Targeted high-resolution taxonomic identification of *Bifidobacterium longum* subsp. *infantis* using human milk oligosaccharide metabolizing genes

Lauren Tso[1†], Kevin S. Bonham[1†], Alyssa Fishbein[1], Sophie Rowland[1], RESONANCE consortium, Vanja Klepac-Ceraj[1*]

[1] Department of Biological Sciences, Wellesley College
[†]These authors have contributed equally to this work and share first authorship

***Correspondence:**
Vanja Klepac-Ceraj
vklepacc@wellesley.edu

Keywords: human milk oligosaccharide metabolizing genes, *Bifidobacterium infantis*, metagenomics, comparative genomics, child development, pediatric cohort, metabolic potential

Running title: *Bifidobacterium infantis* HMO-metabolizing genes

Abstract

Bifidobacterium longum subsp. *infantis* (*B. infantis*) is one of few microorganisms capable of metabolizing human breast milk and is a pioneer colonizer in the guts of breastfed infants. One current challenge is differentiating *B. infantis* from its close relatives, *B. longum* and *B. suis*, by molecular methods. These two organisms are classified in the same species group as *B. infantis* but do not share the ability to metabolize human milk oligosaccharides (HMOs). Here, we compared HMO-metabolizing genes across different *Bifidobacterium* genomes to develop *B. infantis* specific primers and determine if they alone can be used to quickly characterize *B. infantis* with shotgun metagenomic sequencing data. We showed that *B. infantis* is uniquely identified by the presence of five HMO-metabolizing gene clusters, used this characterization to test for its prevalence in infants, and validated the results using the *B. infantis*-specific primers. By examining stool samples from a cohort of US children and pregnant women using shotgun metagenomic sequencing, we observed that only 18 of 204 (8.8%) of children under 2 years old harbored *B. infantis*. These results highlight the importance of developing and improving approaches to identify *B. infantis.* A more accurate characterization may provide insights into regional differences of *B. infantis* prevalence in infant gut microbiota.

1 INTRODUCTION

The human microbiota has coevolved with humans for tens of thousands of years (Lebreton et al., 2017). This coexistence requires a tight partnership between the host and its resident microorganisms. For example, the abundant oligosaccharides found in human milk cannot be digested by the infant who feeds on them. Instead, these are used to support the growth of the pioneer gut colonizers, who then, in turn, support the development of the infant (Frese et al., 2017; Le Doare et al., 2018). In particular, taxa in the *Bifidobacterium* genus often have genes that allow them to metabolize the complex carbohydrates in their hosts (Milani et al., 2014, 2016; Sun et al., 2015). The subspecies *Bifidobacterium longum* subsp. *infantis* (*B. infantis*), for example, can import and degrade a variety of complex human milk oligosaccharides (HMOs) (Garrido et al., 2011). Organisms belonging to the same *longum* subgroup, *B. longum* subsp. *suis* and *B. longum* subsp. *longum,* are unable to metabolize HMOs.

The functional potential of *B. infantis* to metabolize HMOs lies in five genomic loci dedicated to the import of HMOs and their processing (LoCascio et al., 2010). The 79 genes within the five clusters reveal substantial glycolytic potential (LoCascio et al., 2010), allowing for the transmembrane import of milk oligosaccharides into the cytosol and later cleaving into monosaccharides (Sela et al., 2008). The proximity of these genes indicates a role in the regulation of a common metabolic pathway that has coevolved with milk components to outcompete other microbes in the infant gut (LoCascio et al., 2010). In the developing infant gut, where milk is the primary source of nutrients, this ability provides a selective advantage (Sela et al., 2008). Given that other taxa of interest may also have differential functional potential, high-resolution identification of all species in the community is desirable. One of the most widely used culture-independent methods to taxonomically profile gut microbiomes, 16S rRNA amplicon sequencing, cannot differentiate *B. infantis* from its closest relatives (Satokari et al., 2002; Bonjoch et al., 2004; Matsuki et al., 2004; Johnson et al., 2019). Whole metagenome (“shotgun”) sequencing can provide additional taxonomic resolution, but existing tools for assembly-free taxonomic assignment also commonly fail to differentiate subspecies (Truong et al., 2015; Wood et al., 2019).

While the HMO-metabolizing gene suite may distinguish *B. infantis* from other *B. longum* subspecies, it remains unclear whether this HMO metabolic potential is confined to a single lineage or is shared across closely related organisms (e.g., through horizontal gene transfer). The origins of these gene cassettes also remain obscure, though other species such as species belonging to *Bifidobacterium*, *Lactobacillus* and *Bacteroides* genera have HMO utilization potential (Sela et al., 2008; Marcobal et al., 2010, 2011; Ruiz-Moyano et al., 2013; Yu et al., 2013; Garrido et al., 2015; Thongaram et al., 2017). Understanding the distribution of these genes in microbial species may enable using metagenomic sequencing for the interrogation of *B. infantis* colonization, succession, and distribution in the infant microbiome in new stool metagenomes as well as those that are already available (Sela et al., 2008). In this study, we

tested the *in vitro* and *in silico* use of these 79 HMO-metabolizing genes and determined that they alone can identify *B. infantis* in a stool sample and assess its prevalence in the population. We then applied this characterization of *B. infantis* to search for its presence in a US pediatric cohort with shotgun metagenomic sequencing. We validated these results by designing and benchmarking subspecies-specific PCR primers targeting a subset of these HMO-metabolizing genes.

2 MATERIALS AND METHODS

2.1 Bacterial genomes

The genome sequences of 387 *Bifidobacterium* species and subspecies were downloaded in FASTA format from the National Center for Biotechnology Information (NCBI) database. Genomes included 4 *Bifidobacterium breve* and 383 *Bifidobacterium longum* species of *Bifidobacterium*, spanning all three subspecies of *B. longum* (43 *B. longum* subsp. *infantis*, 167 *B. longum* subsp. *longum*, 3 *B. longum* subsp. *suis*). In addition, 170 of the genomes were unspecified *B. longum* species and were also included in the analysis (Supplementary Table 1).

2.2 Calculation of ANI values

All pairwise, whole-genome sequence comparisons were performed using average nucleotide identity (ANI). ANI is an *in silico* substitute for DNA-DNA hybridization (DDH) and thus is useful for delineating species boundaries (Konstantinidis and Tiedje, 2005). The Python module pyani (v 0.2.7) was used to calculate ANI values (Pritchard et al., 2016). Specifically, the genomic sequence from one of the genomes in a pair was aligned against the other using the matching algorithm MUMmer (v 4.0.0) through the ANIm command (Kurtz et al., 2004). The calculated ANI value was the average percentage nucleotide identity for all matching regions between a pair of genomes. A percentage identity table returned from this analysis was then converted to a distance matrix using R 3.6.0 (Supplementary Table 2). The neighbor-joining ANI tree consisting of 387 *Bifidobacterium* genomes was built using the default parameters with the phytools (v 0.6-99) package. Each tree was rooted by the four *B. breve* genomes included as an outgroup and bootstrapped with values assigned to internal edges over 100 iterations.

2.3 Genomic analysis of HMO-metabolizing clusters

Specificity of the HMO-metabolizing gene cluster was determined in published genomes from NCBI and sequenced metagenomes from fecal samples. The presence of HMO-metabolizing clusters in each of the *Bifidobacterium* genomes from NCBI was determined by BLASTN hits and filtered for 95% identity. Heatmaps were colored by the presence of at least one BLAST hit to a gene. The presence of HMO-metabolizing clusters in the non-redundant database from NCBI was determined by BLASTP hits and filtered for 80% identity (Altschul et al., 1990).

