 doi:10.20944/preprints202106.0736.v1

Article

Architecture insight of bifidobacterial α-L-fucosidases

José Antonio Curiel*; Ángela Peirotén; José María Landete; Ana Ruiz de la Bastida; Susana Langa; Juan Luis Arqués

Departamento de Tecnología de Alimentos, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA-CSIC), Carretera de La Coruña Km 7.5, 28040, Madrid, Spain; joseantonio.curiel@inia.es (J.A.C.); angela.peiroten@inia.es (A.P.); landete.josem@inia.es (J.M.L.); ana.ruiz@inia.es (A.R.B.); langa.susana@inia.es (S.L.); arques@inia.es (J.L.A.)
* Correspondence: joseantonio.curiel@inia.es; Tel.: +34 91 347 68 86

Abstract: Fucosylated carbohydrates and glycoproteins from human breast milk are essential for the development of the gut microbiota in early life because they are selectively metabolized by bifidobacteria. In this regard, α-L-fucosidases play a key role in this successful bifidobacterial colonization allowing the utilization of these substrates. Although a considerable number of α-L-fucosidases from bifidobacteria have been identified by computational analysis, only a few of them have been characterized. Hitherto, α-L-fucosidases are classified into 3 families, GH29, GH95 and GH151 based on their catalytic structure. However, bifidobacterial α-L-fucosidases belonging to a particular family show significant differences in their sequence. Because this fact could underlie distinct phylogenetic evolves, here extensive similarity searches and comparative analyses of the bifidobacterial α-L-fucosidases identified were carried out with the assistance of previous physicochemical studies available. This work reveals 4 and 2 paralogue bifidobacterial fucosidase groups within GH29 and GH95 families, respectively. Moreover, *Bifidobacterium logum* subsp. *infantis* species exhibited the greatest number of phylogenetic lineages in their fucosidases clustered in every family GH29, GH95 and GH151. Since α-L-fucosidases phylogenetically descended from other glycosyl hydrolase families, we hypothesized that could exhibit additional glycosidase activities other than fucosidase, raising the possibility about their application to transfucosylate other substrates than lactose in order to synthesis novel prebiotics.

Keywords: Bifidobacteria, fucosidases, glycosyl hydrolases, conserved domains, human milk

1. Introduction

The impact of human milk glycobiome on the gut microbiota of infants is well established [1]. While great part of the components of breast milk provide nutrients to the infant, human milk oligosaccharides (HMOs) and human milk glycoproteins (HMGs) selectively favor the colonization and growth of bifidobacteria in the infant intestine, contributing to the development of the gut microbiota [1, 2]. In this regard, *Bifidobacterium* species are considered key actors in the multifaceted process of gut development and maturation of the immune system [3]. In fact, during the first months of birth, the loss of bifidobacterial or the gain of other bacteria can significantly alter the progression of the healthy microbial community with negative consequences for the infant, including a predisposition to autoimmune and/or metabolic diseases such as allergies and childhood obesity [4, 5]. Concerning to that, fucosylated HMOs (FHMOs) and fucosylated HMGs (FHMGs) constitute a great part of the glycobiome of the breast milk [6] (Figure 1) and have been proposed essential in the development of the microbiota [7].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2021 doi:10.20944/preprints202106.0736.v1

