



Identification of Differentiating Metabolic Pathways between Infant Gut Microbiome Populations Reveals Depletion of Function-Level Adaptation to Human Milk in the Finnish Population

 Jan Majta,^{a,b}  Krzysztof Odrzywolek,^{a,c}  Bozena Milanovic,^{a,d} Vladyslav Hubar,^{a,e}  Sonia Wrobel,^{a,f} Emilia Strycharz-Angrecka,^a Szymon Wojciechowski,^a  Kaja Milanowska^a

^aArdigen, Krakow, Poland

^bDepartment of Computational Biophysics and Bioinformatics, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Krakow, Poland

^cDepartment of Computer Science, Faculty of Computer Science, Electronics and Telecommunications, AGH University of Science and Technology, Krakow, Poland

^dDepartment of Biophysics, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Krakow, Poland

^eFaculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Krakow, Poland

^fDepartment of Medical Physics, Institute of Physics, Astronomy and Applied Computer Sciences, Jagiellonian University, Krakow, Poland

ABSTRACT A variety of autoimmune and allergy events are becoming increasingly common, especially in Western countries. Some pieces of research link such conditions with the composition of microbiota during infancy. In this period, the predominant form of nutrition for gut microbiota is oligosaccharides from human milk (HMO). A number of gut-colonizing strains, such as *Bifidobacterium* and *Bacteroides*, are able to utilize HMO, but only some *Bifidobacterium* strains have evolved to digest the specific composition of human oligosaccharides. Differences in the proportions of the two genera that are able to utilize HMO have already been associated with the frequency of allergies and autoimmune diseases in the Finnish and the Russian populations. Our results show that differences in terms of the taxonomic annotation do not explain the reason for the differences in the *Bifidobacterium/Bacteroides* ratio between the Finnish and the Russian populations. In this paper, we present the results of function-level analysis. Unlike the typical workflow for gene abundance analysis, BiomeScout technology explains the differences in the *Bifidobacterium/Bacteroides* ratio. Our research shows the differences in the abundances of the two enzymes that are crucial for the utilization of short type 1 oligosaccharides.

IMPORTANCE Knowing the limitations of taxonomy-based research, there is an emerging need for the development of higher-resolution techniques. The significance of this research is demonstrated by the novel method used for the analysis of function-level metagenomes. BiomeScout—the presented technology—utilizes proprietary algorithms for the detection of differences between functionalities present in metagenomic samples.

KEYWORDS bioinformatics, gut microbiome, infant microbiome, machine learning, microbiome

Depletion of oligosaccharides from human milk (HMO)-adapted bifidobacterial strains reduces their ability to dominate the gut microbiota in the presence of human milk. It allows other taxa, e.g., *Bacteroides*, to reach the dominant position. The lack of some strains of the *Bifidobacterium* genus has been linked by other researchers with the presence of more genes that encode virulence factors. Additionally, the extensive growth of *Bacteroides* has been described as a potential factor for the development of allergies and autoimmunity. Therefore, the described findings support

Citation Majta J, Odrzywolek K, Milanovic B, Hubar V, Wrobel S, Strycharz-Angrecka E, Wojciechowski S, Milanowska K. 2019. Identification of differentiating metabolic pathways between infant gut microbiome populations reveals depletion of function-level adaptation to human milk in the Finnish population. *mSphere* 4:e00152-19. <https://doi.org/10.1128/mSphereDirect.00152-19>.

Editor Katherine McMahon, University of Wisconsin—Madison

Copyright © 2019 Majta et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Kaja Milanowska, kaja.milanowska@ardigen.com.

Solicited external reviewers: Zbigniew Lazar, Wrocław University of Environmental and Life Sciences; Chengsheng Zhu, Bromberg Lab, Rutgers University.

This paper was submitted via the [mSphereDirect™](https://mspheredirect.com) pathway.

 Function-level analysis done with BiomeScout reveals foundation of infant microbiota dysbiosis. @janmajta @kaja_milanowska

Received 1 March 2019

Accepted 6 March 2019

Published 20 March 2019

the hypothesis that the depletion of HMO-utilizing bifidobacterial strains causes low levels of *Bifidobacterium* in Finnish infants' microbiota.

Importance of the gut microbiome. Understanding of the role of commensal microbiota in human health and well-being has developed over the past few years (1–3). The gastrointestinal (GI) tract community is the most deeply investigated microbiome of the human body. A number of pieces of research have indicated its role in the development of various diseases, including but not limited to GI-related conditions like irritable bowel disease (4, 5) or Crohn's disease (6–8). Recent studies have shown the impact of commensal microbiota on the immune system in processes like cancer immunotherapy (9–11) or the development of autoimmune conditions (12–14). Nowadays, a variety of autoimmune and allergy events are becoming increasingly common (15). A well-established hygiene hypothesis connects such diseases with life conditions during the developmental period of life (16). One proposed hypothesis that explains the molecular background of autoimmunity and allergy development is that GI bacteria stimulate the immune system during early childhood (14).

It has now been demonstrated that the compositions of GI microbiota differ in the various populations of infants in the north European cohort. A study by Vatanen et al. showed that *Bacteroides* are a major genus in the Finnish population, while *Bifidobacterium* spp. dominate Russian guts. Nonetheless, it remains unsolved why such a difference occurs as both clades are present in both groups but with different abundances. A potential explanation for such a phenomenon may be the fact that the *Bifidobacterium* strains that colonize the Finnish population do not act as strong HMO metabolizers. Due to poor adaptation to small type 1 oligosaccharides, the GI tracts of infants may be not an optimal niche for them (17). On the other hand, some *Bacteroides* strains are able to take advantage of the presence of some HMOs and outgrow competing *Bifidobacterium* (18).

Infant gut microbiota characteristics. Various studies have reported a spectrum of variants of the composition of GI microbiota that are typical of infants (19). Despite these differences, the bacterial genera that are present in most research are *Bifidobacterium*, *Bacteroides*, *Escherichia*, and *Lactobacillus* (20). Multiple pieces of research have reported an association between the relative abundances of clades and various development conditions.

The infant microbiome profile relies on multiple factors. The most important are delivery type, antibiotic usage in both mother and infant, and formula feeding (21–23). The initial colonization of the infant takes place during and shortly after delivery as different communities are transferred from the mother's body depending on whether birth is performed via vaginal delivery (predominantly vaginal lactic acid bacteria) or C-section (predominantly skin-bound *Staphylococcus*) (24). Furthermore, the history of the mother's antibiotic usage determines the communities that may be transferred to a newborn (23).