Evolutionary pressures were assessed by differences in the substitution rates at nonsynonymous (dN) and synonymous sites (dS) using MEGA X (Kumar et al., 2018; Stecher et al., 2020). A significant difference in the relative abundance of synonymous and

nonsynonymous substitutions implies non-neutral selection, with a higher ratio of nonsynonymous substitutions indicating positive selection and a lower ratio indicating purifying selection (Kryazhimskiy and Plotkin, 2008).

2.4 Subjects and samples

Samples (n=1092) used in this study were collected from the RESONANCE (Brown University Assessment of Myelination and Behavioral Development Across Maturation) longitudinal study looking at the brain and cognitive development at Rhode Island Hospital in Providence, RI, USA. All procedures and experiments in the study were approved by and followed the guidelines of the local review board. Written consent was obtained from all parents or guardians of participants. Participants under two completed a study visit every six months and participants beyond two years of age, once a year. Fecal samples were collected by patients in OMR-200 tubes (OMNIgene GUT, DNA Genotek, Ottawa, Ontario, Canada), brought to the RI lab and frozen within 24 hours at -80℃ in the RI lab.

2.5 Sample processing and sequencing

All samples were processed at Wellesley College (Wellesley, MA). Nucleic acids were extracted from stool samples using the RNeasy PowerMicrobiome kit automated on the QIAcube (Qiagen, Germantown, MD), excluding the DNA degradation steps. Extracted DNA was sequenced using shotgun metagenomic sequencing at the Integrated Microbiome Resource (IMR, Dalhousie University, NS, Canada). A pooled library (max 96 samples per run) was prepared using the Illumina Nextera Flex Kit for MiSeq and NextSeq (a PCR-based library preparation procedure) from 1 ng of each sample where samples were enzymatically sheared and tagged with adaptors, PCR amplified while adding barcodes, purified using columns or beads, and normalized using Illumina beads or manually. Samples were then pooled onto a plate and sequenced on the Illumina NextSeq 550 platform using 150+150 bp paired-end “high output” chemistry, generating ~400 million raw reads and ~120 Gb of sequence. All metagenomic data were deposited in SRA in BioProject PRJNA695570. The list of samples used in this study is listed in Supplementary Table 3.

2.6 Identification of *B. infantis* in metagenomic samples by the presence of HMO-metabolizing clusters

The presence of *B. infantis* in each of the 1092 sample metagenomes included in our study was determined by that of the majority (75%) of the 79 HMO-metabolizing genes in the gene cluster. Metagenomic samples were analyzed using the bioBakery (McIver et al., 2018) family of tools with default parameters. Briefly, KneadData (v 0.7.1) was used to trim and filter raw sequence reads and to separate human reads from bacterial sequences. Samples that passed quality control were taxonomically profiled to the species level using MetaPhlAn (v 3.0.1). Paired-end reads of the 150 bp nucleotide sequences acted as inputs against a reference database including the 79 HMO-metabolizing genes using the alignment tool Bowtie2 (v2.4.2) (Langmead et al., 2019). Bowtie2 was used in these searches, and not in any others, for its specific ability in aligning short reads to relatively long genomes. Presence of gene expression was noted if at least one mapped alignment was found to that gene.

2.7 Primer design and testing

Previously published primers designed to target *B. longum* subsp. *infantis* were either too specific and did not target *B. infantis* or were too broad and targeted other members of the *B. longum* group (Table 1; Tannock et al., 2013; Frese et al., 2017; Lawley et al., 2017). We decided to test the existing primers *in silico* and to develop new primer sets. To find candidate regions, we analyzed 79 genes associated with the five HMO-metabolizing clusters and selected genes that were present in all *B. infantis* and that had gene sections highly conserved across the different strains. We chose four HMO metabolizing genes that aligned >97% to each *B. infantis* but not to any other bacteria and the included: exo-α-sialidase gene (locus tag 'Blon_2348'), glycoside hydrolase gene (locus tag 'Blon_2355'), haloacid dehalogenase-like hydrolase domain-containing protein gene (locus tag 'Blon_2356'), and β-lactamase domain protein gene (locus tag 'Blon_2358'). To test the target specificity of the published primers and to design the new primers, 48 *B. longum* group genomes and 2 *B. suis* genomes were imported into Geneious (v R11.1.5; https://www.geneious.com), see Supplementary Table 1. We aligned the genes from the selected genomes using MUSCLE (v 3.6) alignment in Geneious and used them to design the primer pairs via the Primer3 (v 2.3.7) (Edgar, 2004; Untergasser et al., 2007). Each primer pair was checked using NCBI primer-BLAST to ensure that it only matches *B. infantis* sequences (Ye et al., 2012). All primers targeting HMO-metabolizing genes were checked against 387 genomes (Supplementary Table 1) to assess their coverage (Table 1). Primers were synthesized from Integrated DNA Technologies, Inc (Coralville, Iowa, USA).

To ensure that the primers were adequate for targeted detection of *B. infantis*, we used genomic DNA from isolated cultures of *B. infantis* (DSM 20088), *B. longum* (DSM 20219), *B. breve* (DSM 20213) and *E. coli* K12. The cultures were purchased from the German Collection of Microorganisms and Cell Cultures GmbH (DSMZ, Leibniz, Germany). The *Bifidobacterium* strains were incubated anaerobically at 37˚C for 36 hrs. The cells of each culture were centrifuged at 13,000 rpm for 1.5 min and washed in anaerobic PBS. Genomic DNA was then extracted using DNeasy PowerSoil Pro Kit (Qiagen, Germantown, MD, USA) following manufacturer's instructions. DNA concentrations were determined using an absorbance ratio of 260/280 nm.

2.8 Development of PCR conditions

PCR amplifications were performed using OneTaq Quick-Load 2X Master Mix (New England BioLabs, Ipswich, MA), 10 µM of each primer, and 10 ng of template DNA (25 µL total volume) following manufacturer's instructions. After the initial denaturation step at 94˚C for 30 s, there were 30 cycles of 94˚C for 30 s, the respective annealing temperature for 60 s, and 72˚C for 60 s, and then finally 68˚C for 5 min (see Table 2 for annealing temperatures). The positive control was gDNA of isolated cultures of *B. infantis* (DSM 20088) or of *B. breve* (DSM 20213) and the negative control contained water instead of a template DNA were included with each batch.

2.9 Confirmation of HMO-metabolizing genes in the metagenomes by PCR

To detect *B. infantis* by PCR in fecal samples, we used the *B. infantis* exo-α-sialidase gene (locus tag 'Blon_2348') primer pair (Sia-266F and Sia-676R). We also used 16S rRNA gene primers that target the genus *Bifidobacterium* (g-Bif-F and g-Bif-R) or universal 16S rRNA primers (27F and 1492R). All primers were purchased from Integrated DNA Technologies (IDT, Coralville, IA, USA).

3 RESULTS

The first step in assessing the suitability of functionally differentiating *B. longum infantis* from other *B. longum* subspecies was to assess the phylogenetic relationships among members of the *Bifidobacterium* clade. To accomplish this, we measured the average nucleotide identity (ANI) and shared gene content of all (387) publicly available *B. longum* genomes (Supplemental Table 1). Four *B. breve* genomes (GCF_000158015.1, GCF_000213865.1, GCF_000220135.1, GCF_000226175.1) were included as an outgroup. Genomes in this analysis were stratified by the (sub)species in each pairwise comparison (Figure 1). As expected, genomes of the same species were the most similar, and within *B. longum,* subspecies were all >95% similar to one another (Supplementary Table 2). Subspecies clusters (*B. longum*, *B. infantis*, and *B. suis*) were clearly distinct by ANI, though substantially overlap in terms of shared gene content. By both metrics, *B. longum longum* and *B. longum infantis* were the most similar of the subspecies (Figure 1). Interestingly, each of these subspecies appears to separate into two groups by ANI. As expected, comparisons between the outgroup *B. breve* genomes and *B. longum* subspecies were the least similar.

