

ARTICLE

Received 1 Dec 2015 | Accepted 13 May 2016 | Published 24 Jun 2016

DOI: 10.1038/ncomms11939

OPEN

# A key genetic factor for fucosyllactose utilization affects infant gut microbiota development

Takahiro Matsuki<sup>1,\*</sup>, Kana Yahagi<sup>1,\*</sup>, Hiroshi Mori<sup>2,\*</sup>, Hoshitaka Matsumoto<sup>1,\*</sup>, Taeko Hara<sup>1</sup>, Saya Tajima<sup>1</sup>, Eishin Ogawa<sup>3</sup>, Hiroko Kodama<sup>3</sup>, Kazuya Yamamoto<sup>2</sup>, Takuji Yamada<sup>2</sup>, Satoshi Matsumoto<sup>1,\*\*</sup> & Ken Kurokawa<sup>2,4,5,\*\*</sup>

Recent studies have demonstrated that gut microbiota development influences infants' health and subsequent host physiology. However, the factors shaping the development of the microbiota remain poorly understood, and the mechanisms through which these factors affect gut metabolite profiles have not been extensively investigated. Here we analyse gut microbiota development of 27 infants during the first month of life. We find three distinct clusters that transition towards Bifidobacteriaceae-dominant microbiota. We observe considerable differences in human milk oligosaccharide utilization among infant bifidobacteria. Colonization of fucosyllactose (FL)-utilizing bifidobacteria is associated with altered metabolite profiles and microbiota compositions, which have been previously shown to affect infant health. Genome analysis of infants' bifidobacteria reveals an ABC transporter as a key genetic factor for FL utilization. Thus, the ability of bifidobacteria to utilize FL and the presence of FL in breast milk may affect the development of the gut microbiota in infants, and might ultimately have therapeutic implications.

<sup>1</sup>Yakult Central Institute, 5-11 Izumi, Kunitachi-shi, Tokyo 186-8650, Japan. <sup>2</sup>Department of Biological Information, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan. <sup>3</sup>Teikyo University School of Medicine, 2-11-1 Kaga, Itabashi-Ku, Tokyo 117-8605, Japan. <sup>4</sup>Earth-Life Science Institute, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan. <sup>5</sup>National Institute of Genetics, Center for Information Biology, Yata 1111, Mishima, Shizuoka 411-8540, Japan. \* These authors contributed equally to this work. \*\* These authors jointly supervised this work. Correspondence and requests for materials should be addressed to T.M. (email: takahiro-matsuki@yakult.co.jp).

It is becoming increasingly apparent that the bacterial ecosystem in our gut has a profound influence on human health and disease. The gut microbiota contributes to immune system maturation, energy harvesting and sympathetic nervous system development. In particular, the composition and metabolite profiles of gut microbiota have been associated with pathogen resistance<sup>1–3</sup>, inflammatory responses<sup>4</sup> and adiposity<sup>5,6</sup>.

Initial gut microbe colonization begins immediately after birth, and bacterial ecosystems develop within the first few days. Previous studies have reported that the composition of the infant gut microbiota differs from that of adults<sup>7–9</sup>, that substantial variation occurs between individuals<sup>6,10,11</sup> and that bifidobacteria predominate in most infants<sup>11–13</sup>. Recent studies also demonstrated that environmental factors including the mode of delivery and feeding affect the gut microbiota assemblage and that the process is not random<sup>6,13,14</sup>. Furthermore, it has been indicated that the gut microbiota development during infancy can have long-lasting effects on the individual's future health<sup>15–18</sup>. However, little is known about their pattern of progression, factors that drive the assembly of infant gut microbiota and how these factors affect metabolite profiles.

Here we investigated gut microbiota compositions and metabolic profiles for 217 stool samples obtained from 27 infants during their first month of life (202 samples from 12 infants were analysed longitudinally and 15 samples from 15 infants were studied in follow-up). The dynamics and equilibria of the developing microbiota were investigated, and their associations with metabolites were evaluated. We subsequently analysed phenotypes and genotypes of isolated bifidobacteria, and found a key genetic factor affecting infant gut microbiota composition and metabolite profile.

## Results

**Early development of gut microbiota.** To investigate the dynamics of gut microbiota immediately after birth, we analysed the sequences of the V1–V2 region of the 16S rRNA genes obtained from 12 infants born by normal delivery (Supplementary Table 1) using the 454 GS Junior platform. We obtained stool samples every day during the first week after birth and every other day thereafter until 1 month of age (~17 stool samples per infant; 202 samples in total). A total of 588,293 pyrosequencing reads (average  $2,912 \pm 1,397$  reads per sample; Supplementary Table 2 and Supplementary Fig. 1) were analysed using an open-source Quantitative Insights Into Microbial Ecology (QIIME) software pipeline<sup>19</sup> (Supplementary Table 3).

Figure 1a shows an age-dependent, gut microbiota composition heatmap for each subject at the bacterial family level. The analysis demonstrated that there are major variations in both the composition and dynamic progression among individuals. Overall, the composition of the infant's microbiota was relatively simple, being composed of only a few dominant bacterial families. The displacement of predominant bacteria occurred within only a few days. We observed an increased average abundance of Bifidobacteriaceae,  $\alpha$ -diversities and total bacterial cell counts, as well as decreased average abundances of Enterobacteriaceae and Staphylococcaceae (Supplementary Fig. 2).

Characteristics of the taxonomic composition observed among the samples were clearly distinguished by principal coordinate analysis (PCoA) and partitioning around medoids (PAM)<sup>20</sup> on the basis of bacterial family composition data (Fig. 1b and Supplementary Data 1). Values of the Calinski-Harabasz (CH) index with PAM clustering suggest that the infant microbiota could be divided into three clusters (Supplementary Fig. 3), which were characterized by the predominance of Bifidobacteriaceae, Enterobacteriaceae or Staphylococcaceae (Fig. 1c).

We subsequently observed sequential transitions occurring from Staphylococcaceae- to Enterobacteriaceae- and/or Enterobacteriaceae- to Bifidobacteriaceae-dominated microbiota, with considerable individual variation in the day of the transition (Fig. 1d). Transitions in the opposite direction were rarely observed. The results suggest that the best-adapted bacterial family in the infant gut may be Bifidobacteriaceae, followed by Enterobacteriaceae and Staphylococcaceae.

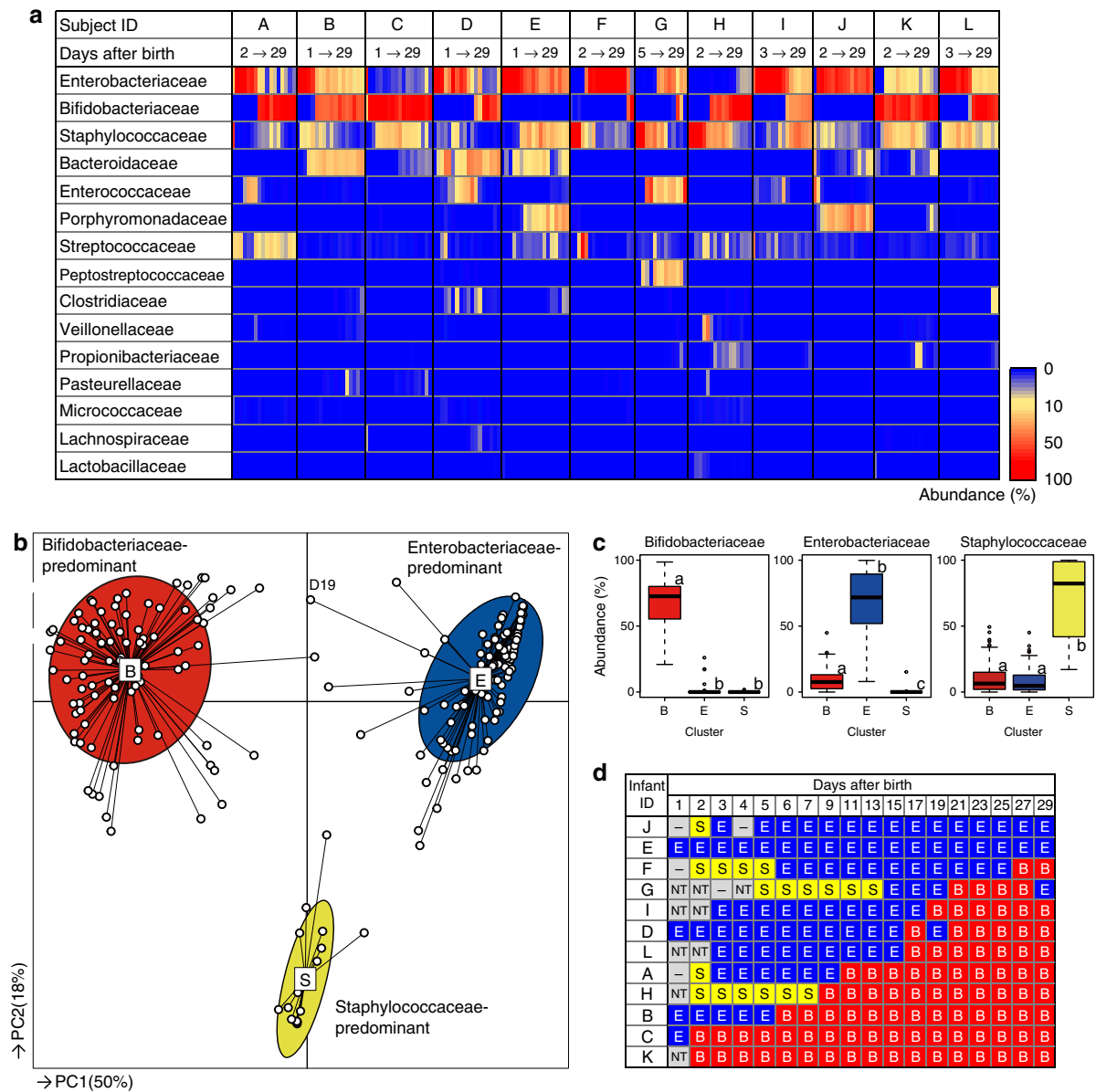
**Microbiota in 1-month-old infants.** Subsequently, we performed 16S rRNA gene-library analysis using additional faecal samples; in addition to the faecal samples from 12 infants (subjects A–L, shown in Fig. 1 on day 29), we obtained faecal samples from their parents ( $n = 22$ ) and another 15 breast-fed infants at approximately 1 month after birth (Supplementary Tables 4 and 5). A total of 147,010 high-quality reads (average  $3,058 \pm 1,232$  reads per sample; Supplementary Data 2) were analysed, and the resulting family composition heatmap is shown in Fig. 2a.