During the 20th century, the composition of the GI microbiota of infants gradually changed. This can be noticed in comparisons of the pH levels of infant stool samples reported in a number of publications from the 20th century. Additionally, analyzing historical descriptions of the fecal microbiome, it can be easily noticed that there are clear differences (25) between the preantibiotic era and current research.

It seems that of the multiple species that are commensal to infants, *Bifidobacterium* spp. have the strongest association with infants' health. Additionally, a significant amount of research reports that *Bifidobacterium* spp. represent the original inhabitants of infants' GI tract (25–27). It was already proven that the presence of specific *Bifidobacterium* bacterial strains is correlated with a low level of virulence factor genes in the whole microbiota of infants (28).

The research by Vatanen et al. on the DIABIMMUNE (<http://www.diabimmune.org/>) cohort showed differences in autoimmunity and allergy events. The paper concluded that differing compositions of GI microbiota are responsible for this difference (14).

HMO as a key prebiotic. The primary food of a newborn is breast milk. Apart from water, breast milk is mostly composed of carbohydrates, and the predominant sugar present in breast milk is lactose (around 70 g/liter) (29). Breast milk also contains indigestible oligosaccharides that act as prebiotics for coevolved GI microbiota members—human milk oligosaccharides (HMOs)—that are present with an approximate concentration of 20 g/liter. HMOs can be primarily divided into two groups: type 1 oligosaccharides, e.g., lacto-n-biose or lacto-n-tetraose, which are characterized by Gal β 1-3GlcNAc linkage, and type 2 HMOs, which possess Gal β 1-4GlcNAc linkage (17). What is specific only to human milk is that type 1 chain oligosaccharides are predominant in the mixture (17, 30).

While present in breast milk with a full spectrum of different molecules, HMOs were originally not a primary objective of formula manufacturing. Currently, different oligosaccharide mixtures are being added to formula to mimic the presence of HMO (31); however, most formulas on the market do not offer full mimicking of a rich HMO composition.

The ability of certain strains of *Bifidobacterium* to utilize some HMOs seems to be an evolutionary adaptation to GI tract conditions. Additionally, these strains tend to possess enzymes specific to type 1 HMOs (32). Studies on breast-fed and formula-fed infants show different compositions of the GI microbiota of infants, with a decrease of *Bifidobacterium* within formula-fed groups (33).

Bacteroides and Bifidobacterium species utilize HMOs in different pathways. A variety of bacterial species are reported to grow on a medium with HMOs. In addition to *Bifidobacterium*, *Bacteroides* species have also been reported to exhibit growth on a full HMO spectrum (34). Comparison between the species growth dynamics of *Bifidobacterium* and *Bacteroides* shows that *Bifidobacterium* spp. are able to grow faster than *Bacteroides*; thus, they become a dominant community member (18).

Analyses of growth on different HMO fractions have shown that short HMOs are utilized exclusively by some strains of *Bifidobacterium* (e.g., *B. longum* subsp. *infantis*, *B. bifidum*). On the other hand, more-complex HMOs are metabolized by a number of species, including gut-inhabiting *Bacteroides fragilis* (18).

The proposed source of this difference is the fact that *Bifidobacterium* spp., which coevolved with GI tract conditions, are able to adapt to the utilization of the HMOs that are present in the intestines of breast-fed infants. This hypothesis is supported by the fact that the major HMO molecules that are present in human milk are of short chain type 1. *Bacteroides* spp., on the other hand, utilize longer type 2 HMOs by using the mucin glycan-degrading pathway. Due to the similarity of host mucin glycans to some HMOs, *Bacteroides* are able to utilize a large fraction of breast milk; however, the ability of *Bacteroides* to grow with HMOs as a carbon source has limitations due to the fact that the HMO proportion in human milk is better suited to *Bifidobacterium* (18, 35).

Utilization of HMOs depends on a number of enzymes. The most general HMO-related enzymes are collocated in the so-called HMO cluster that is present in a number of *Bifidobacterium* species. The members of this cluster that are used for utilization of HMOs are beta-galactosidase, alpha-fucosidase, and sialidase, transporters that allow the transport of predigested oligosaccharides to bacterial cells and other proteins (36).

In addition to the enzymes that are common to a number of species, there are a variety of strain-specific enzymes that play key roles in certain metabolic pathways. *B. longum* subsp. *infantis*—the most-studied HMO-metabolizing strain—possesses additional β -galactosidase—lacto-n-biosidase (LNB)- β -galactosidase (Bga42A)—that is not collocated with other HMO cluster genes. This enzyme is used exclusively to utilize short type 1 HMOs as it has high specificity for lacto-N-tetraose (LNT) (37).

The same phenomenon is observed in *B. bifidum*—another HMO-utilizing bacterial species that possesses a metabolic pathway for type 1 HMOs and uses specific LNB (38).

As both *Bifidobacterium* and *Bacteroides* spp. are able to utilize breast milk prebiotics, it seems that different strains of each genus, each with different modes of metabolism, are present in both populations. The research by Vatanen et al. described equal

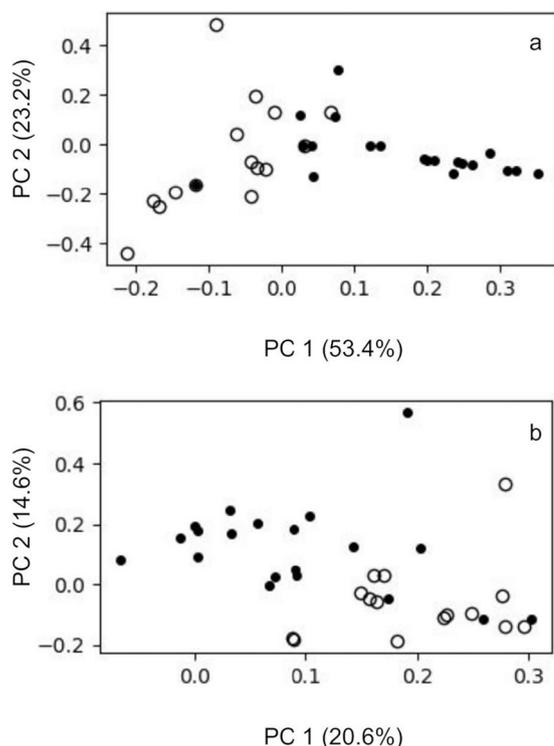


FIG 1 Principal-coordinate analysis done for the taxonomy profiling results of all the samples, with the Russian (solid dots) and Finnish (open circles) groups done with genus-level taxonomy assignment and deepest-achieved-taxonomy assignment.

abundances of HMO cluster genes in both groups. The aim of our study was to detect the genomic features associated with the strain-specific enzymes crucial for short type 1 HMO metabolic pathways and to explain the differential abundances of both HMO-utilizing genera in the Finnish and the Russian group.