Though *B. infantis* is resolvable from other *B. longum* subspecies by ANI, performing ANI calculations on metagenome assemblies is computationally expensive and requires high sequencing depth (Jain et al., 2018). We therefore set out to determine whether individual genes might be suitable for distinguishing these subspecies. We explored the possibility that the presence of five genomic loci containing HMO metabolizing genes could be used as a proxy for subspecies differences (Figure 2). In order to do so, we used BLAST to quantify the presence of the five subspecies-specific gene clusters within the same 387 *Bifidobacterium* genomes. Genomes previously characterized as *B. infantis* largely corresponded with both ANI-based clustering and with the presence of HMO-metabolizing genes, with 36 of 43 *B. infantis*-labeled genomes (Supplementary Table 1), found in a distinct clade (Figure 2, top clade) and also containing at least 93% of HMO-metabolizing genes. Crucially, seven out of these 43 *B. infantis*-labeled genomes including 5 strains (CCUG 52486, CECT 7210, JCM 11660, KCTC 5934, 157F) do not contain the majority of HMO-metabolizing genes, but also do not cluster with other *B. infantis* genomes based on ANI, suggesting that they may have been misclassified. These results persist when HMO-metabolizing genes were excluded from ANI calculations (Figure 2, Supplementary Table 4), precluding the possibility that the ANI similarities are solely attributable to the presence or absence of those genes. Conversely, five unspecified *B. longum*

genomes (GCF_900157085.1, GCF_900157065.1, GCF_900157185.1, GCF_900157075.1, GCF_902381625.1) clustered tightly with those from *B. infantis* taxa and should, perhaps, be more specifically labeled as this particular subspecies (Figure 2). These genomes, although labeled as *B. longum,* contained HMO-metabolizing genes characteristic for *B. infantis*.

The congruence of ANI and presence of HMO-metabolizing genes suggests that *B. infantis* diverged from other *B. longum* subspecies since the acquisition of these genes. In order to explore the history of the HMO-metabolizing gene clusters in more detail, we investigated their presence within and beyond the *Bifidobacterium* genus. We used BLAST to search for homologs of the 78 protein-coding genes from the cluster against all available proteins in the NCBI database. Hits (see methods) for each of the HMO-metabolizing genes came largely from *B. longum* and *B. breve* genomes (Figure 3A, Supplementary Table 5). Cluster 1 displayed the highest percentage of hits to genomes from both of these species at 64% and 24% respectively.

To explore the evolutionary pressures that might have driven the phylogenetic distribution of the HMO-metabolizing clusters and their specificity to *Bifidobacterium* taxa, we compared nucleotide differences in the substitution rates at nonsynonymous (dN) and synonymous sites (dS) using MEGA X (Kumar et al., 2018; Stecher et al., 2020). A significant difference in the relative abundance of synonymous and nonsynonymous substitutions implies non-neutral selection, with a higher ratio of nonsynonymous substitutions indicating positive selection and a lower ratio indicating purifying selection (Kryazhimskiy and Plotkin, 2008). Of the 78 genes across the five clusters, 61 have undergone stabilizing selection (Supplementary Table 6, Figure 3B). Twelve genes underwent positive selection and five did not show evidence of selection (neutral selection) (Figure 3B).

After establishing that the five HMO-metabolizing gene clusters can distinguish *B. infantis* from the rest of the *B. longum* subspecies, we assessed the suitability of using HMO-metabolizing genes to generate primers specific to *B. infantis.* The development of the new primer sets was motivated by realizing that the existing primers to profile *B. infantis* were either too broad (e.g., 16S rRNA gene (Hong and Chen, 2007)) or too narrow (e.g., primers targeting sialidase gene (Lawley et al., 2017)). Forty six out of 79 genes covered were found to be in at least 85% of *B. infantis* strains. Two of these genes (Blon_0218 and Blon_0219) were found exclusively in the *B. infantis* clade (Figure 2, top clade) and might serve as potential future targets for primer development. Eleven additional genes (Blon_2322, Blon_2335, Blon_2336, Blon_2337, Blon_2338, Blon_2348, Blon_2349, Blon_2355, Blon_0220, Blon_0222, Blon_0232) were present in all genomes within the *B. infantis* clade and one additional *B. infantis* genome (GCF_902167615.1) right outside the clade. This genome (GCF_902167615.1) contains a majority of HMO-metabolizing genes found in Cluster 1 (Figure 2) and would be classified as *B. infantis* by our analysis as well if it were limited to this HMO-metabolizing gene cluster.

Out of the 11 genes, we selected four genes (Blon_2348, Blon_2355, Blon_2356 and Blon_2358) that aligned well with the clade of 41 genomes we consider to be *B. infantis* (Supplementary Table 1). Genes exo-alpha-sialidase (Blon_2348) and glycoside-hydrolase

(Blon_2355) were present in all genomes within the clade and one additional *B. infantis* genome (GCF_902167615.1) right outside of it. We also selected genes Blon_2356 and Blon_2358 that were present in 34 (83%) of the genomes within this clade and the additional *B. infantis* genome (GCF_902167615.1) as well. All four genes also had conserved regions among the *B. infantis*, making them good candidates for primer development.

Primer pairs for these four genes were tested *in silico* against the 387 downloaded genomes (Table 1, Supplementary Table 1). The primer specificity to the *B. infantis* clade matched clustering of strains based on ANI (Figure 2). The same six *B. infantis*-labeled genomes that failed to cluster with *B. infantis* by ANI provided no hits to our primer pairs (Supplementary Table 1), providing additional evidence that these *B. infantis*-named strains may have been misclassified. Similarly, the five unspecified *B. longum* genomes (GCF_900157085.1, GCF_900157065.1, GCF_900157185.1, GCF_900157075.1, GCF_902381625.1) that were clustered within the *B. infantis* clade (Figure 2) perfectly hybridized *in silico* to our primer sets.

To confirm the specificity of the designed primer pairs *in vitro*, PCR assays were performed with three *Bifidobacterium* reference strains, an E. coli K12 strain and a negative control. The amplification product was again specific to *B. infantis* (Table 1, the first 4 primer pairs). All four primers aligned 100% to each *B. infantis* strain and not to any other bacteria. These primers were then applied to verify *B. infantis* presence in sequenced gecal samples.

Finally, we used the five HMO-metabolizing gene clusters to distinguish *B. infantis* from the rest of the *B. longum* subspecies in the gut microbiomes of human children using fecal samples in the RESONANCE cohort of young, healthy children (Bonham et al., 2020). We applied the specificity of this HMO-metabolizing cluster to identify *B. infantis* from metagenomic analysis of our samples. Using shotgun metagenomic sequencing, we observed that 15 of 203 (7.4%) samples from children under the age of two harbor *B. infantis* (Figure 4). Each sample was taken from a unique child subject. The gut microbiomes of three of the subjects were sampled at later time points and no longer contained *B. infantis,* suggesting a decrease in abundance around one year of age. Lastly, we validated the metagenomic identification of one HMO-metabolizing gene (the sialidase gene Blon_2348) *in vitro* using PCR. We supplemented this analysis with tests for *B. breve* and the *Bifidobacterium* genus as a positive control in order to confirm our conclusions. Sequences belonging to the *Bifidobacterium* genus were found in the metagenomes of most (214 of 222 or 96%) of the samples from all children analyzed. Of these, 477 (93%) and 375 (91%) of samples from children and mothers, respectively, contained *Bifidobacterium*. Only 13 of 162 (8.0%) samples taken from children under two and 4 of 411 (1.0%) samples taken from mothers were found to have *B. infantis*.