The results of PCoA and PAM clustering based on bacterial family compositions suggest that the microbiota of 1-month-old infants stratified into two clusters that were distinct from the adult cluster (Fig. 2b and Supplementary Fig. 4). The majority of infants ( $n = 18$ , designated as cluster B) were characterized as having a significantly high abundance of Bifidobacteriaceae, and the minor cluster ( $n = 9$ , designated as cluster E) had a significantly high abundance of facultative anaerobes such as Enterobacteriaceae, Enterococcaceae and Staphylococcaceae (Fig. 2c and Supplementary Figs 5 and 6). The 22 adult samples formed a single cluster (designated cluster AD), showing significantly higher abundances of Lachnospiraceae, 'Clostridiales incertae sedis XIV', Bacteroidaceae, Ruminococcaceae and Peptostreptococcaceae, as well as higher  $\alpha$ -diversity compared with the two infant microbiota clusters (Fig. 2c and Supplementary Figs 5–7).

The correlation coefficients observed between bacterial family abundances in infant and adult microbiota are illustrated in network diagrams (Fig. 2d and Supplementary Figs 8 and 9). Analysis of the infant stool samples indicated that the abundance of predominant Bifidobacteriaceae negatively correlated with those of Enterobacteriaceae, Enterococcaceae, Clostridiaceae and Staphylococcaceae. Furthermore, the network diagram for adults indicated that bacterial family compositions and their associations were different from those of infants.

**Bacterial lineages and gut environments.** To understand how gut microbiota affect the host's physiology, it is important to understand the entire gut ecosystem, not only in terms of microbiota compositions but also in terms of the metabolites produced by the bacteria. Therefore, we investigated the pH and organic acid concentrations of each infant's stool (Supplementary Data 3) and assessed the correlation between these parameters and bacterial family abundances (Fig. 3a and Supplementary Fig. 10). We found that increased Bifidobacteriaceae abundance positively correlated with organic acid concentrations and total bacterial counts but negatively correlated with pH. In contrast, the abundance of facultative anaerobes such as Enterobacteriaceae and Staphylococcaceae correlated with decreased organic acid concentrations, decreased total bacterial counts and increased pH (Fig. 3a), as well as decreased Bifidobacteriaceae abundance (Fig. 2d).

Previous studies have reported that bifidobacteria produce acetate and lactate as their metabolites<sup>1,21</sup>, and that some of these strains (for example, *Bifidobacterium longum* ss. *infantis* ATCC 15697) can efficiently utilize some human milk oligosaccharide (HMO) components<sup>22–25</sup>. Therefore, we hypothesized that

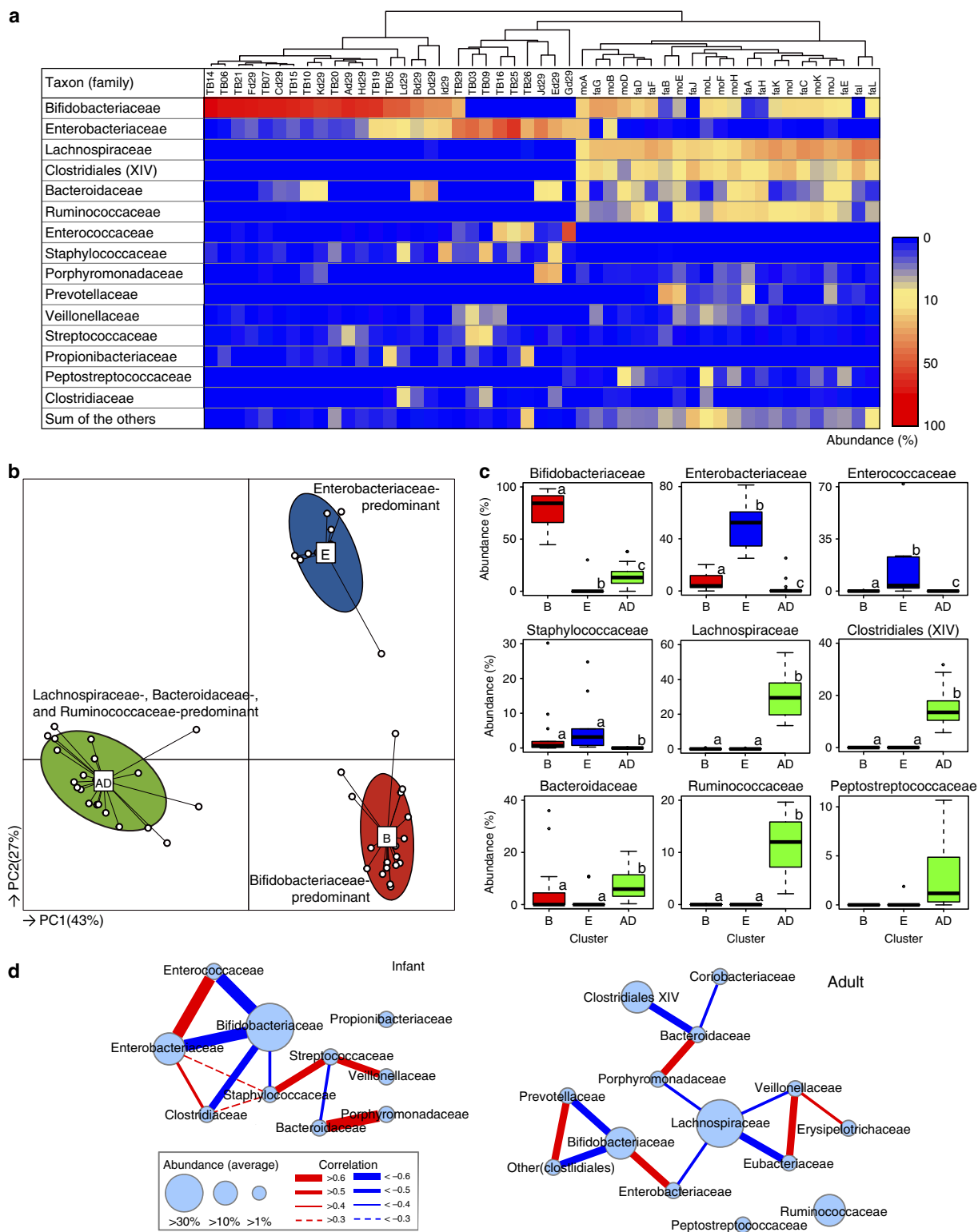


**Figure 1 | Infant gut microbiota community profiles during the first month of life.** (a) Microbiota profiles in stool samples from 12 subjects ( $n = 202$ ;  $\sim 17$  sampling days per subject), temporally ordered from left to right. Each row represents taxonomic groups at the family level. The top 15 families are displayed and sorted according to relative abundance. Abundances are represented using the colour scale. (b) Characteristics of infant gut microbiota, illustrated by PCoA and PAM clustering analyses. Data from individuals (points) were clustered, and the centres of gravity (rectangles) were computed for each class. The coloured ellipses encompass 67% of the samples in each cluster. (c) Box plots showing the relative abundances of the main contributors to each cluster. Different letters (a-c) above the boxes indicate significant differences between clusters ( $P < 0.05$ , Mann-Whitney  $U$ -test with Bonferroni's correction). (d) Temporal shift from Staphylococcaceae- or Enterobacteriaceae-dominant microbiota to Bifidobacteriaceae-dominant microbiota. S, Staphylococcaceae-dominated (yellow); E, Enterobacteriaceae-dominated (blue); B, Bifidobacteriaceae-dominated (red); NT, not tested; —, sample not provided.