RESULTS AND DISCUSSION

Taxonomy-based analyses. As reported by Vatanen et al., the Russian and Finnish groups can be characterized by their different compositions of GI microbiota. Data from principal-coordinate analysis (PCA) done on the genus level (Fig. 1a) and on the “deepest achieved” taxonomic level (Materials and Methods and Fig. 1b) show separation of the Finnish and Russian groups; nevertheless, no significant difference in GI microbiota diversity was detected. Operational taxonomic unit (OTU) richness analyzed for both the deepest achieved taxonomic level and the genus level was higher in the Finnish group ($P = 0.02$ and $P = 0.01$, respectively). However, exclusion of extremely rare ($<0.005\%$ abundance) taxa resulted in insignificant differences in OTU richness ($P = 0.68$ and $P = 0.25$ for the deepest achieved taxonomic level and genus level, respectively). While richness is highly dependent on sequencing depth and the actual composition of the community, inclusion of Shannon index allows better understanding of a sample’s diversity. The Shannon index (39) values do not differ significantly between populations, both at the genus level ($P = 0.8$) and at the deepest achieved level ($P = 0.77$).

The major differences between the groups were the levels of *Bacteroides* and *Bifidobacterium* ($P < 0.005$) (Fig. 2; see also Table S1 in the supplemental material). Finnish infants were mostly colonized by *Bacteroides* species, while the Russian infants were hosts for a variety of *Bifidobacterium* species.

However, while *Bacteroides* species are rare in the Russian population (1.3% abundance at the 75th percentile), the GI microbiome composition in the Finnish population contains a significant amount of *Bifidobacterium* spp. (8.7% abundance at the 75th

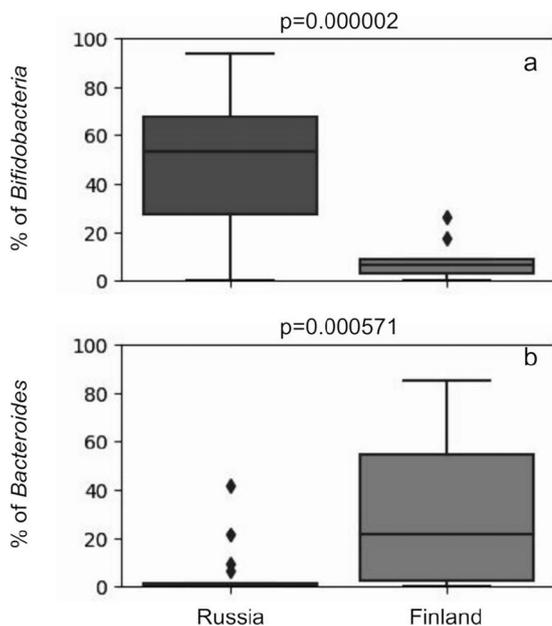


FIG 2 Relative abundances of *Bifidobacterium* (a) and *Bacteroides* (b) genera in samples of the Russian (left) and the Finnish (right) groups.

percentile). It can be clearly seen that despite their GI tracts being dominated by *Bacteroides*, a significant amount of *Bifidobacterium* spp. is present in the Finnish population. Therefore, it is not the lack of colonization of infants' GI tract by members of *Bifidobacterium* genus that results in the dominance of *Bacteroides*. Moreover, the levels of major clades (*Bifidobacterium* and *Bacteroides*) differ in the Russian and Finnish populations, respectively. While *Bifidobacterium* reaches 69.7% at the 75th percentile in the Russian population, the Finnish infants have 58% *Bacteroides* at the 75th percentile. Additionally, as reported by Vatanen et al., differences in *Bifidobacterium* abundance within a cohort should not be related to differences in breastfeeding duration, as the breastfeeding period in the Russian group was shorter. The reported abundance analyses show that *Bifidobacterium* gut-colonizing strains are stronger colonizers of infants' GI tracts than *Bacteroides* strains. Moreover, the presence of a significant amount of *Bifidobacterium* in the Finnish population may support the hypothesis that the strains that colonize these infants are different from those colonizing Russians. However, observations of such taxonomic differences require function-level evidence to fully support this hypothesis.

Genomic feature-based analysis. Genomic features were extracted from a data set comprising approximately 400 M reads in 35 samples (see Materials and Methods). The constructed feature space had roughly 1.2 M genomic features. The BiomeScout Feature Selection module selected ~50,000 genomic features, all of which were significantly ($P < 0.05$) enriched in one but not both of the groups. In contrast to the lack of differences in compositional diversity between groups (shown by comparison of Shannon indices of samples from the Russian and the Finnish groups), the majority of enriched genomic features were found in the Russian population (Fig. 3).

Detection of the majority of selected genomic features that were enriched in the Russian population suggests the presence of unique genomic components in genomes of members of that population's commensals. We also noticed that there are very few genomic features that are specific to microbes that inhabit the guts of Finnish infants. The fact that there were relatively few unique genomic features in the Finnish guts made it possible to infer that the majority of bacteria present there were highly similar to those that colonize the GI tract of Russians.

Comparison of these findings with data from taxonomy-based compositional analysis suggests that the Russian microbiome-specific genomic features may come from