4 DISCUSSION

The human gut microbiome, though better characterized than most microbial communities, remains poorly understood. This is especially true in very young children, whose gut microbial communities are seeded at birth, shaped profoundly by the method of their birth and the nature of their diet, and in constant flux. The identification nearly 91 years ago of a component of human breast milk (HMO) that appears to have evolved solely to feed the developing microbiomes of infants hints at the critical importance of microbiome development as part of overall child development (Kunz, 2012). Here, we have shown that, by focusing on the functional differences between microbial subspecies, in particular those differences that provide a critical metabolic function such as HMO utilization, we can increase taxonomic resolution through culture-independent means.

The *Bifidobacterium* genus is incredibly common in children's guts (96% of children have them in this study, as confirmed by PCR assay and metagenomic search), and many species harbor genes that allow them to make use of complex polysaccharides. Yet there are important functional differences between members of this genus, and even, as we have shown here, between the subspecies *B. longum longum* and *B. longum infantis*. This is not uncommon; a recent study showed an average of 13% (and as much as 25%) difference in gene content between common gut microbiome species from different individuals (Zhu et al., 2015). Using metagenomic sequencing and a gene function-centered approach allowed us to show that only 18 of 204 (8.8%) children under 2 years old in a cohort of US children harbored the subspecies *B. infantis*, and we showed that this is synonymous with harboring five HMO metabolizing gene clusters. These results are consistent with a recent study on microbial dysbiosis in American infants that showed remarkably low prevalence of HMO metabolizing capacity that accompanied other indicators of unhealthy microbiomes (Casaburi et al., 2021).

Other culture-independent methods for identifying microbial taxa have serious limitations for studying *Bifidobacterium infantis*. Amplicon (16S rRNA gene) sequencing is typically only capable of resolving genus-level differences between microbes. Though PCR amplification of genes of interest has been the gold standard for detecting them, PCR is not scalable to large numbers of samples or to identifying many gene functions at once. Here, we showed that metagenomic sequencing largely agrees with PCR-based amplification, at least for one HMO metabolizing gene. A custom primer pair designed to target the HMO-metabolizing sialidase gene, Blon_2348, in *B. infantis* was used in PCR in this study. Forty six of the 79 HMO-metabolizing genes were also found in all genomes within the *B. infantis* ANI cluster; a subset of these might be used for future primer development. Further using PCR to investigate the presence of many genes in multiple gene clusters, however, would have been impractical, and the concordance of dozens of genes using metagenomic sequencing lends increased credence to the presence of these genes in the samples identified through this method.

We were able to identify six genomes that were previously identified as *B. infantis* that should in fact be classified as *B. longum longum*, as well as five genomes that do not have subspecies resolution in Genbank that likely should be classified as *B. infantis.* The high concordance between average nucleotide identity and the presence of all 5 HMO utilization clusters suggests

that acquisition or loss of these genes occurred in a single common ancestor and continued divergence with positive selection due to occupying the HMO-utilization niche. But this need not always be the case; horizontal gene transfer (HGT) can lead to gene function sharing across species boundaries (Smillie et al., 2011; Brito et al., 2016). In human microbiome epidemiology, what likely matters in shaping health outcomes is functional variation, for which taxonomic profiling is an imperfect proxy. We believe that a function-based metagenomic approach is essential for untangling the complexity of the gut microbiome and its relationship to human health.

AUTHOR CONTRIBUTIONS

LT, KB, and VKC designed the study; the samples were collected as a part of an ongoing NIH ECHO study done by RESONANCE consortium, SR and AF processed the samples; LT and KB designed and wrote the code and analyzed the data; LT, KB, and VKC wrote, edited, and finalized the manuscript. All authors approved the manuscript's final version.

ACKNOWLEDGMENTS

We thank families participating in the RESONANCE cohort. We thank Libusha Kelly and Sarah Comstock for reviewing the manuscript and providing helpful comments. We thank Ivica Ceraj for help with data analysis. The project was funded by the NIH UG3 OD023313 (VKC) and the American Society for Microbiology Undergraduate Research Fellowship (LT).

DATA AND CODE AVAILABILITY

Metagenome data were deposited in NCBI's Sequence Read Archive (SRA), under BioProject PRJNA695570 and accession numbers SAMN17618458-9435. Code for data analysis, statistics, and figures generated for this manuscript is available on OSF.io (Bonham et al., 2021).

SUPPLEMENTARY MATERIAL

The supplementary material for this manuscript can be found online at:

CONFLICT OF INTEREST

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

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2.

Bonham, K. S., Bruchhage, M. M. K., Rowland, S., Volpe, A. R., Dyer, K., Consortium, R., et al. (2020). Gut microbes and their genes are associated with brain development and cognitive function in healthy children. *bioRxiv*, 2020.02.13.944181. doi:10.1101/2020.02.13.944181.

Bonham, K., Tso, L., and Klepac-Ceraj, V. (2021). Bifido_HMO. doi:10.17605/OSF.IO/ZKVJH.

Bonjoch, X., Balleste, E., and Blanch, A. R. (2004). Multiplex PCR with 16S rRNA Gene-Targeted Primers of Bifidobacterium spp. To Identify Sources of Fecal Pollution. *APPL. ENVIRON. MICROBIOL.* 70, 6.

Brito, I. L., Yilmaz, S., Huang, K., Xu, L., Jupiter, S. D., Jenkins, A. P., et al. (2016). Mobile genes in the human microbiome are structured from global to individual scales. *Nature* 535, 435–439. doi:10.1038/nature18927.

Casaburi, G., Duar, R. M., Brown, H., Mitchell, R. D., Kazi, S., Chew, S., et al. (2021). Metagenomic insights of the infant microbiome community structure and function across multiple sites in the United States. *Scientific Reports* 11, 1472. doi:10.1038/s41598-020-80583-9.

Chaplin, A. V., Efimov, B. A., Smeianov, V. V., Kafarskaia, L. I., Pikina, A. P., and Shkoporov, A. N. (2015). Intraspecies Genomic Diversity and Long-Term Persistence of Bifidobacterium longum. *PLOS ONE* 10, e0135658. doi:10.1371/journal.pone.0135658.

Desbonnet, L., Garrett, L., Clarke, G., Kiely, B., Cryan, J. F., and Dinan, T. G. (2010). Effects of the probiotic Bifidobacterium infantis in the maternal separation model of depression. *Neuroscience* 170, 1179–1188. doi:10.1016/j.neuroscience.2010.08.005.

Duar, R. M., Henrick, B. M., Casaburi, G., and Frese, S. A. (2020). Integrating the Ecosystem Services Framework to Define Dysbiosis of the Breastfed Infant Gut: The Role of B. infantis and Human Milk Oligosaccharides. *Front Nutr* 7. doi:10.3389/fnut.2020.00033.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32, 1792–1797. doi:10.1093/nar/gkh340.