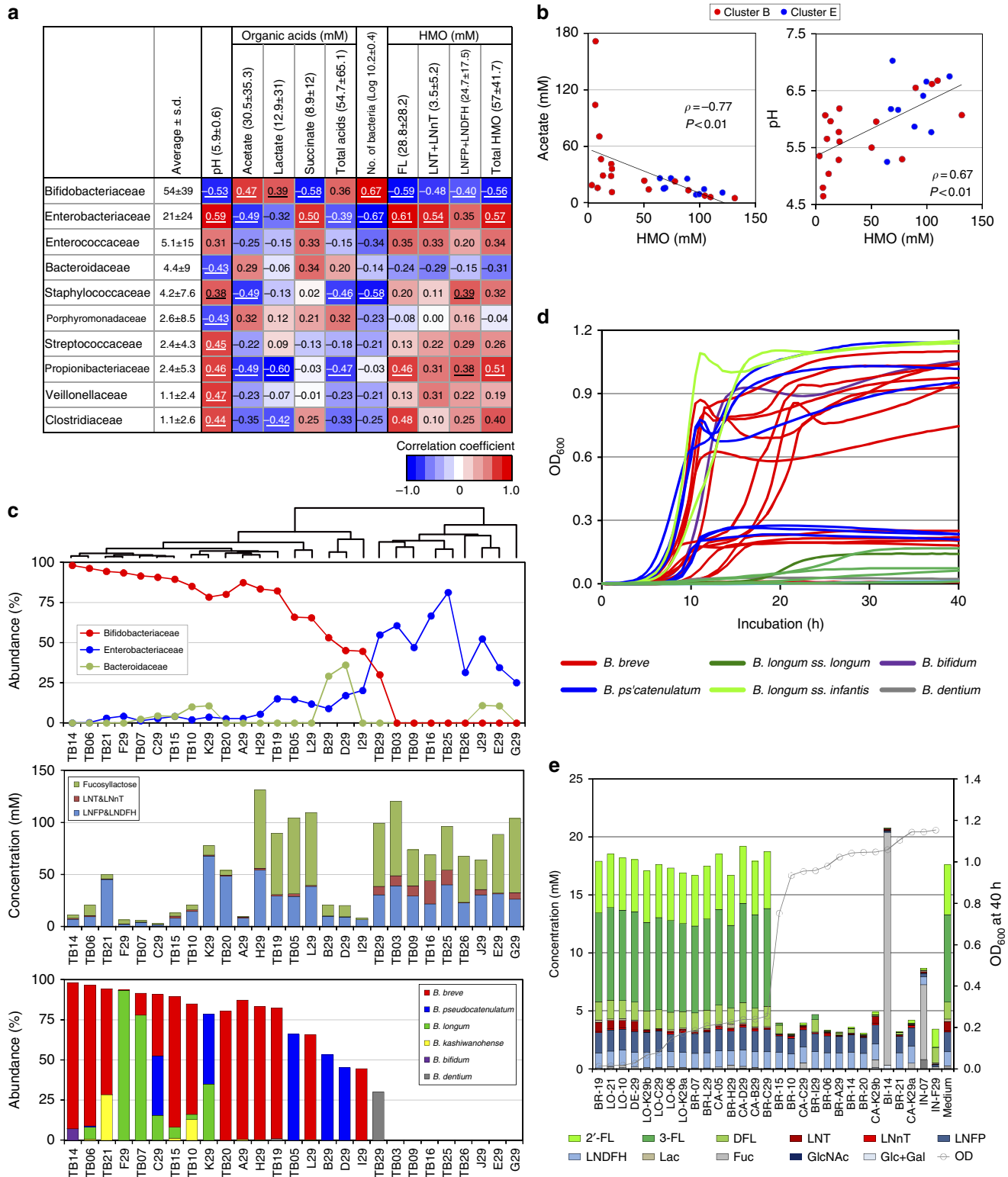
bifidobacteria consume the remaining oligosaccharides in the infant gut, causing elevated acetate and lactate concentrations and decreased pH. Furthermore, we examined the concentrations of major residual HMO components in infant stools by high-performance liquid chromatography (HPLC; Supplementary Data 3 and Supplementary Figs 11 and 12)<sup>26</sup>. As expected, HMO consumption in the gut was associated with increased abundances of Bifidobacteriaceae (Fig. 3a), increased organic acid concentrations and decreased pH values (Fig. 3b). However, some infants showed high faecal oligosaccharide concentrations despite the presence of Bifidobacteriaceae (Fig. 3c). Therefore, we examined the species composition of bifidobacteria in these

infants (Fig. 3c). We also analysed the oligosaccharide profiles of breast milk provided by their mothers, and found that three infants (subjects I29, TB16 and TB19) received breast milk from non-secretor mothers (lacking 2'-fucosyllactose (2'-FL) and 2', 3-difucosyllactose (DFL), Supplementary Table 6). However, we found it difficult to explain these discordant results.

These observations prompted us to isolate bifidobacteria from the infant faeces to assess their ability to utilize HMOs *in vitro*. The 29 isolated strains were cultured in medium containing HMOs as the carbon source and their growth was monitored (Fig. 3d and Supplementary Figs 13 and 14). Interestingly, 14 of the 29 strains exhibited remarkable growth in HMO medium



**Figure 2 | Gut microbiota community profiles of 27 1-month-old infants and 22 adults.** (a) Bacterial families representing more than 1% (on average) of the microbiota in infants or adults are shown on the colour scale. Samples were hierarchically clustered by measuring Euclidean distances with complete-linkage clustering, as shown in the upper tree. (b) Characteristics of infant and adult gut microbiota, as illustrated by PCoA and PAM clustering analyses. Cluster B, Bifidobacteriaceae-predominant; cluster E, Enterobacteriaceae-predominant; cluster AD, adult-type microbiota. (c) Abundances of the main contributors to each cluster. Different letters (a–c) indicate significant differences between clusters ( $P < 0.05$ , Mann–Whitney  $U$ -test with Bonferroni’s correction). Differences in other bacterial families are shown in Supplementary Fig. 7). (d) Network diagram showing co-occurrence relationships among the main contributors in infants and adults. Node sizes indicate the abundances of each bacterial family, and the widths of the edges reflect the calculated Spearman’s rank correlation coefficient.



**Figure 3 | Relationships between bacterial family abundances and gut environments.** (a) Spearman's rank correlation coefficients between bacterial family abundances and gut environmental factors such as pH, organic acid concentrations and faecal oligosaccharide concentrations are shown in numerical and colour-scale formats. (b) Spearman's rank correlations of oligosaccharide concentrations with pH values and acetate concentrations. (c) Relationships between bacterial abundances, faecal oligosaccharide concentrations and the relative abundances of bifidobacterial species. The upper tree shows hierarchical clustering on the basis of the bacterial family compositions. (d) Growth curves of 29 bifidobacterial strains in medium containing HMOs (see Supplementary Fig. 14 for more details). (e) Glycoproteomic profiles of bacterial supernatants after 40 h of cultivation. Samples are ordered based on their OD<sub>600</sub> values after 40 h of cultivation.

(saturating  $OD_{600} > 0.7$ ), but 15 strains did not (saturating  $OD_{600} < 0.3$ ). Culture supernatants were collected to investigate the remaining oligosaccharides. Although most bifidobacterial strains utilized lacto-*N*-tetraose (LNT), there were considerable differences in the utilization of fucosyllactose (sum of 2'-FL, 3-fucosyllactose and DFL), which is the main component of HMOs<sup>27,28</sup> (Fig. 3e and Supplementary Fig. 15). These results indicated that efficient FL utilization is not a universal property of infant bifidobacteria and is instead strain-dependent.

**Bifidobacterial genomes and FL utilization.** To gain insight into how the strains showed differences in FL utilization, we determined the draft genomes of all 29 strains (Table 1 and Supplementary Table 7) and subsequently performed OrthoMCL clustering analysis using bi-directional BLAST alignments (Supplementary Fig. 16a). Initially, we investigated the presence of fucosidase genes (glycoside hydrolase family classifications: GH95 and GH29) and confirmed that all strains exhibiting robust growth in an HMO-containing medium possessed at least one fucosidase gene (Table 1 and Supplementary Fig. 16b). However, we found that 6 out of 15 strains showing limited growth in the HMO-containing medium possessed a fucosidase gene. We analysed the phylogeny and subcellular localization of the fucosidase genes. We found that most fucosidases, except for that of *B. bifidum* BI-14, are intracellular enzymes, with only minor sequence differences occurring between FL-utilizing and non-FL-utilizing strains (Supplementary Fig. 17).

Subsequently, we searched for other genes responsible for FL utilization and discovered that the presence of a homologous group (HG\_2571) corresponded with the FL-utilization phenotype in all strains, except for *B. bifidum* BI-14 (Table 1 and Supplementary Fig. 16c). BLAST searches against the KEGG database indicated that the homologous sequences were highly similar to substrate-binding protein (SBP), which participates in the multiple-sugar ABC transporter system (KEGG entry K02027; Supplementary Fig. 16c). Furthermore, we found that two permease genes of the multiple-sugar ABC transporter system (K02025 and K02026) and the fucosidase gene (GH95) were adjacent to the SBP gene in most strains (Fig. 4a). On the basis of these findings, we hypothesized that the ABC transporter mediates FL transportation into bacterial cells. This hypothesis could explain the absence of the transporter in the *B. bifidum* BI-14 strain, which utilizes FL because that strain

expresses extracellular membrane-bound fucosidases (Supplementary Fig. 17).

To determine whether the putative ABC transporter SBP for FL (denoted FL-SBP) mediates FL utilization, the gene was knocked out in *B. breve* BR-A29 using homologous recombination (Supplementary Fig. 18). After confirming that the *FL-SBP* gene was knocked out, growth of the knockout strain was investigated in the HMO medium. In contrast with the original BR-A29 strain, the FL-SBP gene-knockout strain showed limited growth in the HMO medium (Fig. 4b). Furthermore, we confirmed that FL was not utilized by the FL-SBP gene-knockout strain (Fig. 4c), demonstrating that the FL-SBP is responsible for FL utilization.

**HMO-utilizing bifidobacteria affect gut ecosystems.** Having identified the differences in FL utilization among bifidobacteria, we subdivided the Bifidobacteriaceae-dominated microbiota (cluster B) into those colonized by FL-utilizing bifidobacteria (designated cluster B1;  $n = 11$ ) and those dominated with non-FL-utilizing bifidobacteria (cluster B2;  $n = 7$ ; Supplementary Fig. 19). Furthermore, we compared the faecal organic acids, pH, HMO and microbiota compositions of these two subgroups with those of Enterobacteriaceae-dominated microbiota (cluster E;  $n = 9$ ). Compared with clusters B2 and E, cluster B1 showed significantly higher acetate concentrations and lower pH and residual oligosaccharide concentrations ( $P < 0.05$ , Mann-Whitney *U*-test with Bonferroni's correction; Fig. 5 and Supplementary Fig. 20). In addition, cluster B1 had significantly higher Bifidobacteriaceae and lower Enterobacteriaceae abundances. In contrast, there were no significant differences in faecal acetate concentrations, pH or oligosaccharide concentrations between clusters B2 and E.