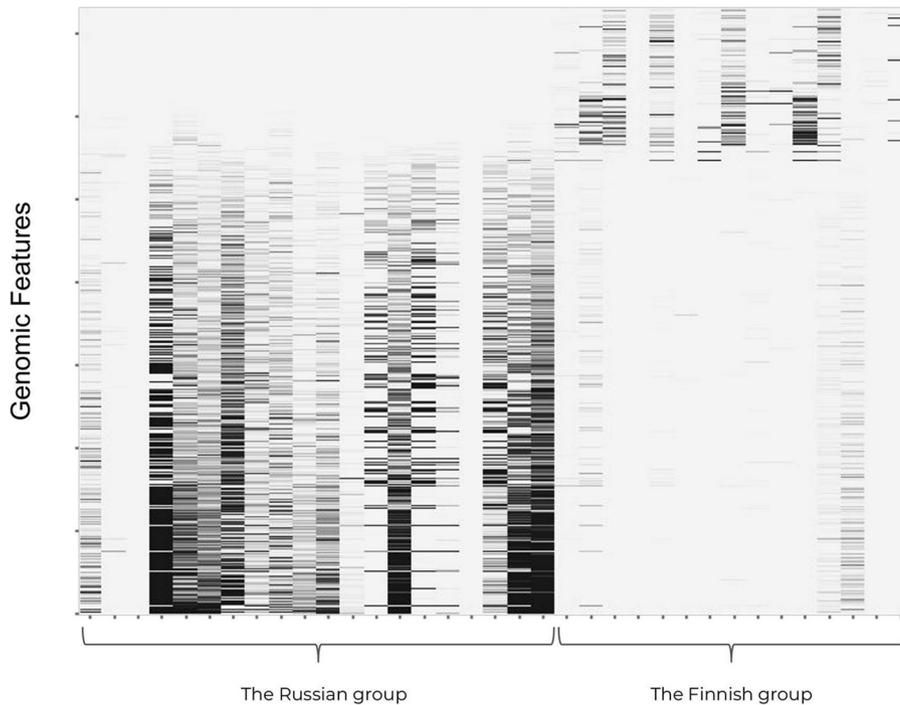


FIG 3 Heat map showing a normalized BiomeScout score level in samples for selected ($P < 0.05$) genomic features. Genomic features were sorted by an average normalized BiomeScout score in the Russian group (highest at the bottom).

specific strains of *Bifidobacterium* that overgrow other clades. This is in accordance with the previous description of the breast milk-boosted growth of *Bifidobacterium* (17).

Functional analysis. For the functional analysis, all genomic features were associated with the known bacterial genes. Annotation was performed with regard to the sequence comparison with a curated UniProtKB database (<http://uniprot.org/>).

The analysis of selected genomic features showed that a number of them were associated with HMO-processing functionalities (Table 1). Genes known to be collocated within a bifidobacterial HMO cluster (fucosidase, sialidase, galactosidase) were associated with the multiple genomic features enriched in the Russian population.

Apart from the fact that such functionalities were associated with the genomic features typical of Russian infants, the same functionalities were also associated with a

TABLE 1 List of top scored genomic features significantly differentiating in two populations (the Russian and the Finnish) that were associated with proteins playing key role in HMO metabolism (located in bifidobacterial HMO cluster)

Genomic feature ID ^a	P value	Annotated name
SRX2234520_64394	0.005	6-Phospho-beta-galactosidase
SRX2199553_754	0.005	Beta-galactosidase; evolved beta-galactosidase subunit alpha
SRX2199674_8463	0.007	Beta-galactosidase; evolved beta-galactosidase subunit alpha
SRX2199674_13733	0.007	6-Phospho-beta-galactosidase
SRX2199763_1108	0.022	Alpha-L-fucosidase 2; probable alpha-fucosidase A
SRX2199670_2849	0.024	Alpha-L-fucosidase 2; probable alpha-fucosidase A
SRX2199670_12138	0.024	Probable alpha-fucosidase A
SRX2199300_19278	0.024	Probable alpha-fucosidase A
SRX2199674_15004	0.012	Sialidase; exo-alpha-sialidase
SRX2199674_12906	0.015	Exo-alpha-sialidase; sialidase-1
SRX2199670_6933	0.016	Sialidase
SRX2199553_8800	0.025	Sialidase
SRX2199608_26754	0.033	N-Acetylneuraminate lyase
SRX2234520_48491	0.041	N-Acetylneuraminate lyase
SRX2234453_7550	0.043	N-Acetylneuraminate lyase

^aID, identifier.

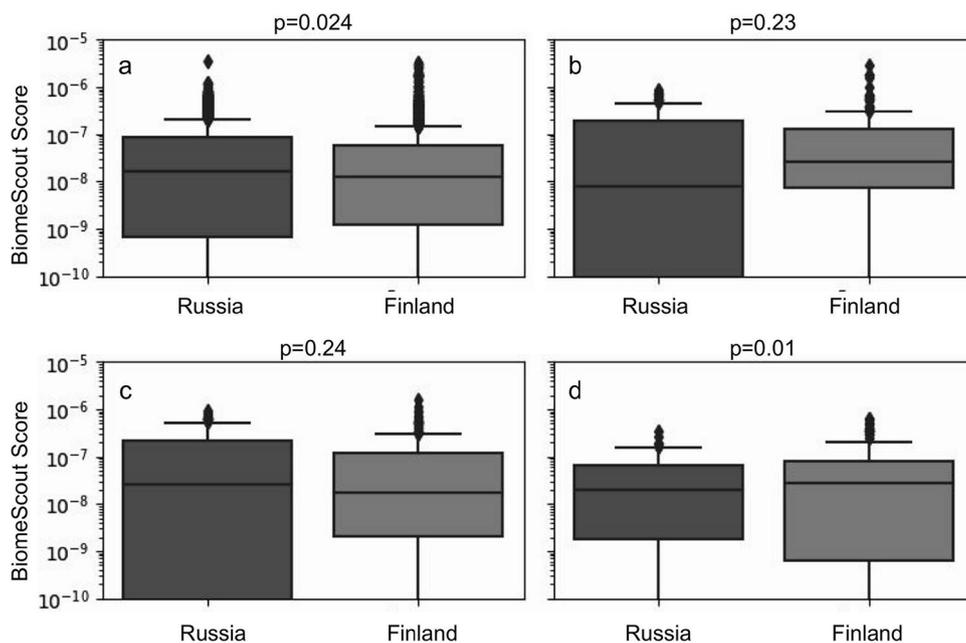


FIG 4 BiomeScout scores for enzymes present in the HMO cluster. Mean values for those enzymes were approximately equal for the Russian and the Finnish groups and were as follows: for beta-galactosidase, $8.1e-8$: $8.7e-8$ (a); for alpha-fucosidase $1.3e-7$: $1.8e-7$ (b); for sialidase $1.4e-7$: $1.1e-7$ (c); for *N*-acetylneuraminase lyase, $4.5e-8$: $7.5e-8$ (d). All genomic features detected in the whole cohort were used.

number of other genomic features. Consequently, these functional annotations revealed the presence of approximately equal levels of genomic features that are related to genes corresponding to the HMO cluster in the Russian and Finnish populations (Fig. 4).

Detection of genomic features associated with the genes present within the HMO cluster and of equal distributions of them in the two populations is consistent with the analysis by Vatanen et al. In their 2016 paper (14) on the same cohort, they reported such equality with regard to HMO cluster genes. In the same research, they found that such genes are mostly present in *Bacteroides* within the Finnish population and in *Bifidobacterium* within the Russian population. Such findings are consistent with our findings based on the BiomeScout technology, but taxonomy-agnostic genomic features do not imply the source organism at this step of analysis.

Detection of enzyme enrichment. To verify the hypothesis that the bifidobacterial strains present in the Russian population possess functionalities typical of enzymes reported as strain specific, we analyzed different HMO-processing enzymes. We chose two enzymes that are key for the utilization of the type 1 HMOs that are most common in human milk in the two key bifidobacterial species: lacto-*n*-biosidase (LNB) of *B. bifidum* and β -galactosidase from *B. longum* subsp. *infantis* (Bga42A).

