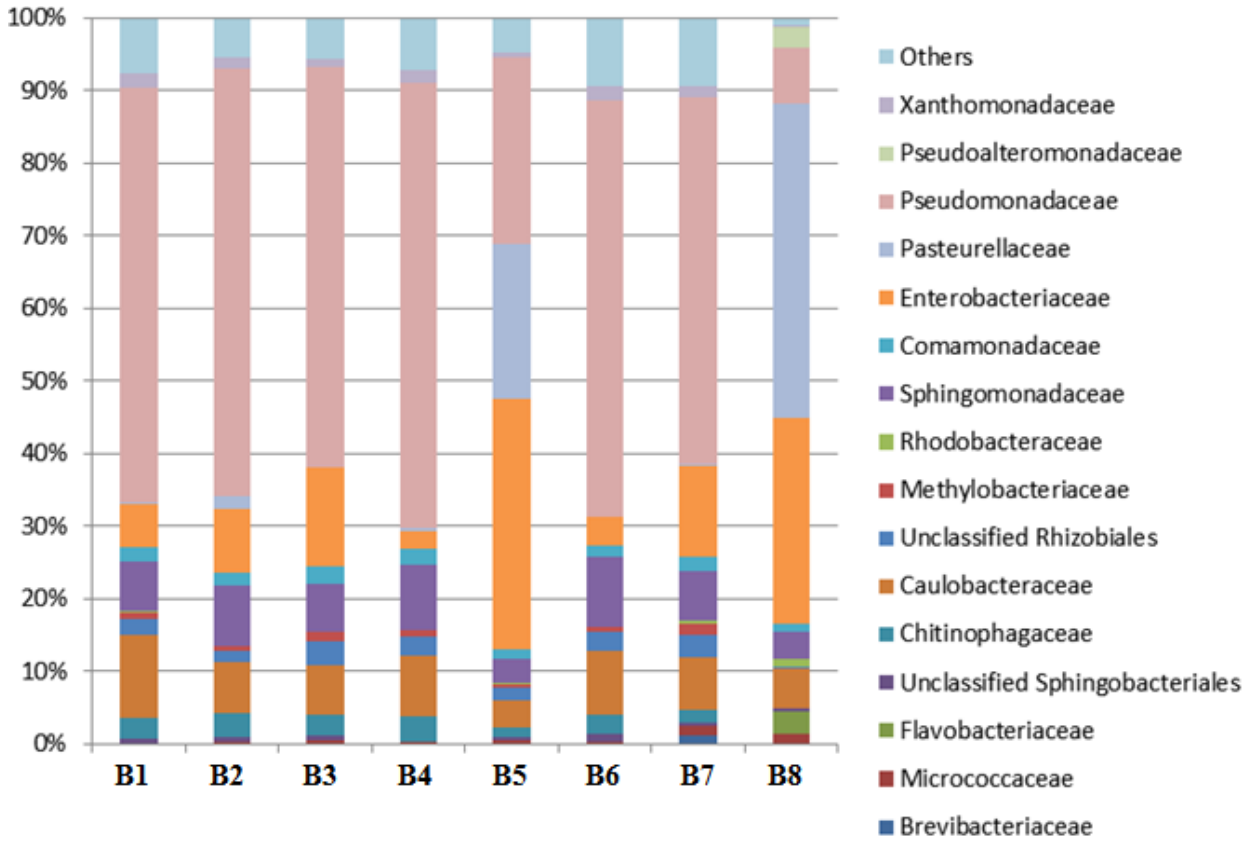

Figure 4. Relative abundance of dominant bacterial families in each bile sample. Only families >1% in any one group was included.

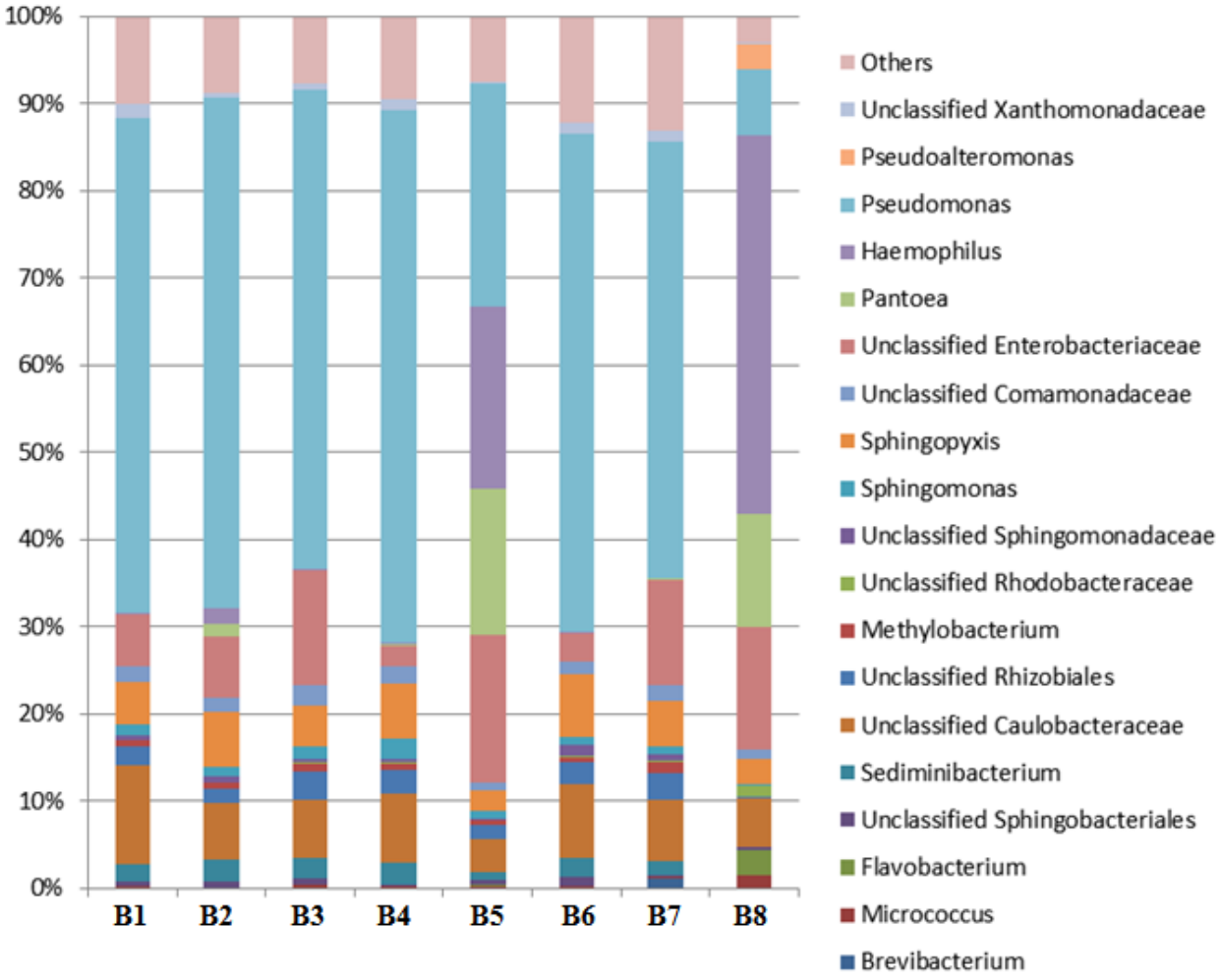

Figure 5. Relative abundance of dominant bacterial genera in each bile sample. Only genera >1% in any one group was included.