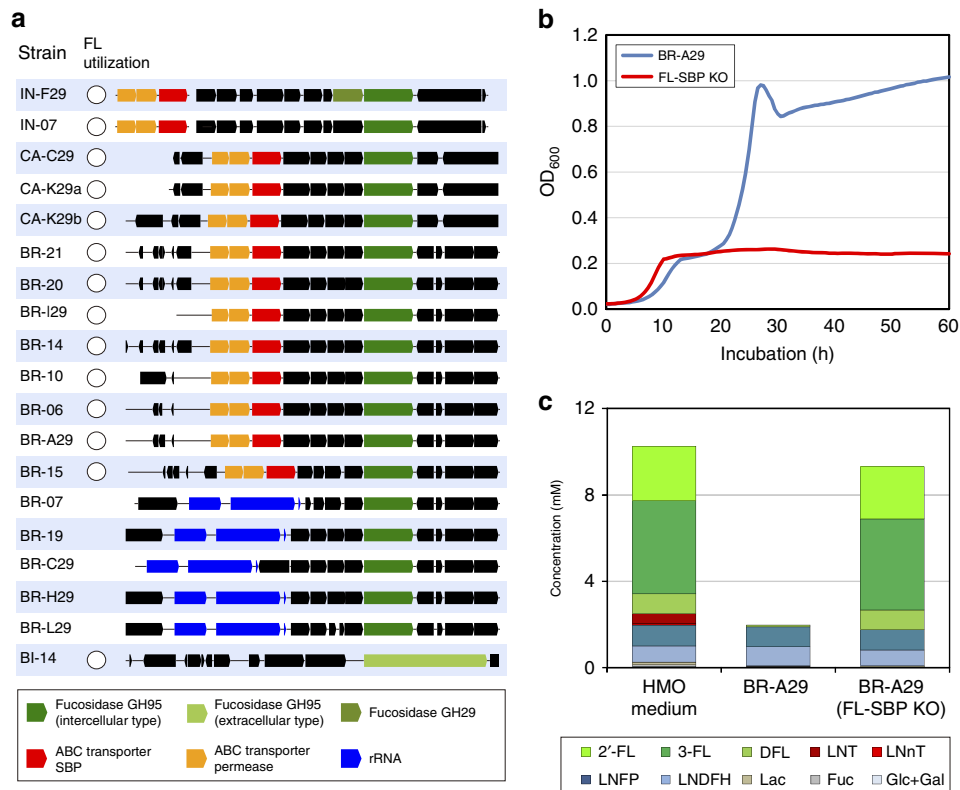
## Discussion

Gut microbiota development in healthy, full-term infants has been investigated using culture-based enumeration<sup>29</sup>, molecular analysis with 16S rRNA gene-targeting primers or probes<sup>10–12,30</sup>, or, more recently, using metagenomic approaches<sup>8,9,13,31</sup>. These intensive observational studies have demonstrated that the modes of delivery<sup>10,13</sup>, feeding<sup>10,13,29</sup> and use of antibiotics<sup>12</sup> can influence the development of the infants' microbiota. In addition, recent investigations have suggested that maternal HMO secretion type<sup>31</sup>, environmental exposures<sup>10,14</sup> and bacterial transmission and propagation<sup>30</sup> can also alter the

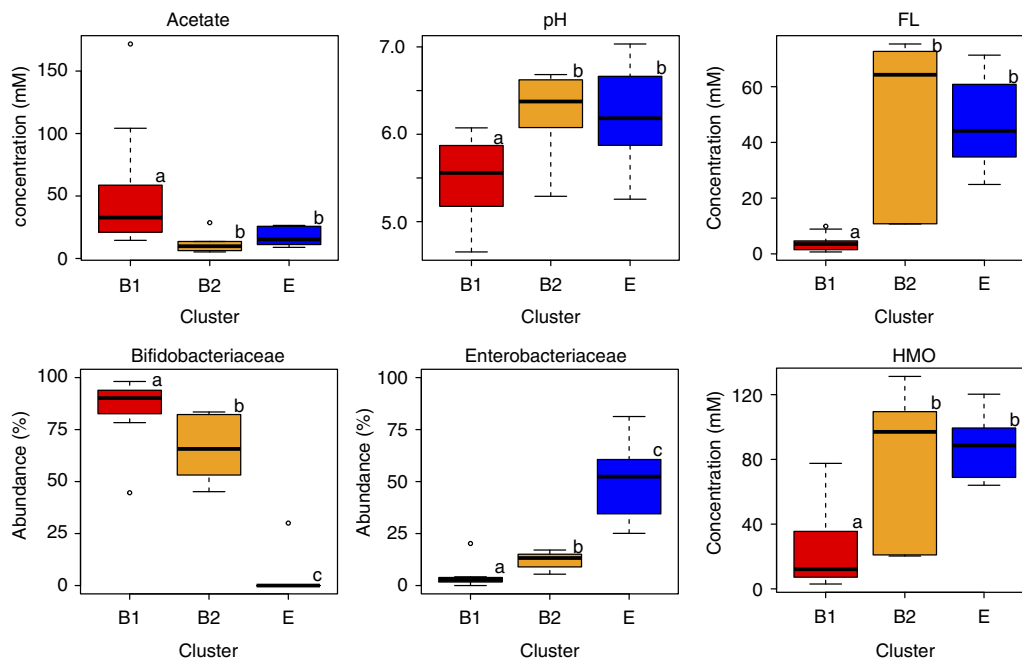
**Table 1 | Summary of the draft genomes of the 29 bifidobacterial strains.**

Strains	Genome size (Mbp)	No. of CDSs	Growth with HMO ( $OD_{600} > 0.7$ )	$\alpha$ -L-fucosidase		HG_2571 SBP of ABC transporter
				GH29	GH95	
<i>B. breve</i> BR-06, BR-10, BR-14, BR-15, BR-20, BR-21, BR-A29, BR-I29 (eight strains)	2.2–2.7	1,928–2,368	+	–	+	+
<i>B. breve</i> BR-07, BR-19, BR-C29, BR-H29, BR-L29 (five strains)	2.2–2.5	1,955–2,224	–	–	+	–
<i>B. longum</i> ss. <i>infantis</i> IN-07, IN-F29 (two strains)	2.6–2.7	2,356–2,441	+	+	+	+
<i>B. longum</i> ss. <i>longum</i> LO-06, LO-10, LO-21, LO-C29, LO-K29a, LO-K29b (six strains)	2.4–2.7	1,987–2,209	–	–	–	–
<i>B. pseudocatenulatum</i> CA-C29, CA-K29a, CA-K29b (three strain)	2.2–2.5	1,825–2,168	+	–	+	+
<i>B. pseudocatenulatum</i> CA-05, CA-B29, CA-D29 (three strains)	2.2–2.3	1,896–1,904	–	–	–	–
<i>B. bifidum</i> BI-14	2.2	1,779	+	+	+	–
<i>B. dentium</i> DE-29	2.6	2,127	–	+	–	–

HMO, human milk oligosaccharide; SBP, substrate-binding protein.



**Figure 4 | Identification of bifidobacterial genes responsible for HMO utilization.** (a) Syntenic relationships of putative FL-utilization gene clusters identified among the bifidobacteria. The genes and their orientations are depicted with arrows. Strains lacking fucosidase genes (10 strains among 29) are not represented. (b) Growth curves of the *B. breve* BR-A29 strain and the corresponding FL-SBP gene knockout strain in medium containing HMOs. (c) Glycoprofiles of BR-A29 and FL-SBP gene-knockout strains after 40 h of cultivation.



**Figure 5 | Impact of FL-utilizing bifidobacteria on gut microbial ecosystems.** Box plots showing differences among Bifidobacteria-dominant microbiota, with or without FL-utilizing bifidobacteria (clusters B1 and B2, respectively) and Enterobacteriaceae-dominant microbiota (cluster E). Different letters (a-c) above the boxes indicate significant differences between clusters ( $P < 0.05$ , Mann-Whitney *U*-test with Bonferroni's correction).

assemblage of infants' microbiota. However, the overall understanding of gut microbiota development is still limited because of the small number of subjects studied<sup>9</sup>, the low

frequency of sample analysis<sup>8,13,31</sup> and/or the limitations in the enumeration methods used<sup>10-12</sup>. In addition, metabolite profiles, bacterial strain isolation and related phenotype and genotype