As was already noticed in the case of HMO cluster genes, analyses of significantly enriched genomic features may be misleading; therefore, all genomic features were used in the analysis of strain-specific functionalities. Among these features, we chose those associated with specific HMO-processing enzymes: LNB and Bga42A (Materials and Methods).

BiomeScout analysis showed a significantly higher level of genomic features associated with gene-encoding enzymes specific to short type 1 HMOs in the Russian population. Genomic features associated with lacto-*n*-biosidase, present predominantly in *B. bifidum* and some bifidobacterial strains, were found to be significantly enriched ($P = 3.6e-11$) in the Russian population (Fig. 5a).

Additionally, genomic features associated with Bga42A that are present specifically in *B. longum* subsp. *infantis* were found to be present at a significantly higher level ($P = 5.2e-19$) in that population (Fig. 5b).

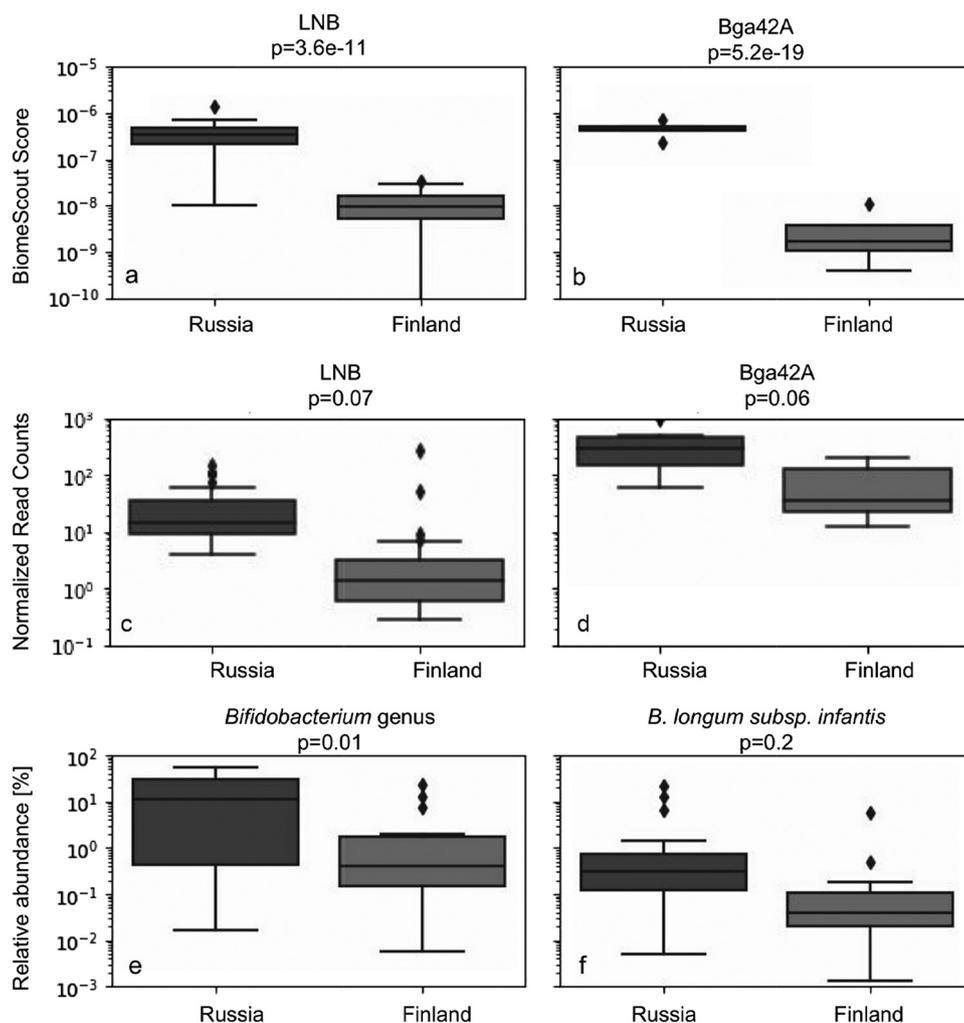


FIG 5 BiomeScout scores of all genomic features annotated as matching lacto-N-biosidase (a) and beta-galactosidase Bga42A (b) significantly differentiate the Russian and the Finnish groups. In contrast, the counted reads mapped against those genomic features (c and d, respectively) were shifted in the same way, but the difference was insignificant. Differential presence of Bga42A in samples cannot be inferred from the abundance of *B. longum* subsp. *infantis*. The strain that contains this enzyme was not detected with significantly higher abundance in the Russian population (f), while bacteria of the *Bifidobacterium* genus were significantly more abundant (e).

Differences in the genomic features associated with short type 1 HMOs reveal the basis of the differences between the *Bifidobacterium* abundances in the guts of Russian and Finnish infants. *Bifidobacterium* strains that lack the ability to utilize short type 1 HMOs have to compete with other community members that are able to utilize a broad HMO fraction, mostly *Bacteroides*. Additionally, over 26% of the whole human HMO composition consists of LNT (40); thus, the lack of ability to metabolize this compound is a significant handicap. As described by Marcobal et al. (18), under conditions of use of porcine mucin glycans as a sole source of carbon, *Bacteroides* overgrow *Bifidobacterium*. The proposed explanation for such a difference was the fact that the compounds in a large fraction of HMO are similar to mucin glycans. *Bacteroides* spp. use such oligosaccharides as their main carbon source, while *Bifidobacterium* spp. are adapted to the use of the small oligosaccharides of human milk. Therefore, our findings are in accordance with the growth dynamics on short type 1 HMOs, complex glycans, and full HMOs that were described by Marcobal et al.

Such a finding shows the possibility of detecting enzyme-specific function-level differences between two populations. Such an analysis, performed with a straightfor-

ward pipeline that counts reads mapped to each reference gene, shows no significant differences between the two groups in the levels of these genes (Fig. 5c and d).

Enrichment of LNB could be inferred from analysis of the taxonomic content of samples, as the Russian group was significantly enriched with *B. bifidum* (averages of 15.8% and 0.8% in the Russian and Finnish groups, respectively, $P = 0.003$), which possesses such an enzyme. Importantly, in the case of Bga42A from *B. longum* subsp. *infantis*, genomic feature-based analysis showed differences that were not revealed by the traditional taxonomy-based approach. Distinguishing *B. longum* subspecies—even with shotgun sequencing data—is an extremely complex task because these bacteria may share over 99% of genomic sequences. The fractions of microbiome members assigned as *B. longum* species and of those specifically assigned as *B. longum* subsp. *longum* were significantly enriched in the Russian population ($P = 0.01$ and $P = 0.03$, respectively). However, no significant difference in *B. longum* subsp. *infantis* abundance was detected between the Russian and the Finnish populations ($P = 0.2$) (Fig. 5e and f). Importantly, strain level was annotated only rarely.