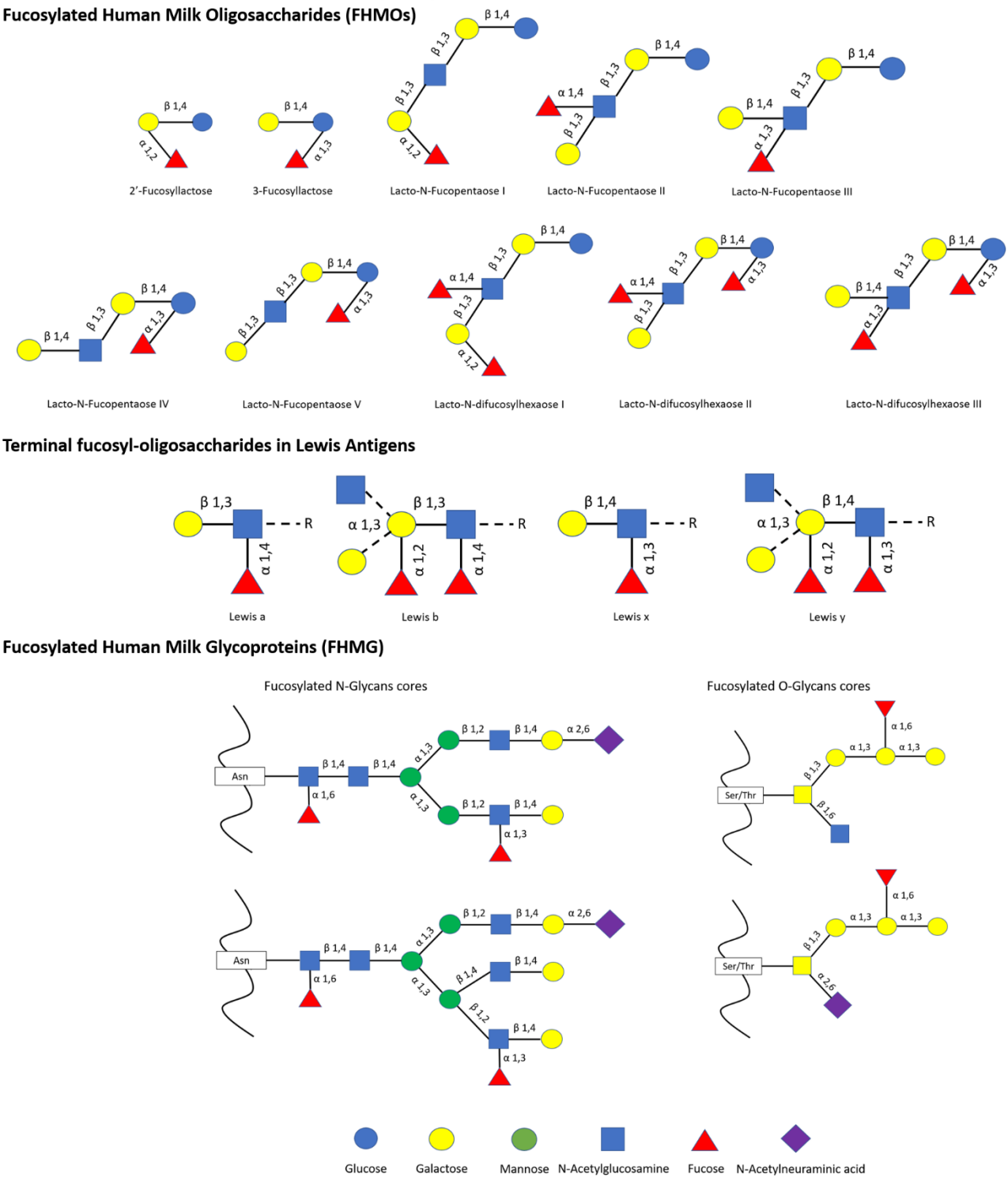

Figure 1. List of main Fucosylated Human Milk Oligosaccharides (FHMOs) and Fucosylated Human Milk Glycoproteins (FHMG) reported in bibliography.

FHMOs constitute the largest fraction of human milk oligosaccharides, and although they showed a small number of different conformations, they can mean up to 70% of the total in an individual mother's milk [6]. The fucosylated trisaccharide 2'-fucosyllactose is the most abundant FHMO, representing from 12 to 45% of the total HMO content in breastmilk, while 3-fucosyllactose is less abundant, from 0.5% to 3% [8]. On the other hand, there are several FMHGs investigated, and contrary to FHMOs, appear

 doi:10.20944/preprints202106.0736.v1

at lower concentration but showing a higher number of different forms including lactoferrin (17%), immunoglobulins IgG (<1%), IgM (<1%), and secretory IgA (11%) [9, 10, 11]. Both FHMOs and FHMGs stand out for their ability to stimulate the growth of bifidobacteria [7, 12], whose metabolism transforms fucosylated oligosaccharides into short-chain fatty acids (SCFAs) such as acetate, formate, lactate, pyruvate [13], which in turn stimulate the immune system by inducing the differentiation of T-regulatory cells via inhibition of histone deacetylase [14].

The great influence of fucosylated compounds present in breast milk on bifidobacteria is due to their ability to metabolize them, being α-L-fucosidases (henceforth fucosidases) indispensable tools that allow shaping the gut microbiome in the first months of life.