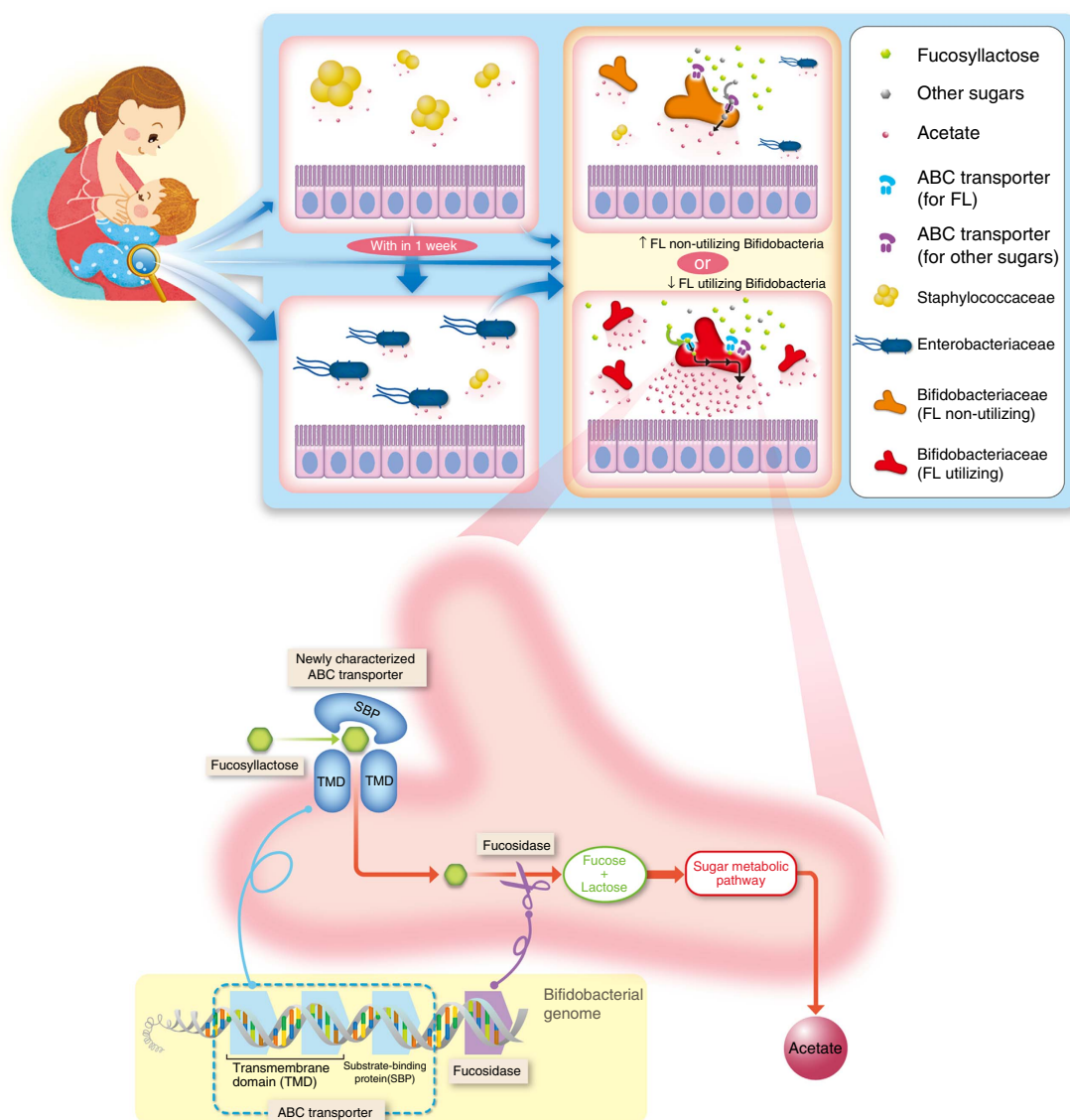
investigations were not performed in these studies. Thus, the presence of infant microbiota equilibria, key bacterial lineages and genetic factors regulating gut metabolite profiles and the molecular mechanisms that influence the gut microbiota in early life remain poorly understood.

Among the considerable number of metabolic pathways, we demonstrated the importance of the FL-utilization pathway, which is associated with high concentrations of acetate in the gut and alters the microbiota composition (Fig. 6). To our knowledge, this is the first study suggesting that a single bacterial gene may affect the gut metabolic profile and microbiota composition in a human cohort. Previous metagenomics and metabolomics analyses in infants have suggested that gut microbial communities consistently serve the same major functions, regardless of variations in their compositions<sup>18</sup>. In contrast, we focused on the utilization of HMOs, the most abundant carbohydrates in the infant gut, and found that the important gut microbial metabolic

activity was associated with a specific group of infant gut microbes in this study.

Genomic analysis of bifidobacteria and subsequent experiments with target gene knockout strain identified an ABC transporter that plays an essential role in FL utilization. Previous studies have demonstrated that some bifidobacteria can efficiently utilize HMO components, such as LNT<sup>22,23</sup>. Differences in FL utilization have also been observed between different species and strains<sup>32,33</sup>. However, the mechanism and importance of FL transport into bacterial cells have not been demonstrated.

Recently, Lewis *et al.*<sup>31</sup> investigated the relationship between maternal secretor status and the development of infant microbiota, and reported that infants fed by non-secretor mothers are delayed in the establishment of microbiota dominated by bifidobacteria. However, the FL-utilizing properties of bifidobacteria and the molecular mechanisms of FL utilization characterized in this study were not taken account.



**Figure 6 | Infant microbiota development and molecular mechanisms of FL utilization by bifidobacteria.** Infant gut microbiota showed 3 distinct clusters, which underwent directional transition to Bifidobacteriaceae-dominant microbiota, and displayed individual variations in the pace of progression. Isolated bifidobacterial strain showed differences in FL utilization. Colonization of FL-utilizing bifidobacteria are associated with altered gut acetate concentrations, pH, and Enterobacteriaceae and Bifidobacteriaceae abundances in the cohort, which have been previously shown to affect infant health. We subsequently identified a SBP of the multiple-sugar ABC transporter system is a key genetic factor of FL utilization.



In this study, we found that 3 out of 27 infants received breast milk from non-secretor mothers, and one infant (subject I, Supplementary Fig. 19) harboured FL-utilizing bifidobacteria. The infant exhibited considerably lower Bifidobacteriaceae abundance, higher Enterobacteriaceae abundance, lower acetate concentrations and higher pH (45%, 20%, 16.2 mM and 6.2, respectively; Supplementary Data 2 and 3) compared with the other cluster B1 infants (Fig. 5). These results support our notion that the presence of FL in breast milk and colonization of FL-utilizing bifidobacteria induce altered gut microbiota composition and metabolite profiles, consistent with the observation by Lewis *et al.*<sup>31</sup> Larger infant cohorts are needed to identify the relationship between maternal secretor status and colonization of FL-utilizing and non-utilizing bifidobacteria in order to draw more robust conclusion.

A recent study reported that the ‘commensal colonization factors’ encoding a unique class of polysaccharide utilization genes play important roles in stable colonization of *Bacteroides*, one of the most predominant genera in the human adult microbiota<sup>34</sup>. In the current study, we identified the FL transporter as a key stable colonization factor of gut microbes in infants. More importantly, the FL transporter was shown to cause changes in the gut microbiome and the production of metabolites, which may confer benefits to infants. Thus, these data indicate that the FL transporter is a symbiotic colonization factor of infant bifidobacteria, which is conceptually distinct from the commensal colonization factor of *Bacteroides* colonizing the adult gut.

Changes in metabolite profiles and microbiota compositions caused by FL-utilizing bifidobacteria have been suggested to have a variety of beneficial effects on the host<sup>1,4,35–38</sup>. It has been demonstrated that the acetate produced by bifidobacteria protects against *Escherichia coli* O157 infection by improving the gut barrier<sup>1</sup> or inhibiting Shiga-toxin production<sup>35</sup> in animal models. Other studies have reported that the short-chain fatty acids produced by gut microbes are associated with host energy balance<sup>5,36</sup>, inflammation resistance<sup>4</sup> and sympathetic nervous system development<sup>37</sup>. In addition, correlations between decreased Enterobacteriaceae abundance and lower susceptibility to infection have been demonstrated in animal models<sup>2</sup> and human cohorts<sup>3</sup>. Thus, the ABC transporter that is present in HMO-utilizing bifidobacteria may benefit the host by facilitating acetate production and limiting the abundance of Enterobacteriaceae.

Given the importance of infant gut microbiota for the long-term health of the individual, the FL-utilizing properties of bifidobacteria and related molecular mechanisms may ultimately have implications in therapeutic approaches or preventive medicine. Our findings suggest that the presence of FL in breast milk and its utilization by bifidobacteria help fostering the mutually beneficial relationship between humans and the main constituents of the infant gut microbiota. Since the ABC transporter is present in only a subset of bifidobacteria, future studies should investigate the effects of FL-utilizing bifidobacteria on the health of individuals using human interventions or animal models. The phenotypic and genotypic properties of bifidobacteria should be considered during the development of probiotics targeting the gut microbiota of infants.

## Methods

**Subject recruitment and sample collection.** The study was approved by the ethical committees of Yakult Central Institute (subjects A to L; Supplementary Table 1) and Teikyo University (subjects TB–XX; Supplementary Table 4). Written informed consent was obtained from the parents before enrolment. No infants received antibiotics or probiotics before or during the study period. Infant stool samples were collected by the parents, ~17 times each from 12 infants (subjects A to L) and once from 15 infants (Supplementary Table 4). Additional stool samples

were collected from most parents of infants A to L (Supplementary Table 5). The samples were frozen at  $-20^{\circ}\text{C}$ , transferred to the laboratory and stored at  $-80^{\circ}\text{C}$ . In total, 239 samples were collected for gut microbiota analysis, including 217 stool samples from infants and 22 stool samples from parents. Parents were instructed to record changes in diet, medications, hospitalizations, birth weight and gestation age at the time of delivery (Supplementary Tables 1 and 4). HMOs were purified from breast milk provided by nine healthy volunteers ( $32.6 \pm 3.9$  years of age) at approximately 1 year following delivery.

**DNA extraction.** DNA was extracted as described in ref. 39. Briefly, 20 mg of each faecal sample (equivalent to 100  $\mu\text{l}$  of fivefold homogenate) was suspended in 500  $\mu\text{l}$  of extraction buffer (100 mM Tris-HCl, pH 9.0, 40 mM EDTA and 1% SDS; final concentrations) in a 2-ml screw-cap tube. Then, 300 mg of 0.1-mm-diameter glass beads and 500  $\mu\text{l}$  of TE buffer-saturated phenol were added to the suspension. Microbial cells were lysed by mechanical disruption using a FastPrep FP 120 (BIO 101, Vista, CA, USA) at a power level of 5.0 for 30 s. The mixtures were centrifuged at 4,500g for 5 min, and the upper layers were subjected to phenol/chloroform/isoamyl alcohol extraction, followed by isopropanol and ethanol precipitation. Finally, the dried DNA samples were suspended in 1 ml TE buffer and stored at  $-35^{\circ}\text{C}$  until they were pyrosequenced.