In this paper, we report the ability to determine enzyme-scale differences between two populations using next-generation sequencing data. What is more, our approach makes it possible to detect differences that are hard to detect with standard pipelines for both taxonomy analysis and gene catalog annotation.

Importantly, BiomeScout analysis is done with no prior taxonomy assignment analysis—Genomic feature data are extracted from the raw sequencing data. Therefore, we were able to also utilize reads of ambiguous origin without bias toward information stored in the available databases.

The described use case allows detection of the differences between the bifidobacterial strains that colonize the Russian and the Finnish infants' GI tract in terms of the presence of enzymes specific to short type 1 HMOs. As such oligosaccharides represent a large portion of the whole HMO mixture, we hypothesize that the lack of dedicated enzymes in the strains colonizing the Finnish infants allows *Bacteroides* to overgrow *Bifidobacterium*. Further studies are necessary to more deeply investigate the phenotypic differences between the dominating and dominated *Bifidobacterium* strains.

MATERIALS AND METHODS

The data set analyzed in this study represents a subset of the DIABIMMUNE cohort (<http://www.diabimmune.org>) (14). The chosen samples were sequenced during the first 180 days of the infants' lives. Infants from the DIABIMMUNE cohort were subsampled such that the groups were balanced. Sample collection in the Russian population started later (around month 3 of life); therefore, the Finnish samples from the earlier period of life were excluded from analyses. For each cohort member, only the sample from the first time point was included in the analysis. Only the Finnish and the Russian groups were chosen, as the Estonian population has previously been described as being in a "transition zone" between west and east due to political changes over the last 30 years. For a better understanding of early colonizing strains, only samples collected within the first 180 days of life were used for this study. During this period breast milk is the primary or only nutrition for infants. Additionally, such a period is crucial for the development of the immune system of infants. Consequently, 35 shotgun metagenomic sequencing samples were chosen for further investigation.

Taxonomy binning and profiling were done with Kraken2 (<https://ccb.jhu.edu/software/kraken2/>) (41). For binning of reads and taxonomy profiling of samples, a standard database was built according to the tool's manual. Lineages were built for all assigned taxonomies with seven canonical levels (kingdom, phylum, class, order, family, genus, and species) and subspecies if assigned. Such annotation is referred to as the deepest achieved taxonomic level. Given the fact that multiple lineages were shorter than the subspecies level, lineages truncated above the genus level were also used. The abundances corresponding to the resulting identical assignments were summed.

Initial assembly of metagenomic samples was done with Megahit v1.1.3 (<https://github.com/voutcn/megahit/>) (42). Genomic feature-based analysis was done with BiomeScout tool v.1.0.2. In BiomeScout's nomenclature, "genomic feature" is a general term that refers to any part of a genomic sequence. To construct the genomic feature space, BiomeScout used preassembled metagenomes from the whole studied cohort as an input-only data set.

Each feature was scored with respect to the two methodologies: counting mapped reads and BiomeScout. In the first, all reads mapped on each genomic feature were counted. BiomeScout analysis used its own internal proprietary scoring algorithm. The two scoring results were normalized with respect to the number of sequenced reads.

The scored genomic features were then passed to the feature selection algorithms, which choose a set of genomic features that significantly differentiate groups. Each selected genomic feature had a P

value assigned by BiomeScout's Feature Selection module using Student's *t* test. The same test was used in the whole analysis where *P* values are reported.

Functional annotation was done with bioinformatics analysis of each genomic feature. The similarity of the sequence to full or partial sequences of known proteins was calculated. Reviewed entries from the UniProtKB v. 2018_09 database (<http://uniprot.org>) (43) were used. Each genomic feature was matched with a set of known genes using DIAMOND v0.9.22 (<https://github.com/bbuchfink/diamond>) (44). Identity threshold analysis was based on a paper by Addou et al., which describes 80% identity as a threshold for 100% consensus of a four-digit EC number (45, 46). The same percentage was chosen for query coverage threshold based on the UniRef database clustering approach (47, 48).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSphereDirect.00152-19>.

TABLE S1, XLSX file, 1.4 MB.

ACKNOWLEDGMENTS

We thank Monika Brzychczy-Włoch for valuable discussions and expertise.

This work was funded by NCRD Fast Track Grant POIR.01.01.01-00-0347/17 awarded to Ardigen S.A.