According to CAZy database, hitherto, more than 10000 sequences have been identified *in silico* as fucosidases from more than 2000 bacteria species (www.cazy.org). This database classifies fucosidases into 3 families (GH29, GH95 and GH151) according to their catalytic structures. GH29 fucosidases act through a retaining mechanism and have a broader substrate specificity, including hydrolysis of Fuc-α1,3/4/6 linkages [15]. Moreover, Family GH29 fucosidases have been subclassified into two subfamilies. The subfamily A contains α-fucosidases with relatively relaxed substrate specificities, able to hydrolyze *p*-nitrophenyl-α-L-fucopyranoside (pNP-fucose), while the members of subfamily B are specific to α1,3/4-glycosidic linkages and are practically unable to hydrolyze pNP-fucose [16]. Although GH29 fucosidases also could exhibit hydrolysis of Fuc-α1,2 linkages, that activity is mainly attributed to GH95 family, which catalyze the hydrolysis of fucose linkages through an inverting mechanism, resulting in the inversion of the anomeric configuration [17, 18]. Finally GH151 family has poor activity on fucosylated substrates, reason why it is currently questioned whether they are genuine fucosidases [19, 20, 21].

Even though species of the *Bifidobacterium* genus dominate the infant gut microbiota in early life, and given the importance of their metabolism of fucosylated conjugates, there are only few bifidobacterial species studied extensively at both cellular and genomics level for their ability to utilize fucosylated carbohydrates including *B. bifidum* and *Bifidobacterium longum* subsp. *infantis* [22, 23, 24]. However, different strain-dependent metabolic abilities have been unraveled for the use of fucosylated conjugates and likely determined by their fucosidases diversity [25]. Indeed, agreeing to the evolution and phylogenetics of fucosidases previously studied in metazoan fucosidases [26], bifidobacterial fucosidase sequences listed in CAZy reveal substantially *in silico* differences regarding to their conserved domains, even those ones clustered in the same GH, revealing different adaptation/specialization ranges as well as their origin. Therefore, this work pretends to unravel the diverse conserved architectures of bifidobacterial fucosidases and cluster them by activity and phylogenetic evolution in order to propose a novel classification within the GH groups already listed in CAZy.

2. Results

2.1. Bifidobacterial GH29 fucosidases

GH29 fucosidases from bifidobacteria listed in CAZy are shown in Table S1. Based on *in silico* studies concerning to conserved domains released by NCBI Conserved Domains Database (CDD) bifidobacterial GH29 fucosidases could be classified into 4 different phylogenetically groups (Table S1). That differentiation was also confirmed through sequence homology PCA and cluster analyses (Figure 2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2021 doi:10.20944/preprints202106.0736.v1