Frese, S. A., Hutton, A. A., Contreras, L. N., Shaw, C. A., Palumbo, M. C., Casaburi, G., et al. (2017). Persistence of Supplemented Bifidobacterium longum subsp. infantis EVC001 in Breastfed Infants. *mSphere* 2, mSphere.00501-17, e00501-17. doi:10.1128/mSphere.00501-17.

Garrido, D., Kim, J. H., German, J. B., Raybould, H. E., and Mills, D. A. (2011). Oligosaccharide Binding Proteins from Bifidobacterium longum subsp. infantis Reveal

a Preference for Host Glycans. *PLOS ONE* 6, e17315. doi:10.1371/journal.pone.0017315.

Garrido, D., Ruiz-Moyano, S., Lemay, D. G., Sela, D. A., German, J. B., and Mills, D. A. (2015). Comparative transcriptomics reveals key differences in the response to milk oligosaccharides of infant gut-associated bifidobacteria. *Scientific Reports* 5, 13517. doi:10.1038/srep13517.

Gertz, E. M., Yu, Y.-K., Agarwala, R., Schäffer, A. A., and Altschul, S. F. (2006). Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. *BMC Biol* 4, 41. doi:10.1186/1741-7007-4-41.

Gut microbes and their genes are associated with brain development and cognitive function in healthy children | bioRxiv Available at: https://www.biorxiv.org/content/10.1101/2020.02.13.944181v3 [Accessed February 11, 2021].

Huda, M. N., Lewis, Z., Kalanetra, K. M., Rashid, M., Ahmad, S. M., Raqib, R., et al. (2014). Stool Microbiota and Vaccine Responses of Infants. *Pediatrics* 134, e362–e372. doi:10.1542/peds.2013-3937.

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* 9, 5114. doi:10.1038/s41467-018-07641-9.

Johnson, J. S., Spakowicz, D. J., Hong, B.-Y., Petersen, L. M., Demkowicz, P., Chen, L., et al. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun* 10, 1–11. doi:10.1038/s41467-019-13036-1.

Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30, 772–780. doi:10.1093/molbev/mst010.

Kim, H.-B., Kim, E., Yang, S.-M., Lee, S., Kim, M.-J., and Kim, H.-Y. (2020). Development of Real-Time PCR Assay to Specifically Detect 22 Bifidobacterium Species and Subspecies Using Comparative Genomics. *Front. Microbiol.* 11. doi:10.3389/fmicb.2020.02087.

Kryazhimskiy, S., and Plotkin, J. B. (2008). The Population Genetics of dN/dS. *PLOS Genetics* 4, e1000304. doi:10.1371/journal.pgen.1000304.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* 35, 1547–1549. doi:10.1093/molbev/msy096.

Kunz, C. (2012). Historical Aspects of Human Milk Oligosaccharides1234. *Adv Nutr* 3, 430S-439S. doi:10.3945/an.111.001776.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, 9.

Le Doare, K., Holder, B., Bassett, A., and Pannaraj, P. S. (2018). Mother's Milk: A Purposeful Contribution to the Development of the Infant Microbiota and Immunity. *Front. Immunol.* 9. doi:10.3389/fimmu.2018.00361.

Lebreton, F., Manson, A. L., Saavedra, J. T., Straub, T. J., Earl, A. M., and Gilmore, M. S. (2017). Tracing the Enterococci from Paleozoic Origins to the Hospital. *Cell* 169, 849-861.e13. doi:10.1016/j.cell.2017.04.027.

LoCascio, R. G., Desai, P., Sela, D. A., Weimer, B., and Mills, D. A. (2010). Broad Conservation of Milk Utilization Genes in Bifidobacterium longum subsp. infantis as Revealed by Comparative Genomic Hybridization. *Applied and Environmental Microbiology* 76, 7373–7381. doi:10.1128/AEM.00675-10.

Lugli, G. A., Milani, C., Turroni, F., Duranti, S., Ferrario, C., Viappiani, A., et al. (2014). Investigation of the Evolutionary Development of the Genus Bifidobacterium by Comparative Genomics. *Appl. Environ. Microbiol.* 80, 6383–6394. doi:10.1128/AEM.02004-14.

Marcobal, A., Barboza, M., Froehlich, J. W., Block, D. E., German, J. B., Lebrilla, C. B., et al. (2010). Consumption of Human Milk Oligosaccharides by Gut-related Microbes. *J Agric Food Chem* 58, 5334–5340. doi:10.1021/jf9044205.

Marcobal, A., Barboza, M., Sonnenburg, E. D., Pudlo, N., Martens, E. C., Desai, P., et al. (2011). Bacteroides in the Infant Gut Consume Milk Oligosaccharides via Mucus-Utilization Pathways. *Cell Host & Microbe* 10, 507–514. doi:10.1016/j.chom.2011.10.007.

Matsuki, T., Watanabe, K., Fujimoto, J., Kado, Y., Takada, T., Matsumoto, K., et al. (2004). Quantitative PCR with 16S rRNA-Gene-Targeted Species-Specific Primers for Analysis of Human Intestinal Bifidobacteria. *Applied and Environmental Microbiology* 70, 167–173. doi:10.1128/AEM.70.1.167-173.2004.

Mattarelli, P., Bonaparte, C., Pot, B., and Biavati, B. (2008). Proposal to reclassify the three biotypes of Bifidobacterium longum as three subspecies: Bifidobacterium longum subsp. longum subsp. nov., Bifidobacterium longum subsp. infantis comb. nov. and Bifidobacterium longum subsp. suis comb. nov. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* 58, 767–772. doi:10.1099/ijs.0.65319-0.

McIver, L. J., Abu-Ali, G., Franzosa, E. A., Schwager, R., Morgan, X. C., Waldron, L., et al. (2018). bioBakery: a meta'omic analysis environment. *Bioinformatics* 34, 1235–1237. doi:10.1093/bioinformatics/btx754.

Milani, C., Lugli, G. A., Duranti, S., Turroni, F., Bottacini, F., Mangifesta, M., et al. (2014). Genomic Encyclopedia of Type Strains of the Genus Bifidobacterium. *Appl. Environ. Microbiol.* 80, 6290–6302. doi:10.1128/AEM.02308-14.

Milani, C., Turroni, F., Duranti, S., Lugli, G. A., Mancabelli, L., Ferrario, C., et al. (2016). Genomics of the Genus Bifidobacterium Reveals Species-Specific Adaptation to the Glycan-Rich Gut Environment. *Appl. Environ. Microbiol.* 82, 980–991. doi:10.1128/AEM.03500-15.

Polz, M. F., Alm, E. J., and Hanage, W. P. (2013). Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends in Genetics* 29, 170–175. doi:10.1016/j.tig.2012.12.006.

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24. doi:10.1039/C5AY02550H.

Robinson, E. R., Walker, T. M., and Pallen, M. J. (2013). Genomics and outbreak investigation: from sequence to consequence. *Genome Medicine* 5, 36. doi:10.1186/gm440.

Ruiz-Moyano, S., Totten, S. M., Garrido, D. A., Smilowitz, J. T., German, J. B., Lebrilla, C. B., et al. (2013). Variation in Consumption of Human Milk Oligosaccharides by Infant Gut-Associated Strains of Bifidobacterium breve. *Appl. Environ. Microbiol.* 79, 6040–6049. doi:10.1128/AEM.01843-13.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406–425. doi:10.1093/oxfordjournals.molbev.a040454.

Sakata, S. (2002). Unification of Bifidobacterium infantis and Bifidobacterium suis as Bifidobacterium longum. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* 52, 1945–1951. doi:10.1099/ijs.0.02221-0.