REFERENCES

- Shreiner AB, Kao JY, Young VB. 2015. The gut microbiome in health and in disease. *Curr Opin Gastroenterol* 31:69. <https://doi.org/10.1097/MOG.0000000000000139>.
- D'Argenio V, Salvatore F. 2015. The role of the gut microbiome in the healthy adult status. *Clin Chim Acta* 451:97–102. <https://doi.org/10.1016/j.cca.2015.01.003>.
- Ghaisas S, Maher J, Kanthasamy A. 2016. Gut microbiome in health and disease: linking the microbiome-gut-brain axis and environmental factors in the pathogenesis of systemic and neurodegenerative diseases. *Pharmacol Ther* 158:52–62. <https://doi.org/10.1016/j.pharmthera.2015.11.012>.
- Kostic AD, Xavier RJ, Gevers D. 2014. The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* 146:1489–1499. <https://doi.org/10.1053/j.gastro.2014.02.009>.
- Ni J, Wu GD, Albenberg L, Tomov VT. 2017. Gut microbiota and IBD: causation or correlation? *Nat Rev Gastroenterol Hepatol* 14:573. <https://doi.org/10.1038/nrgastro.2017.88>.
- Jangi S, Gandhi R, Cox LM, Li N, von Glehn F, Yan R, Patel B, Mazzola MA, Liu S, Glanz BL, Cook S, Tankou S, Stuart F, Melo K, Nejad P, Smith K, Topçuoğlu BD, Holden J, Kivisäkk P, Chitnis T, De Jager PL, Quintana FJ, Gerber GK, Bry L, Weiner HL. 2016. Alterations of the human gut microbiome in multiple sclerosis. *Nat Commun* 7:12015. <https://doi.org/10.1038/ncomms12015>.
- Bhargava P, Mowry EM. 2014. Gut microbiome and multiple sclerosis. *Curr Neurol Neurosci Rep* 14:492. <https://doi.org/10.1007/s11910-014-0492-2>.
- Zhu C, Miller M, Marpaka S, Vaysberg P, Rühlemann MC, Wu G, Heinsen F-A, Tempel M, Zhao L, Lieb W, Franke A, Bromberg Y. 2018. Functional sequencing read annotation for high precision microbiome analysis. *Nucleic Acids Res* 46:e23. <https://doi.org/10.1093/nar/gkx1209>.
- Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpnits TV, Prieto PA, Vicente D, Hoffman K, Wei SC, Cogdill AP, Zhao L, Hudgens CW, Hutchinson DS, Manzo T, de Macedo MP, Cotechini T, Kumar T, Chen WS, Reddy SM, Szczepaniak Sloane R, Galloway-Pena J, Jiang H, Chen PL, Shpall EJ, Rezvani K, Alousi AM, Chemaly RF, Shelburne S, Vence LM, Okhuysen PC, Jensen VB, Swennes AG, McAllister F, Marcelo Riquelme Sanchez E, Zhang Y, Le Chatelier E, Zitvogel L, Pons N, Austin-Breneman JL, Haydu LE, Burton EM, Gardner JM, Sirmans E, Hu J, Lazar AJ, Tsujikawa T, Diab A, et al. 2018. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* 359:97. <https://doi.org/10.1126/science.aan4236>.
- Matson V, Fessler J, Bao R, Chongsuwat T, Zha Y, Alegre M-L, Luke JJ, Gajewski TF. 2018. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science* 359:104–108. <https://doi.org/10.1126/science.aao3290>.
- Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, Fluckiger A, Messaoudene M, Rauber C, Roberti MP, Fidelle M, Flament C, Poirier-Colame V, Opolon P, Klein C, Iribarren K, Mondragón L, Jacquelot N, Qu B, Ferrere G, Clémenson C, Mezquita L, Masip JR, Naltet C, Brosseau S, Kaderbhai C, Richard C, Rizvi H, Levenez F, Galleron N, Quinquis B, Pons N, Ryffel B, Minard-Colin V, Gonin P, Soria J-C, Deutsch E, Loriot Y, Ghiringhelli F, Zalcman G, Goldwasser F, Escudier B, Hellmann MD, Eggermont A, Raouf D, Albiges L, Kroemer G, Zitvogel L. 2018. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* 359:91–97. <https://doi.org/10.1126/science.aan3706>.
- Fujimura KE, Lynch SV. 2015. Microbiota in allergy and asthma and the emerging relationship with the gut microbiome. *Cell Host Microbe* 17:592–602. <https://doi.org/10.1016/j.chom.2015.04.007>.
- Vatanen T, Franzosa EA, Schwager R, Tripathi S, Arthur TD, Vehik K, Lernmark Å, Hagopian WA, Rewers MJ, She J-X, Toppari J, Ziegler A-G, Akolkar B, Krischer JP, Stewart CJ, Ajami NJ, Petrosino JF, Gevers D, Lähdesmäki H, Vlamakis H, Huttenhower C, Xavier RJ. 2018. The human gut microbiome in early-onset type 1 diabetes from the TEDDY study. *Nature* 562:589–594. <https://doi.org/10.1038/s41586-018-0620-2>.
- Vatanen T, Kostic AD, d'Hennezel E, Siljander H, Franzosa EA, Yassour M, Kolde R, Vlamakis H, Arthur TD, Hämäläinen A-M, Peet A, Tillmann V, Uibo R, Mokurov S, Dorshakova N, Ilonen J, Virtanen SM, Szabo SJ, Porter JA, Lähdesmäki H, Huttenhower C, Gevers D, Cullen TW, Knip M, DIABIMMUNE Study Group, Xavier RJ. 2016. Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell* 165:842–853. <https://doi.org/10.1016/j.cell.2016.04.007>.
- Lerner A, Jeremias P, Matthias T. 2016. The world incidence and prevalence of autoimmune diseases is increasing. *Int J Celiac Dis* 3:151–155. <https://doi.org/10.12691/ijcd-3-4-8>.
- Okada H, Kuhn C, Feillet H, Bach J-F. 2010. The “hygiene hypothesis” for autoimmune and allergic diseases: an update. *Clin Exp Immunol* 160:1–9. <https://doi.org/10.1111/j.1365-2249.2010.04139.x>.
- Bode L. 2012. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* 22:1147–1162. <https://doi.org/10.1093/glycob/cws074>.
- Marcobal A, Barboza M, Sonnenburg ED, Pudlo N, Martens EC, Desai P, Lebrilla CB, Weimer BC, Mills DA, Bruce German J, Sonnenburg JL. 2011. Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* 10:507–514. <https://doi.org/10.1016/j.chom.2011.10.007>.
- Milani C, Duranti S, Bottacini F, Casey E, Turroni F, Mahony J, Belzer C, Delgado Palacio S, Arbolea Montes S, Mancabelli L, Lugli GA, Rodriguez JM, Bode L, de Vos W, Gueimonde M, Margolles A, van Sinderen D, Ventura M. 2017. The first microbial colonizers of the human gut: composition, activities, and health implications of the infant gut microbiota. *Microbiol Mol Biol Rev* 81:e00036-17. <https://doi.org/10.1128/MMBR.00036-17>.
- Bergström A, Skov TH, Bahl MI, Roager HM, Christensen LB, Ejlerskov KT, Mølgaard C, Michaelsen KF, Licht TR. 2014. Establishment of intestinal microbiota during early life: a longitudinal, explorative study of a large