**16S rRNA gene amplification and pyrosequencing.** The V1–V2 regions of 16S rRNA gene were selected for pyrosequencing analysis because that region offers enough sequence variation for species-level discrimination of bifidobacteria, which are reported as the main constituent of infant gut microbiota. The 16S rRNA genes of each sample were amplified using a forward 63Fm-TAG-linker A primer (5'-CCATCTCATCCCTGCGTGTCTCCGAC-TCAG-NNNNNNNNNN-GCYTAAAYACATGCAAGTMTGA-3') and a reverse 338R-linker B primer (5'-CCTATCCCTGTGTGCCTTGGCAGTCTCAG-GCTGCCWCCCGTAGG-WGT-3').

The italicized sequence represents 454 Life Sciences linkers A and B, respectively; TCAG in forward primer was inserted as a key sequence as recommended by the manufacturer; NNNNNNNNNN in forward primers represents the unique 10-base barcode for tagging each PCR product; and the underlined sequence is a broadly conserved sequence in bacterial 16S rRNA genes. We used universal primer 63F (ref. 40) with modification since bifidobacteria have a two-base-pair mismatch with respect to the universal primer 27F (ref. 41). The taxonomic coverage of the 63Fm and 338Rm primer sequences was evaluated with the Probe Match programme in the Ribosomal Database Project (RDP; Release 11, Update 2)<sup>42</sup>. A total of 1,354,916 16S rRNA gene sequences in the RDP were retrieved using the following parameters: strain = both, source = both, size  $\geq 1,200$ , quality = good. The sequences retrieved were used as reference sequences for the Probe Match programme to evaluate the taxonomic coverage of the primer pair. The taxonomic coverage of the primer pair for each phylum was calculated as the percentage of genera for which more than half of the member sequences in each genus were found to match the candidate sequence with no more than one mismatch<sup>43</sup>. The PCR mixture (50  $\mu\text{l}$  total volume) contained  $1 \times$  SYBR Premix Ex Taq (Takara Bio, Osaka, Japan), 100 nM of each primer and 1  $\mu\text{l}$  of template DNA. The thermocycling conditions used were as follows:  $95^{\circ}\text{C}$  for 5 min, followed by 25–40 cycles of  $95^{\circ}\text{C}$  for 30 s,  $55^{\circ}\text{C}$  for 30 s and  $72^{\circ}\text{C}$  for 1 min. Amplification was performed using an ABI PRISM 7500 Real-Time PCR System (Applied Biosystems, Framingham, MA, USA), and thermal cycling was stopped before the amplification curves reached a plateau to avoid introducing biases to the microbiota composition and generation of chimera. Amplicons were purified with the AMPure XP Kit (Beckman Coulter Genomics GmbH, Bernried, Germany) and quantified using the Quant-iT PicoGreen dsDNA Kit (Invitrogen, Leek, the Netherlands). Equal amounts of the amplicons were pooled and sequenced using a Roche 454 GS Junior pyrosequencer with a GS FLX Titanium emPCR Kit (Lib-L) according to the manufacturer's protocols (Roche Applied Science, Bavaria, Germany).

**16S rRNA gene sequence-based microbiota analysis.** The sequences generated from the 454 GS Junior platform were analysed using the open-source software package QIIME<sup>19</sup>. The resultant 16S rRNA gene sequences were assigned to operational taxonomic units (OTUs) using the USEARCH algorithm<sup>44</sup>, with a 97% identity threshold. Simultaneously, UCHIME<sup>45</sup> was employed to remove potential chimeric sequences from a representative set of OTUs, using the reference mode (against ‘Gold’ database<sup>46</sup>). The taxonomy of each representative OTU sequence was assigned using the RDP-naive Bayesian Classifier with a minimum bootstrap threshold of 50% (ref. 47). A single representative from each OTU was aligned using the MUSCLE alignment tool<sup>48</sup>, and a phylogenetic tree was constructed using FastTree<sup>49</sup>.  $\alpha$ -diversities (the number of OTU observed, Shannon index and phylogenetic diversity) were estimated for 1,000 randomly selected sequences to account for differences in sampling effort between the samples. Detailed QIIME commands and options are summarized in Supplementary Table 3.

PCoA and between-class analysis were performed as described in ref. 20. Data generated by QIIME at taxonomic level 5 (family) were used to calculate the Jensen–Shannon distances between samples. The PAM clustering algorithm<sup>50</sup> was applied to cluster the profiles. Estimation of the number of clusters in infant gut microbiota by PAM clustering according to the method in ref. 20 with modifications. We examined the number of clusters exhibiting the highest CH

index<sup>51</sup> in the randomly subsampled data set (see Supplementary Figs 3 and 4 for more details). This trial was repeated 1,000 times. A cluster number most frequently exhibiting the highest CH index in the 1,000 trials was defined as the optimal number of microbiota cluster.

**Total bacterial counts.** Approximately 1 g of faecal samples were diluted fivefold with PBS, homogenized, and a portion of the homogenates (100 µl) was fixed with 300 µl of 4% paraformaldehyde at 4 °C for 16 h. The paraformaldehyde-treated samples were smeared on a MAS-coated slide glass (Matsunami, Osaka, Japan) and stained with Vectashield Mounting Medium with 4,6-diamidino-2-phenylindole (Vector Laboratories, Burlingame, CA, USA). The fluorescent cells were counted in an automated format using a Leica MM AF microscope (Leica, Wetzlar, Germany) and Image-Pro Plus Image Analysis Software (version 5.1; Media-Cybernetics, Silver Spring, MD, USA). Bacterial counts were expressed as the mean values of 10 fields for each sample.

**pH values and organic acid concentrations.** The pH values of the faecal samples were measured directly with a handheld pH meter (model IQ150) equipped with a PH17SS electrode (IQ Scientific Instruments, San Diego, CA, USA). To determine the concentration of organic acids, 180 µl of faecal homogenates (fivefold, prepared as described above) were mixed with 20 µl 10% HClO<sub>4</sub> and incubated at 4 °C overnight. Subsequently, the samples were centrifuged for 5 min at 14,000g and the supernatants were filtered through Centriscut W-MO membrane filters (0.45 µm; Kurabo, Tokyo, Japan). The concentrations of organic acids were determined with a HPLC system equipped with 432 electroconductivity detectors (Waters, Milford, MA, USA) and a Rspak KC-811 column (Showa Denko KK, Tokyo, Japan)<sup>52</sup>.

**Statistics.** Statistical analyses were performed using the R software, version 3.1.0 (<http://www.r-project.org/>) and Excel 2013 (Microsoft). Differences in bacterial abundance,  $\alpha$ -diversities and other host metadata between the clusters were evaluated by the Mann–Whitney *U*-test, with Bonferroni's correction. Differences between clusters were represented with box plots. The middle bar of each plot indicates the median, while the top and bottom of each box indicate the third and first quartiles, respectively. Whiskers denote the lowest and highest values within 1.5 times the interquartile range from the first and third quartiles, respectively. Circles denote outliers beyond the whiskers. Correlations between host metadata, microbiota (bacterial family abundances,  $\alpha$ -diversities and total bacterial counts) and gut environmental factors (pH, organic acid and oligosaccharide concentrations) were analysed by calculating Spearman's rank correlation coefficients. To construct the network diagram, Spearman's correlations were computed between the bacterial families representing more than 1% (on average) of the microbiota (Supplementary Figs 8 and 9). Only relationships having an absolute Spearman's correlation above 0.3 with a *P*-value less than 0.05 were selected. The yEd Graph Editor (<https://www.yworks.com/products/yed>) was then used to construct the network figures.

**Preparation of HMOs.** Breast milk samples (~11 total) were combined, mixed with 4 volumes of chloroform:methanol (2:1) and stirred vigorously for 1 min. After centrifugation at 9,000g, the upper layer was collected, evaporated and dissolved in H<sub>2</sub>O. Two volumes of ethanol were added to it and incubated overnight at 4 °C, and the precipitated lactose was removed by passing samples through filter paper. The ethanol was removed by rotary evaporation, after which the residue was dissolved in H<sub>2</sub>O and applied to an XK column (5 cm in diameter and 100 cm in length; GE Healthcare, Buckinghamshire, UK). Elution was performed overnight using water as the mobile phase, with a flow rate of 1.5 ml min<sup>-1</sup>. Fractions were collected and analysed with HPLC, and those fractions containing HMOs were combined.

**Isolation of bifidobacteria from infant faeces.** Faecal homogenates were diluted 10<sup>5</sup>–10<sup>8</sup>-fold with sterile saline (0.85% NaCl) and spread on bifidobacteria-specific TOS propionate agar plates (Yakult Honsha Co., Tokyo, Japan) supplemented with 50 mg l<sup>-1</sup> mupirocin. The plates were incubated at 37 °C for 3 days in an anaerobic chamber (Coy Laboratory Products, Grass Lake, MI, USA) with 88% N<sub>2</sub>, 5% CO<sub>2</sub> and 7% H<sub>2</sub>. All resulting colonies appearing on the plates inoculated with the highest and second-highest dilutions of faecal homogenate (at least 10 colonies per plate) were then streaked on GAM agar plates (Nissui Seiyaku, Tokyo, Japan). These isolates were further classified by random amplified polymorphic DNA profiling, as described in ref. 53. Isolates with identical random amplified polymorphic DNA patterns using two primers (p1254, 5'-CCGACGCCAA-3'; p1281, 5'-AACGCGCAAC-3') were regarded as identical strain. Each strain was grown overnight in GAM broth (Nissui Seiyaku), centrifuged, suspended in GAM broth with 20% glycerol and stored at -80 °C. For further analysis, the strains were passaged twice in GAM broth and grown in an anaerobic chamber.