Satokari, R. M., Vaughan, E. E., Favier, C. F., Doré, J., Edwards, C., and Vos, W. M. de (2002). Diversity of Bifidobacterium and Lactobacillus spp. in Breast-Fed and Formula-Fed Infants as Assessed by 16S rDNA Sequence Differences. *Microbial Ecology in Health and Disease* 14, 97–105. doi:10.1080/08910600260081748.

Sela, D. A. (2011). Bifidobacterial utilization of human milk oligosaccharides. *International Journal of Food Microbiology* 149, 58–64. doi:10.1016/j.ijfoodmicro.2011.01.025.

Sela, D. A., Chapman, J., Adeuya, A., Kim, J. H., Chen, F., Whitehead, T. R., et al. (2008). The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. *Proceedings of the National Academy of Sciences* 105, 18964–18969. doi:10.1073/pnas.0809584105.

Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., and Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480, 241–244. doi:10.1038/nature10571.

Stecher, G., Tamura, K., and Kumar, S. (2020). Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular Biology and Evolution* 37, 1237–1239. doi:10.1093/molbev/msz312.

Sun, Z., Zhang, W., Guo, C., Yang, X., Liu, W., Wu, Y., et al. (2015a). Comparative Genomic Analysis of 45 Type Strains of the Genus Bifidobacterium: A Snapshot of Its Genetic Diversity and Evolution. *PLOS ONE* 10, e0117912. doi:10.1371/journal.pone.0117912.

Sun, Z., Zhang, W., Guo, C., Yang, X., Liu, W., Wu, Y., et al. (2015b). Comparative Genomic Analysis of 45 Type Strains of the Genus Bifidobacterium: A Snapshot of Its Genetic Diversity and Evolution. *PLOS ONE* 10, e0117912. doi:10.1371/journal.pone.0117912.

Thongaram, T., Hoeflinger, J. L., Chow, J., and Miller, M. J. (2017). Human milk oligosaccharide consumption by probiotic and human-associated bifidobacteria and lactobacilli. *Journal of Dairy Science* 100, 7825–7833. doi:10.3168/jds.2017-12753.

Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., et al. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods* 12, 902–903. doi:10.1038/nmeth.3589.

Underwood, M. A., German, J. B., Lebrilla, C. B., and Mills, D. A. (2015). Bifidobacterium longum subspecies infantis : champion colonizer of the infant gut. *Pediatric Research* 77, 229–235. doi:10.1038/pr.2014.156.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research* 35, W71–W74. doi:10.1093/nar/gkm306.

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol* 20, 257. doi:10.1186/s13059-019-1891-0.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13, 134. doi:10.1186/1471-2105-13-134.

Youn, S. Y., Seo, J. M., and Ji, G. E. (2008). Evaluation of the PCR method for identification of Bifidobacterium species. *Letters in Applied Microbiology* 46, 7–13. doi:10.1111/j.1472-765X.2007.02263.x.

Yu, Z.-T., Chen, C., and Newburg, D. S. (2013). Utilization of major fucosylated and sialylated human milk oligosaccharides by isolated human gut microbes. *Glycobiology* 23, 1281–1292. doi:10.1093/glycob/cwt065.

Zhu, A., Sunagawa, S., Mende, D. R., and Bork, P. (2015). Inter-individual differences in the gene content of human gut bacterial species. *Genome Biol* 16, 82. doi:10.1186/s13059-015-0646-9.

TABLES

Table 1: Primers targeting HMO metabolizing genes and 16S rRNA.

exo-alpha-sialidase (Blon_2348)	*B. longum infantis*	Sia-266F	GACGAGGAGGAATACAGCAG	100	This study	410
		Sia-676R	CACGAACAGCGAATCATGGATT			
glycoside-hydrolase (Blon_2355)	*B. longum infantis*	GH-750F	GCGCCATCCTGGTGATGTTATT	100	This study	498
		GH-1248R	CTACGTGATCTGGGAGAGTTTC			
haloacid dehalogenase domain protein hydrolase (Blon_2356)	*B. longum infantis*	HH-60F	CCACAATGTCATCGACCATCTG	83	This study	474
		HH-534R	CCGAAGTATTCGGATGCCTATG			
glycoside hydrolase (Blon_2358)	*B. longum infantis*	GH-492F	CGATGATGTGCTGGATTCGTTC	74	This study	510
		GH-1002R	CTCGACCATTCCAAGATGCTCA			
Sialidase	*B. longum infantis*	Inf 2348-F	ATACAGCAGAACCTTGGCCT	50	Lawley et al. (2017)	280
		Inf 2348-R	GCGATCACATGGACGAGAAC			
Major Facilitator Superfamily	*B. longum infantis*	Blon0915-F	CGTATTGGCTTTGTACGCATTT	71	Frese et al. (2017)	118
		Blon0915-R	ATCGTGCCGGTGAGATTTAC			
16S rRNA	*B. longum infantis*, *B. longum longum*, *B. indicum*	Binf-F	CCATCTCTGGGATCGTCGG		Tannock et al. (2013)	562
		Binf-R	TATCGGGGAGCAAGCGTGA			
16S rRNA	*B. infantis*	HCBin-F	AGGATACGTTCGGCGTC		Hong and Chen (2007)	377
		HCBin-R	CGCAAGATTCCTCTAGCA			
16S rRNA	*Bifidobacterium*	g-Bif-F	CTCCTGGAAACGGGTGG		Tannock et al. (2013)	550
		g-Bif-R	GGTGTTCTTCCCGATATCTACA			
16S rRNA	All bacteria	Uni-F	ACTCCTACGGGAGGCAGCAGT		Wang et al. (1996)	174
		Uni-R	ATTACCGCGGCTGCTGGC			

[1]Percent of exact primer match to 41 *B. infantis* strains found in Clade 1 (Figure2)