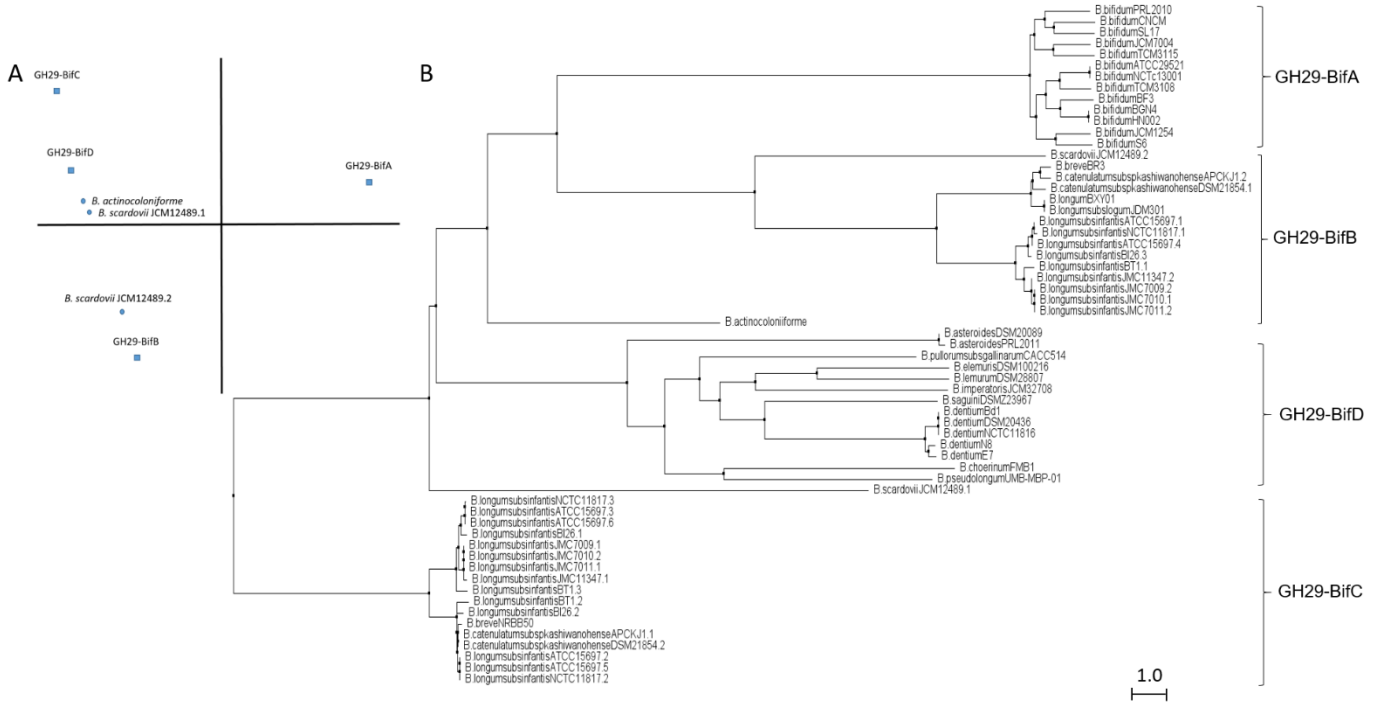

Figure 2. Phylogenetic analysis of bifidobacterial GH29 fucosidases. PCA (A) and cladogram tree (B) distributions of bifidobacterial GH29 fucosidase sequences listed in CAZy released from Jalview 2.11.1.4. software using the neighbor-joining method.

The enzymes included in the proposed GH29-BifA, only found in *B. bifidum* strains, are characterized as large membrane-bound fucosidases (AfuC super family domain; NCBI CDD accession number cl34656) and exhibit an accessory F5/F8 type C domain family (NCBI CDD accession number cl23730) probably involved in recognizing galactose or N-acetyllactosamine [27]. Interestingly, while InterPro database (EMBL-EBI) recognized the F5/F8 type C domain (IPR000421) interpreted the AfuC domain as Glyco_Hydro_29 domain (IPR000933) probably due to the degree of updating of both databases (Table S1). In addition, Ashida et al., 2009 identified a second putative sugar-binding domain in GH29 fucosidase AfcB from *B. bifidum* JCM1254, domain that is frequently found in membrane-bound or cell wall-associated proteins and denominated FIVAR [28]. Those results were here confirmed by SOSUI and HMMTOP databases which allowed the identification of two putative transmembrane helices in GH29-BifA fucosidases (Table S1). Therefore, it has been suggested that both accessory F5/F8 type C and FIVAR domains allow the extracellular character of GH29 fucosidases in *B. bidifum* and could enhance affinity toward fucosyl conjugates [28]. Moreover, in all the N-terminal regions of GH29-BifA fucosidases hydrophobic sequences predicted by SignalP-5.0 to be putative signal peptide with potential cleavage sites were observed (Table S1).

Concerning the AfuC/Glyco_hydro_29 domain, the only representative GH29 fucosidase of GH29-BifA purified and characterized, which is AfcB from *B. bifidum* ATCC 1254, is able to hydrolyze 3-fucosyllactose, Lewis blood group substances (a, b, x, and y types), and lacto-N-fucopentaose II and III. However, the enzyme did not act on glycoconjugates containing α1,2-fucosyl residue or on synthetic pNP-fucose [28].

Supporting the *in silico* characterization of GH29-BifA fucosidases, several studies confirm the ability of *B. bifidum* to hydrolyze extracellularly FHMOs [29]. However, *B. bifidum* appears to prefer the utilization of lactose when growing on FHMO, probably releasing fucose to the environment [29]. This incapacity to consume fucose may be due to the lack of specific transporters. Nevertheless, the extracellular fucosidase activity of *B.*

 doi:10.20944/preprints202106.0736.v1

bifidum could be facilitating the establishment of the bifidobacteria community allowing them to consume the released fucose residues [30].