**In vitro growth of bifidobacterial strains in the presence of HMO as the sole carbon source.** The bacterial strains listed in Fig. 4a were cultured in PY medium (200 mM PIPES, pH 6.7, 2 g l<sup>-1</sup> Peptone, 2 g l<sup>-1</sup> BBL Trypticase Peptone, 2 g l<sup>-1</sup> Bacto Yeast Extract, 8 mg l<sup>-1</sup> CaCl<sub>2</sub>, 19.2 mg l<sup>-1</sup> MgSO<sub>4</sub>·7H<sub>2</sub>O, 80 mg l<sup>-1</sup> NaCl,

4.9 mg l<sup>-1</sup> hemin, 0.5 g l<sup>-1</sup> L-cysteine hydrochloride and 100 ng l<sup>-1</sup> vitamin K<sub>1</sub>), which was supplemented with lactose or HMOs. The PY medium (200 µl) was inoculated with bifidobacterial strains growing at an exponential phase (equivalent to OD<sub>600</sub> = 5 × 10<sup>-4</sup>) and covered with 50-µl sterile mineral oil to prevent evaporation. Growth was monitored by measuring the OD<sub>600</sub> using a PowerWave 340 plate reader (BioTek, Winooski, VT, USA) every 30 min in the anaerobic chamber. Two technical replicates were performed for each strain. The supernatants of the bacterial cultures were collected at the end of the exponential growth phase (40 h) and stored at -80 °C for subsequent glycoprofiling.

**Glycoprofiling of HMOs.** Oligosaccharides in bacterial supernatants or faecal suspensions were labelled with an ultraviolet light-absorbing compound, *p*-aminobenzoic acid ethyl ester (ABEE), and quantified using HPLC, as described in ref. 26. Briefly, 200 µl of bacterial culture or faecal homogenate was used as the specimen, and arabinose was added as an internal standard (1 mM final concentration). The cultured media were mixed with 800 µl of chloroform:methanol (2:1) and stirred vigorously for 1 min. The upper layer was collected after centrifugation at 9,000g for 10 min and concentrated by evaporation at 70 °C under a nitrogen gas flow. The crude oligosaccharides were dissolved in 200 µl of distilled water. A 10-µl portion of this solution was added to 40 µl of freshly prepared reagent mixture (35 mg ABEE, 3.5 mg NaBH<sub>3</sub>CN, 41 µl acetic acid and 350 µl methanol). The mixtures were heated at 80 °C for 45 min in screw-capped vials, after which 500 µl of water was added. The mixture was extracted three times with 500 µl of diethyl ether to remove excess ABEE. The aqueous layer was evaporated, and the residue was dissolved in 300 µl of distilled water. Ten-microlitre samples of labelled HMOs were then separated on a Shimadzu Prominence HPLC System (Kyoto, Japan) with an L-column 2 ODS (Chemicals Evaluation and Research Institute, Japan). Fractions were eluted with an 87:13 (vol:vol) mixture of 100 mM ammonium acetate buffer (pH 4.5) and acetonitrile, using flow rates of 1.0 ml min<sup>-1</sup> (0–33 min) and 2.0 ml min<sup>-1</sup> (33–55 min) at 40 °C. Labelled oligosaccharides were detected with an SPD-20 A ultraviolet detector (Shimadzu) at 304 nm. Standard curves for the major HMO components were generated using solutions containing both the internal standard (1 mM arabinose) and oligosaccharides at varying concentrations. The peak areas of the saccharides detected were standardized and quantified by comparison with the internal standard. 3-fucosyllactose, lacto-*N*-fucopentaose (LNFP) I, lacto-*N*-difucosylhexose (LNDFH) I and LNDFH II were purchased from Dextra Laboratories (Reading, UK). 2'-FL, DFL, LNT, lacto-*N*-neotetraose, LNFP II and LNFP III were purchased from IsoSep AB (Tullinge, Sweden). *N*-acetyl glucosamine was purchased from Nacalai Tesque (Kyoto, Japan). These structures are illustrated in Supplementary Fig. 12.

**Genome sequencing.** The draft genomes of 29 bifidobacterial strains were sequenced with the MiSeq sequencing platform (Illumina, San Diego, USA). Multiplexed shotgun libraries were constructed using the Nextera XT DNA Sample Prep Kit v2 (Illumina). The resulting paired-end sequence reads (250 bp × 2) were assembled into contigs using the assembly software ABySS 1.3.5 (ref. 54). The optimal k-mer size was predicted using VelvetK (<http://bioinformatics.net.au/software/velvetk.shtml>; maximum value = 128). To identify adjacent gene of fucosidase (GH29 and 95) and *FL-SBP*, draft contigs were aligned to the published genome of *B. breve* UCC2003 (Accession Number CP000303)<sup>55</sup> or *B. longum* ss. *infantis* ATCC 15697 (CP001095)<sup>22</sup> using BLAST to predict the order and orientation. Oligonucleotide primers were designed to anneal to each end of the neighbouring contigs, and PCR amplification was performed to confirm the contig ordering. The resulting amplicons were subjected to Sanger sequencing. Draft genome sequences of 29 bifidobacterial strains were deposited in the DDBJ whole-genome shotgun database under BioProject Accession No. PRJDB4041.

**Annotation and comparative genome analysis.** Protein-coding sequences (CDSs) on >500-bp contigs were predicted using the gene prediction software Glimmer 3.02 (ref. 56). Functional annotations of CDSs were performed by BLASTP searches (version 2.2.24)<sup>57</sup> against the KEGG database<sup>58</sup>, with an *e*-value cutoff of 10<sup>-6</sup>. All-against-all BLASTP searches for all proteins from the 29 strains were performed with an *e*-value cutoff of 10<sup>-5</sup> and a minimum identity of 55%, and the results were clustered using the OrthoMCL method<sup>59</sup>. rRNA genes were identified by BLASTN analysis using known bifidobacterial rRNA sequences as queries. The genes annotated as fucosidase (both GH29 and GH95) by BLASTP searching were aligned using Clustal X (ref. 60) and were used as inputs for calculating phylogenetic relationships. Phylogenetic trees were computed with the neighbour-joining method using Clustal X and were visualized using FigTree (<http://tree.bio.ed.ac.uk/software/figtree>; Supplementary Fig. 17). Localization of the fucosidase proteins was predicted based on their amino-acid sequences using the PSORT software (<http://psort.hgc.jp/form.html>).

**Multilocus sequence analysis.** Identification of isolated strains by multilocus sequence analysis was performed by targeting seven reference genes (16S rRNA, *clpC*, *fusA*, *groEL*, *gyrB1*, *purF* and *xfp*). The genome sequences of *B. breve* UCC2003 (Accession Number CP000303) were used as queries for sequence identity searches, as described in ref. 61. The homologue sequences of the target

genes were extracted from the 29 bifidobacterial strains by BLAST, concatenated and aligned with Clustal X (ref. 60). Phylogenetic trees were computed and visualized as described above.

**Construction of FL-SBP insertion mutant in *B. breve* BR-A29.** The insertional mutagenesis procedure used to inactivate the putative FL-SBP gene in *B. breve* BR-A29 is summarized in Supplementary Fig. 18. Briefly, a DNA fragment of the *E. coli* and *Bifidobacterium*-derived shuttle vector pBEA26 (generated from pBEA4 (ref. 62) by replacing the erythromycin-resistance gene *ermE* with the tetracycline-resistance gene *tetW*) was generated by PCR amplification using KOD-plus-DNA polymerase (Toyobo, Osaka, Japan). Amplification proceeded via the following primers: SacI0001F (5'-TCCGAGCTCCAGGTGGCACTTTTCGGGGAAATG-3') and Kpn4067R (5'-GGGGTACCCCAAGGAAATGGCTATCAACGGTA-3'), where the underlined sequences represent SacI and KpnI restriction sites, respectively. To construct the p15aAT-KS plasmid, the PCR product was digested with SacI and KpnI, and ligated with multiple cloning site of pBluescript II KS(+) (Agilent Technologies, CA, USA). To construct the p15aAT-FL-SBP-int plasmid, an internal 1,000-bp fragment of FL-SBP (FL-SBP-int) was amplified from *B. breve* BR-A29 genomic DNA with the primers FL-SBP-Bam0194F (5'-CGGGATCCTGGAGCCTACCGTCAAGGCATTCGA-3') and FL-SBP-Bam1193R (5'-CGGGATCCTCGAATAGACCGAGTTAACGTCCTCG-3'). This amplicon was digested with BamHI and ligated into p15aAT-KS. Thereafter, p15aAT-FL-SBP-int (or the control plasmid, pBEA26) was introduced into *B. breve* BR-A29 by electroporation (18 kV cm<sup>-1</sup>, 200 Ω, 25 μF), and tetracycline-resistant transductants were selected. Site-specific recombination was confirmed by colony PCR with the FL-SBP-0078 F (5'-CGACACTAACGGAAGCCAGGCTA-3') and ATKS-0097R (5'-CAGGGTTATTGTCTCATGAGCGGATAC-3') primers.

**Data availability.** Infant gut 16S rRNA gene microbiome data have been deposited in the DDBJ Sequence Read Archive (DRA) under BioProject accession code PRJDB4038. The draft genome sequences of the 29 bifidobacterial strains have been deposited in the DDBJ whole-genome shotgun database under BioProject accession code PRJDB4597. The FL-SBP gene sequence of *B. breve* BR-A29 has been deposited in the DDBJ database under accession code LC068768. The authors declare that all other data supporting the findings of this study are available within the article and its Supplementary Information Files, or from the corresponding author upon request.