[2]Amplicon size in base pairs

FIGURE LEGENDS

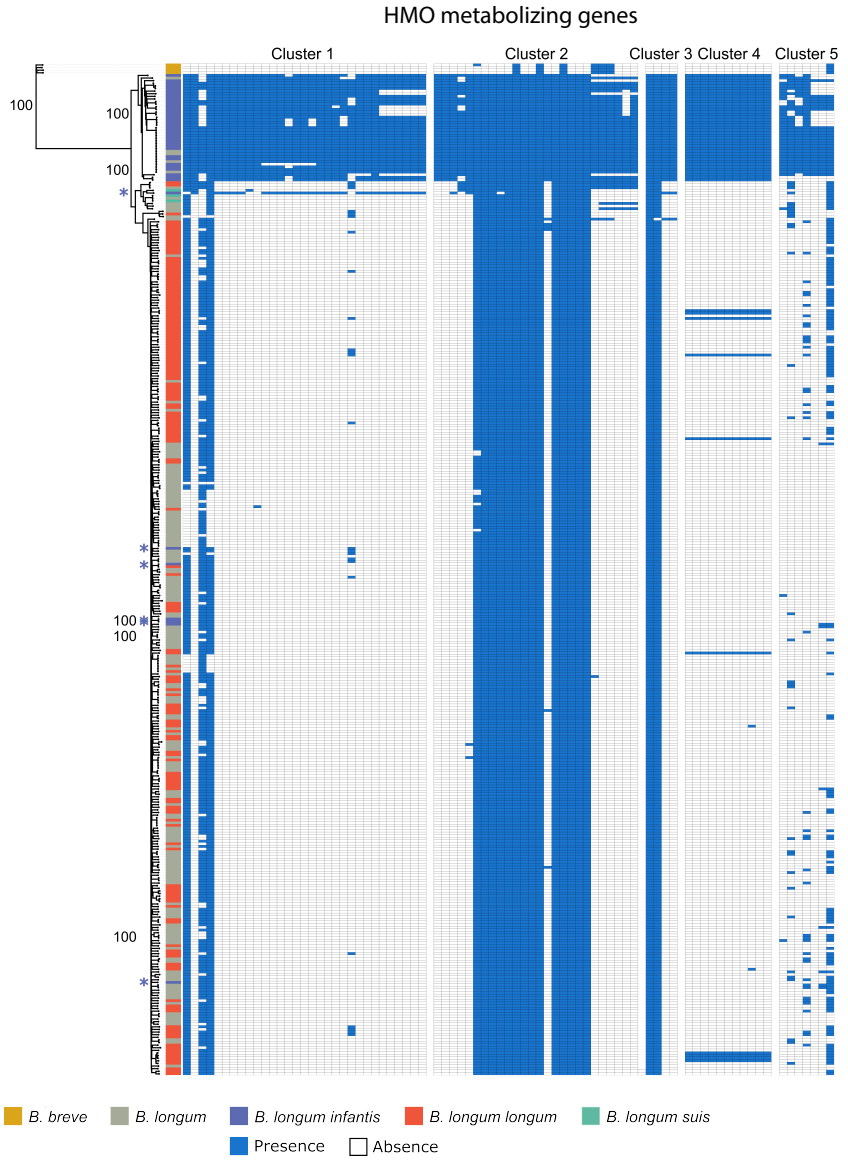
Figure 1. *B. infantis* genomes cluster separately from *B.longum.*
Scatter plot of percent shared gene content versus ANI values across pairwise comparisons of *Bifidobacterium* genomes. Shared gene content calculated from BLAST and ANI values calculated by the python package pyani. Presence of HMO-metabolizing genes determined by the occurrence of at least 60 of 79 (75%) genes in a genome. Dashed line indicates the ANI cutoff score of 95% for genomes belonging to the same species.

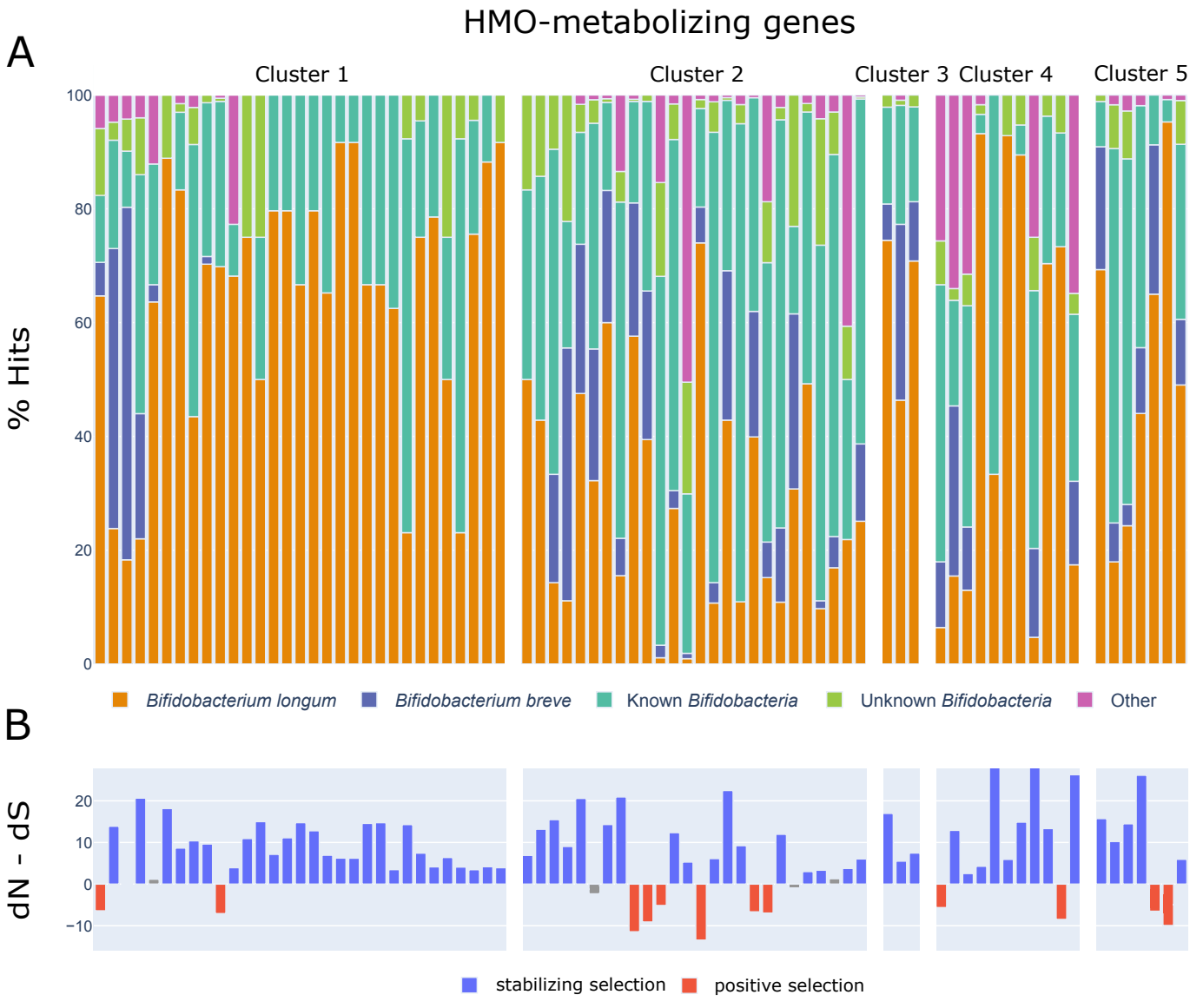
Figure 2. HMO-metabolizing gene clusters mirror phylogenetic separation of *B. infantis*.
Rooted neighbor joining tree of the average nucleotide identity (ANI) values of the same 387 *Bifidobacterium* genomes with HMO-metabolizing genes removed (Supplementary Table 4). The tree is paired with a heatmap denoting the presence of HMO-metabolizing genes by BLASTN alignments (>95% identity). Potentially misclassified *B. infantis* genomes noted with a purple asterisk.