- cohort of Danish infants. *Appl Environ Microbiol* 80:2889. <https://doi.org/10.1128/AEM.00342-14>.
21. Neu J, Rushing J. 2011. Cesarean versus vaginal delivery: long term infant outcomes and the hygiene hypothesis. *Clin Perinatol* 38:321–331. <https://doi.org/10.1016/j.clp.2011.03.008>.
 22. Madan JC, Hoen AG, Lundgren SN, Farzan SF, Cottingham KL, Morrison HG, Sogin ML, Li H, Moore JH, Karagas MR. 2016. Effects of Cesarean delivery and formula supplementation on the intestinal microbiome of six-week old infants. *JAMA Pediatr* 170:212. <https://doi.org/10.1001/jamapediatrics.2015.3732>.
 23. Pärnänen K, Karkman A, Hultman J, Lyra C, Bengtsson-Palme J, Joakim Larsson DG, Rautava S, Isolauri E, Salminen S, Kumar H, Satokari R, Virta M. 2018. Maternal gut and breast milk microbiota affect infant gut antibiotic resistance and mobile genetic elements. *Nat Commun* 9:3891. <https://doi.org/10.1038/s41467-018-06393-w>.
 24. Mueller NT, Bakacs E, Combellick J, Grigoryan Z, Dominguez-Bello MG. 2015. The infant microbiome development: mom matters. *Trends Mol Med* 21:109. <https://doi.org/10.1016/j.molmed.2014.12.002>.
 25. Henrick BM, Hutton AA, Palumbo MC, Casaburi G, Mitchell RD, Underwood MA, Smilowitz JT, Frese SA. 2018. Elevated fecal pH indicates a profound change in the breastfed infant gut microbiome due to reduction of *Bifidobacterium* over the past century. *mSphere* 3:e00041-18. <https://doi.org/10.1128/mSphere.00041-18>.
 26. Underwood MA, Bruce German J, Lebrilla CB, Mills DA. 2015. *Bifidobacterium longum* subspecies *infantis*: champion colonizer of the infant gut. *Pediatr Res* 77:229. <https://doi.org/10.1038/pr.2014.156>.
 27. Harmsen HJ, Wildeboer-Veloo AC, Raangs GC, Wagendorp AA, Klijn N, Bindels JG, Welling GW. 2000. Analysis of intestinal flora development in breast-fed and formula-fed infants by using molecular identification and detection methods. *J Pediatr Gastroenterol Nutr* 30:61–67. <https://doi.org/10.1097/00005176-200001000-00019>.
 28. Casaburi G, Frese SA. 2018. Colonization of breastfed infants by *Bifidobacterium longum* subspecies *infantis* EVC001 reduces virulence gene abundance. *Human Microbiome J* 9:7–10. <https://doi.org/10.1016/j.humic.2018.05.001>.
 29. Ballard O, Morrow AL. 2013. Human milk composition: nutrients and bioactive factors. *Pediatr Clin North Am* 60:49. <https://doi.org/10.1016/j.pcl.2012.10.002>.
 30. Urashima T, Asakuma S, Leo F, Fukuda K, Messer M, Oftedal OT. 2012. The predominance of type I oligosaccharides is a feature specific to human breast milk. *Adv Nutr* 3:473S–482S. <https://doi.org/10.3945/an.111.001412>.
 31. Bode L, Contractor N, Barile D, Pohl N, Prudden AR, Boons G-J, Jin Y-S, Jennewein S. 2016. Overcoming the limited availability of human milk oligosaccharides: challenges and opportunities for research and application. *Nutr Rev* 74:635–644. <https://doi.org/10.1093/nutrit/nuw025>.
 32. LoCascio RG, Ninonuevo MR, Freeman SL, Sela DA, Grimm R, Lebrilla CB, Mills DA, German JB. 2007. Glycoprofiling of bifidobacterial consumption of human milk oligosaccharides demonstrates strain specific, preferential consumption of small chain glycans secreted in early human lactation. *J Agric Food Chem* 55:8914–8919. <https://doi.org/10.1021/jf0710480>.
 33. Lee J-H, O'Sullivan DJ. 2010. Genomic insights into *Bifidobacteria*. *Microbiol Mol Biol Rev* 74:378–416. <https://doi.org/10.1128/MMBR.00004-10>.
 34. Marcobal A, Sonnenburg JL. 2012. Human milk oligosaccharide consumption by intestinal microbiota. *Clin Microbiol Infect* 18(Suppl 4):12–15. <https://doi.org/10.1111/j.1469-0691.2012.03863.x>.
 35. Katayama T. 2016. Host-derived glycans serve as selected nutrients for the gut microbe: human milk oligosaccharides and bifidobacteria. *Biosci Biotechnol Biochem* 80:621–632. <https://doi.org/10.1080/09168451.2015.1132153>.
 36. Odamaki T, Horigome A, Sugahara H, Hashikura N, Minami J, Xiao J-Z, Abe F. 2015. Comparative genomics revealed genetic diversity and species/strain-level differences in carbohydrate metabolism of three probiotic bifidobacterial species. *Int J Genomics* 2015:567809. <https://doi.org/10.1155/2015/567809>.
 37. Yoshida E, Sakurama H, Kiyohara M, Nakajima M, Kitaoka M, Ashida H, Hirose J, Katayama T, Yamamoto K, Kumagai H. 2012. *Bifidobacterium longum* subsp. *infantis* uses two different β -galactosidases for selectively degrading type-1 and type-2 human milk oligosaccharides. *Glycobiology* 22:361–368. <https://doi.org/10.1093/glycob/cwr116>.
 38. Wada J, Ando T, Kiyohara M, Ashida H, Kitaoka M, Yamaguchi M, Kumagai H, Katayama T, Yamamoto K. 2008. *Bifidobacterium bifidum* lacto-N-biosidase, a critical enzyme for the degradation of human milk oligosaccharides with a type 1 structure. *Appl Environ Microbiol* 74:3996–4004. <https://doi.org/10.1128/AEM.00149-08>.
 39. Hill TCJ, Walsh KA, Harris JA, Moffett BF. 2003. Using ecological diversity measures with bacterial communities. *FEMS Microbiol Ecol* 43:1–11. <https://doi.org/10.1111/j.1574-6941.2003.tb01040.x>.
 40. Thurl S, Muller-Werner B, Sawatzki G. 1996. Quantification of individual oligosaccharide compounds from human milk using high-pH anion-exchange chromatography. *Anal Biochem* 235:202–206. <https://doi.org/10.1006/abio.1996.0113>.
 41. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15:R46. <https://doi.org/10.1186/gb-2014-15-3-r46>.
 42. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, Yamashita H, Lam T-W. 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 102:3–11. <https://doi.org/10.1016/j.jmeth.2016.02.020>.
 43. The UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 46:2699. <https://doi.org/10.1093/nar/gky092>.
 44. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60. <https://doi.org/10.1038/nmeth.3176>.
 45. Pearson WR. 2013. An introduction to sequence similarity (“homology”) searching. *Curr Protoc Bioinformatics* Chapter 3:Unit3.1. <https://doi.org/10.1002/0471250953.bi0301s42>.
 46. Addou S, Rentszsch R, Lee D, Orengo CA. 2009. Domain-based and family-specific sequence identity thresholds increase the levels of reliable protein function transfer. *J Mol Biol* 387:416–430. <https://doi.org/10.1016/j.jmb.2008.12.045>.
 47. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH. 2015. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 31:926–932. <https://doi.org/10.1093/bioinformatics/btu739>.
 48. Mirdita M, von den Driesch L, Galiez C, Martin MJ, Söding J, Steinegger M. 2017. Unifluc databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res* 45:D170.