## References

- Fukuda, S. *et al.* Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469**, 543–547 (2011).
- Stecher, B. *et al.* Like will to like: abundances of closely related species can predict susceptibility to intestinal colonization by pathogenic and commensal bacteria. *PLoS Pathog.* **6**, e1000711 (2010).
- Dicksved, J., Ellstrom, P., Engstrand, L. & Rautelin, H. Susceptibility to Campylobacter infection is associated with the species composition of the human fecal microbiota. *MBio* **5**, e01212–e01214 (2014).
- Maslowski, K. M. *et al.* Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461**, 1282–1286 (2009).
- Kimura, I. *et al.* The gut microbiota suppresses insulin-mediated fat accumulation via the short-chain fatty acid receptor GPR43. *Nat. Commun.* **4**, 1829 (2013).
- Dogra, S. *et al.* Dynamics of infant gut microbiota are influenced by delivery mode and gestational duration and are associated with subsequent adiposity. *MBio* **6**, e02419–e02414 (2015).
- Kurokawa, K. *et al.* Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* **14**, 169–181 (2007).
- Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
- Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl Acad. Sci. USA* **108**(Suppl 1): 4578–4585 (2011).
- Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. Development of the human infant intestinal microbiota. *PLoS Biol.* **5**, e177 (2007).
- Tsuji, H. *et al.* Molecular monitoring of the development of intestinal microbiota in Japanese infants. *Benef. Microbes* **3**, 113–125 (2012).
- Penders, J. *et al.* Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* **118**, 511–521 (2006).
- Backhed, F. *et al.* Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* **17**, 690–703 (2015).
- La Rosa, P. S. *et al.* Patterned progression of bacterial populations in the premature infant gut. *Proc. Natl Acad. Sci. USA* **111**, 12522–12527 (2014).
- Heijtz, R. D. *et al.* Normal gut microbiota modulates brain development and behavior. *Proc. Natl Acad. Sci. USA* **108**, 3047–3052 (2011).
- Cho, I. *et al.* Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature* **488**, 621–626 (2012).
- Cox, L. M. *et al.* Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell* **158**, 705–721 (2014).
- Kostic, A. D. *et al.* The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe* **17**, 260–273 (2015).
- Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
- Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
- Scardovi, V. in *Bergey's Manual of Systematic Bacteriology*. (eds Krieg, N. R. & Holt, J. G.) (The Williams & Wilkins Co., 1984).
- Sela, D. A. *et al.* The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome. *Proc. Natl Acad. Sci. USA* **105**, 18964–18969 (2008).
- Wada, J. *et al.* *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 (2008).
- Asakuma, S. *et al.* Physiology of consumption of human milk oligosaccharides by infant gut-associated bifidobacteria. *J. Biol. Chem.* **286**, 34583–34592 (2011).
- Zivkovic, A. M., German, J. B., Lebrilla, C. B. & Mills, D. A. Human milk glyco-biome and its impact on the infant gastrointestinal microbiota. *Proc. Natl Acad. Sci. USA* **108**(Suppl 1): 4653–4658 (2010).
- Matsuura, F. & Imaoka, A. Chromatographic separation of asparagine-linked oligosaccharides labeled with an ultraviolet-absorbing compound, p-aminobenzoic acid ethyl ester. *Glycoconj. J.* **5**, 13–26 (1988).
- Castanys-Munoz, E., Martin, M. J. & Prieto, P. A. 2'-fucosyllactose: an abundant, genetically determined soluble glycan present in human milk. *Nutr. Rev.* **71**, 773–789 (2013).
- Albrecht, S., Schols, H. A., van den Heuvel, E. G., Voragen, A. G. & Gruppen, H. Occurrence of oligosaccharides in feces of breast-fed babies in their first six months of life and the corresponding breast milk. *Carbohydr. Res.* **346**, 2540–2550 (2011).
- Benno, Y., Sawada, K. & Mitsuoka, T. The intestinal microflora of infants: composition of fecal flora in breast-fed and bottle-fed infants. *Microbiol. Immunol.* **28**, 975–986 (1984).
- Makino, H. *et al.* Mother-to-infant transmission of intestinal bifidobacterial strains has an impact on the early development of vaginally delivered infant's microbiota. *PLoS ONE* **8**, e78331 (2013).
- Lewis, Z. T. *et al.* Maternal fucosyltransferase 2 status affects the gut bifidobacterial communities of breastfed infants. *Microbiome* **3**, 13 (2015).
- Sela, D. A. & Mills, D. A. Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. *Trends Microbiol.* **18**, 298–307 (2010).
- Ruiz-Moyano, S. *et al.* Variation in consumption of human milk oligosaccharides by infant-gut associated strains of *Bifidobacterium breve*. *Appl. Environ. Microbiol.* **79**, 6040–6049 (2013).
- Lee, S. M., Donaldson, G. P., Mikulski, Z., Boyajian, S., Ley, K. & Mazmanian, S. K. Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* **501**, 426–429 (2013).
- Asahara, T., Shimizu, K., Nomoto, K., Hamabata, T., Ozawa, A. & Takeda, Y. Probiotic bifidobacteria protect mice from lethal infection with Shiga toxin-producing *Escherichia coli* O157:H7. *Infect. Immun.* **72**, 2240–2247 (2004).
- Samuel, B. S. *et al.* Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proc. Natl Acad. Sci. USA* **105**, 16767–16772 (2008).
- Kimura, I. *et al.* Short-chain fatty acids and ketones directly regulate sympathetic nervous system via G protein-coupled receptor 41 (GPR41). *Proc. Natl Acad. Sci. USA* **108**, 8030–8035 (2011).
- Kamada, N., Chen, G. Y., Inohara, N. & Nunez, G. Control of pathogens and pathobionts by the gut microbiota. *Nat. Immunol.* **14**, 685–690 (2013).
- Matsuki, T. *et al.* Quantitative PCR with 16S rRNA-gene-targeted species-specific primers for analysis of human intestinal bifidobacteria. *Appl. Environ. Microbiol.* **70**, 167–173 (2004).
- Marchesi, J. R. *et al.* Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Appl. Environ. Microbiol.* **64**, 795–799 (1998).
- Lane, D. J. in *Nucleic Acid Techniques in Bacterial Systematics*. (eds Stackebrandt, E. & Goodfellow, M.) (John Wiley & Sons Ltd., 1991).
- Cole, J. R. *et al.* Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **42**, D633–D642 (2013).
- Mori, H. *et al.* Design and experimental application of a novel non-degenerate universal primer set that amplifies prokaryotic 16S rRNA genes with a low possibility to amplify eukaryotic rRNA genes. *DNA Res.* **21**, 217–227 (2014).
- Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).

46. Haas, B. J. *et al.* Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* **21**, 494–504 (2011).
47. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
48. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
49. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
50. Kaufman, L. & Rousseeuw, P. J. in *Partitioning Around Medoids (Program PAM), Finding Groups in Data: An Introduction to Cluster Analysis* (John Wiley & Sons Inc., 1990).
51. Calinski, T. & Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **3**, 1–27 (1974).
52. Matsumoto, K. *et al.* Effects of a probiotic fermented milk beverage containing *Lactobacillus casei* strain Shirota on defecation frequency, intestinal microbiota, and the intestinal environment of healthy individuals with soft stools. *J. Biosci. Bioeng.* **110**, 547–552 (2006).
53. Akopyanz, N., Bukanov, N. O., Westblom, T. U., Kresovich, S. & Berg, D. E. DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD fingerprinting. *Nucleic Acids Res.* **20**, 5137–5142 (1992).
54. Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. & Birol, I. ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**, 1117–1123 (2009).
55. Fanning, S. *et al.* Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection. *Proc. Natl Acad. Sci. USA* **109**, 2108–2113 (2012).
56. Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. Identifying bacterial genes and endosymbiont DNA with glimmer. *Bioinformatics* **23**, 673–679 (2007).
57. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
58. Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* **42**, D199–D205 (2014).
59. Li, L., Stoeckert, Jr. C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
60. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
61. Ventura, M. *et al.* Analysis of bifidobacterial evolution using a multilocus approach. *Int. J. Syst. Evol. Microbiol.* **56**, 2783–2792 (2006).
62. Kiwaki, M. & Sato, T. Antimicrobial susceptibility of *Bifidobacterium breve* strains and genetic analysis of streptomycin resistance of probiotic *B. breve* strain Yakult. *Int. J. Food Microbiol.* **134**, 211–215 (2009).

## Acknowledgements

We thank the infants and their parents for participation in this study; Kazumasa Kimura and Yoko Watanabe (Yakult Central Institute) for providing technical support; Yoshinori Umesaki (Yakult Central Institute), Chihiro Sasakawa and Yoshiyuki Goto (University of Tokyo) for critical reviewing of this manuscript. This research was financially supported by Yakult Central Institute and the National Bioscience Database Center (NBDC) of the Japan Science and Technology Agency (JST).

## Author contributions

T.M. conceived and designed the experiments. K. Yahagi, T.H., S.T. and T.M. performed faecal manipulations. S.T. and T.H. performed DNA sequencing. T.M., H. Mori and S.T. performed microbiome analysis. K. Yahagi and T.H. isolated bifidobacteria and performed HPLC analysis. H. Matsumoto, H. Mori, K. Yamamoto, T.H. and T.M. performed genome analysis. H. Matsumoto generated the FL-SBP knockout strain. T.M., E.O. and H.K. recruited the volunteers. T.Y. and T.M. prepared figure illustrations. T.Y., S.M. and K.K. contributed to the design and execution of this study. T.M. wrote the manuscript. All authors reviewed and approved the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Matsuki, T. *et al.* A key genetic factor for fucosylactose utilization affects infant gut microbiota development *Nat. Commun.* **7**:11939 doi: 10.1038/ncomms11939 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>