In contrast to GH29-BifA, the rest of GH29 fucosidases from bifidobacteria do not have neither putative signal peptides nor transmembrane helices and consequently their mode of action can be considered intracellular. Indeed, GH29-BifB fucosidases are characterized by exhibiting an AfuC super family/ Glyco_Hydro_29 domain (NCBI CDD accession number cl34656/ IPR000933) such as GH29-BifA fucosidases but lacking of F5/8 type C and FIVAR domains. Due to the presence of the same fucosidase domain in both groups of fucosidases (GH29-BifA and GH29-BifB), similar metabolic capacities could be affirmed. In fact, the only characterized bifidobacterial GH29-BifB fucosidase (Blon_2336 from *Bifidobacterium longum* subsp. *infantis* ATCC 15697) revealed similarly activity to AfcB from *B. bifidum* ATCC 1254 (GH29-BifA) against Fuc-α1,3 glucosidic, Fuc-α1,3GlcNAc and Fuc-α1,4GlcNAc linkages [21]. These GH29-BifB fucosidases appear to be distributed along strains of different species, contrarily to GH29-BifA fucosidases, and frequently, strains that exhibit GH29-BifB fucosidases, also show GH29-BifC which are duplicated in some of the sequenced strains (Table S1). Actually, the duplication of GH29 fucosidases has been reported previously and plays an important role in fucosidases evolution [31].

GH29-BifC fucosidases are characterized by showing conserved α-Amylase catalytic domain family (NCBI CDD accession number cl38930). It must be taken into account that this superfamily is present in a large number of GHs able to hydrolyze α1,4/6 glycosidic bonds although in turn they have specific domains unlike the GH29-BifC fucosidases of bifidobacteria [32]. However, and since GH29-BifC fucosidases can catalyze the transformation of fucosidic α1-2Gal/3GlcNAc linkages in LNFP I and III, respectively, and mainly Fuc-α1,6 GlcNAc linkages [33], activity non described in the above fucosidase groups, it is difficult to ensure that its catalytic family proposed is α-Amylase catalytic (NCBI CDD) or Glyco_Hydro_29 (InterPro) (Table S1). In this sense, InterPro database (EMBL-EBI) indicated the presence in GH29-BifC fucosidases of a second catalytic family denominated FUC_metazoa_typ (IPR016286) that is close to eukaryotic fucosidases (Table S1). Probably the presence of this domain is key for these fucosidases to be considered as the most unspecific and versatile fucosidases of bifidobacteria since wide range of substrates have been reported for two different GH29-BifC fucosidases from *B. longum* subsp. *infantis* ATCC 15697 [21, 28].

Both GH29-BifB/C fucosidases described in *B. longum* subsp. *infantis* strains are likely found in the cytosol, therefore efficient transport of oligosaccharides is needed unlike *B. bifidum* [13, 21]. In this context, genomic studies carried on *B. longum* subsp. *infantis* ATCC 15697 have unraveled several putative fucose permeases that may facilitate environmental scavenging when soluble fucose is encountered.

In order to elucidate the roles and fitness of bifidobacterial community to shape the gut microbiome and taking into account the relevance of fucosidases in this regard, their features mentioned above should be updated and expanded to avoid ambiguities in the catalytic domains and relate them to their metabolic properties. Certainly, the rest of the enzymes from different bifidobacterial species need to be characterized in order to distinguish reliably the properties of each group of fucosidases for determining the interaction and mode of actions of bifidobacteria during gut colonization. In this sense, the role of GH29-BifD of fucosidases remains unknown despite having been sequenced and identified in certain *Bifidobacterium* species (Table S1). Unlike to GH29-BifC, GH29-BifD fucosidases exhibit specific α-L-fucosidase main domain (NCBI CDD accession number cl38930). Surprisingly, their accession number is matching with superfamily AmyAc family of group II, suggesting a better accurate and updated *in silico* annotation. However, InterPro database (EMBL_EBI) indicates both catalytic domain Glyco_Hydro_29 and FUC_metazoa_typ (InterPro IPR000933 and IPR016286 respectively). Nevertheless, physicochemical properties, substrate specificity confirmation and their correlation with catalytic domains are still pending to be characterized.

 doi:10.20944/preprints202106.0736.v1

2.2. Bifidobacterial GH95 fucosidases

Similarly to GH29 bifidobacterial fucosidases and according to architecture domains, bifidobacterial GH95 fucosidases collected on CAZy also could be subclassified into 2 main groups (Table S2; Figure 3).