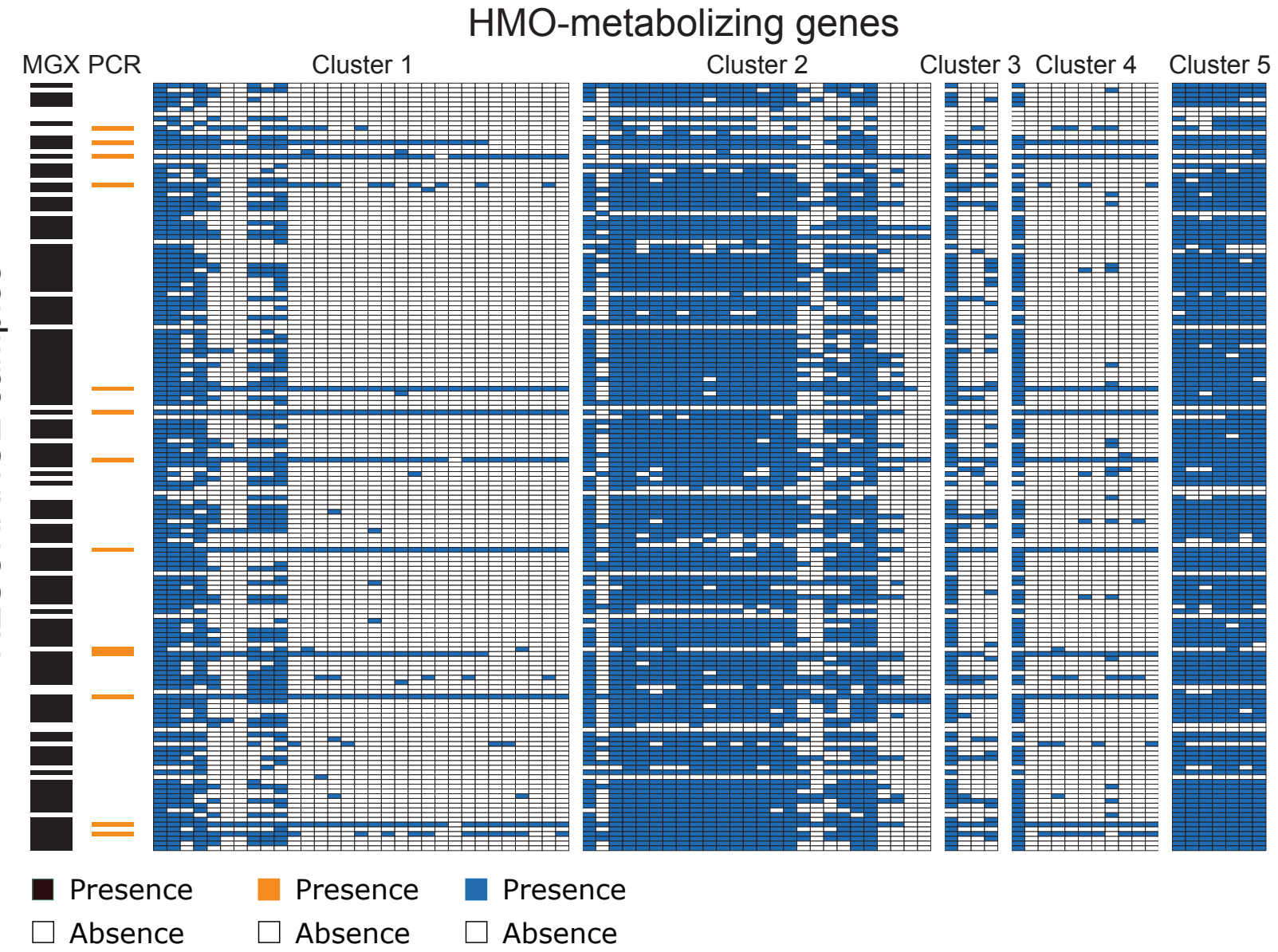
Figure 3. HMO-metabolizing genes are largely specific to the *Bifidobacterium* genus.
(A) Stacked bar plot showing taxonomy of BLASTP alignments from each HMO-metabolizing gene to taxa contained in NCBI's non-redundant database. Known *Bifidobacterium* include any specified species in the *Bifidobacterium* genus other than *B. longum* or *B. breve*. Unknown *Bifidobacterium* include any unspecified *Bifidobacterium* taxa (Bifidobacterium, Bifidobacterium sp.). (B) Statistics from dN - dS analysis showing the probability and certainty that a given gene within the HMO-metabolizing gene clusters are under stabilizing or positive selection. Colored bars indicate statistical significance ($p<0.05$). Analyses were conducted using the Nei-Gojobori method in MEGA X (Kumar et al., 2018; Stecher et al., 2020). All ambiguous positions were removed for each sequence pair (pairwise deletion option).

Figure 4. *B. infantis* presence observed in 18 of 204 (8.8%) samples from children under the age of two via PCR or metagenomics.
Results from tests on samples from children under the age of two for the presence of HMO-metabolizing genes either via metagenomics (n=203) or PCR (n=162). The metagenomic presence of *B. infantis* (n=15) was determined by mapped Bowtie2 alignments to HMO-metabolizing genes. The PCR presence of *B. infantis* (n=13) was determined by PCR using primers designed to target the gene Blon_2348 unique to the subspecies, with presence noted in orange and absence in white. Presence of *Bifidobacterium longum* calculated by MetaPhlAn is also noted on the left hand side in the MGX column in black and absence in white.

Average nucleotide identity (%)
100
98
96
94
92
90
50
60
70
80
90
100
Shared gene content (%)
Genome with HMO-metabolizing genes
Genome without HMO-metabolizing genes
B. longum infantis v. B. longum infantis
B. breve v. B. longum infantis
B. longum longum v. B. longum longum
B. breve v. B. longum longum
B. longum infantis v. B. longum longum


HMO metabolizing genes
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
100
100
100
100
100
100
B. breve
B. longum
B. longum infantis
B. longum longum
B. longum suis
Presence
Absence


HMO-metabolizing genes
A
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
% Hits
100
80
60
40
20
0
Bifidobacterium longum
Bifidobacterium breve
Known Bifidobacteria
Unknown Bifidobacteria
Other
B
dN - dS
20
10
0
−10
stabilizing selection
positive selection


HMO-metabolizing genes
MGX
PCR
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Presence
Absence
Presence
Absence
Presence
Absence

Gene	Target organism	Name	Sequence (5'-3')	% Coverage[1]	References	Length (bp)[2]
exo-alpha-sialidase (Blon_2348)	*B. longum infantis*	Sia-266F	GACGAGGAGGAATACAGCAG	100	This study	410
		Sia-676R	CACGAACAGCGAATCATGGATT			
glycoside-hydrolase (Blon_2355)	*B. longum infantis*	GH-750F	GCGCCATCCTGGTGATGTTATT	100	This study	498
		GH-1248R	CTACGTGATCTGGGAGAGTTTC			
haloacid dehalogenase domain protein hydrolase (Blon_2356)	*B. longum infantis*	HH-60F	CCACAATGTCATCGACCATCTG	83	This study	474
		HH-534R	CCGAAGTATTCGGATGCCTATG			
glycoside hydrolase (Blon_2358)	*B. longum infantis*	GH-492F	CGATGATGTGCTGGATTCGTTC	74	This study	510
		GH-1002R	CTCGACCATTCCAAGATGCTCA			
Sialidase	*B. longum infantis*	Inf 2348-F	ATACAGCAGAACCTTGGCCT	50	Lawley et al. (2017)	280
		Inf 2348-R	GCGATCACATGGACGAGAAC			
Major Facilitator Superfamily	*B. longum infantis*	Blon0915-F	CGTATTGGCTTTGTACGCATTT	71	Frese et al. (2017)	118
		Blon0915-R	ATCGTGCCGGTGAGATTTAC			
16S rRNA	*B. longum infantis, B. longum longum, B. indicum*	Binf-F	CCATCTCTGGGATCGTCGG		Tannock et al. (2013)	562
		Binf-R	TATCGGGGAGCAAGCGTGA			
16S rRNA	*B. infantis*	HCBin-F	AGGATACGTTCGGCGTC		Hong and Chen (2007)	377
		HCBin-R	CGCAAGATTCCTCTAGCA			
16S rRNA	*Bifidobacterium*	g-Bif-F	CTCCTGGAAACGGGTGG		Tannock et al. (2013)	550
		g-Bif-R	GGTGTTCTTCCCGATATCTACA			
16S rRNA	All bacteria	Uni-27F	ACTCCTACGGGAGGCAGCAGT		Wang et al. (1996)	174
		Uni-1492R	ATTACCGCGGCTGCTGGC			

[1] Percent of exact alignment to 42 *B. infantis* strains

[2] Amplicon size in base pairs