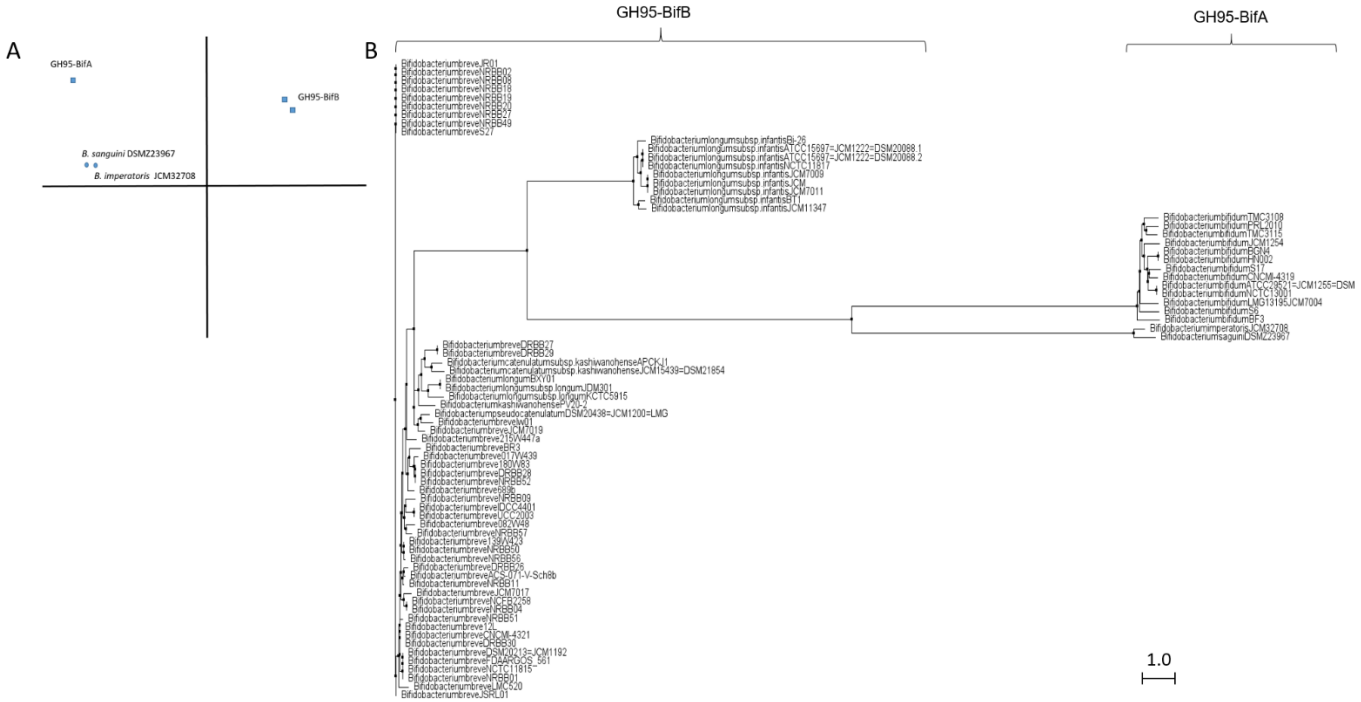

Figure 3. Phylogenetic analysis of bifidobacterial GH95 fucosidases. PCA (A) and cladogram tree (B) distributions of bifidobacterial GH95 fucosidase sequences listed in CAZy released from Jalview 2.11.1.4. software using the neighbor-joining method.

The extracellular character observed in GH29-BifA fucosidases from *B. bifidum* strains is reflected also in their GH95 fucosidases, which are characterized by a putative signal peptide and two predicted transmembrane helices. Among GH95 fucosidases, those features are only found in the proposed GH95-BifA fucosidases from *B. bifidum* with the exception of *Bifidobacterium saguini* DSMZ 23967 fucosidase (Genbank QTB91571.1) which exhibited two putative transmembrane helices (Table S2).

The proposed GH95-BifA was characterized according to the NCBI CDD database by exhibiting Glycosil hydrolase 65 N-terminal (accession number cl22392) as main catalytic domain, while InterPro database analysis (EMBL-EBI) revealed a Glycosyl hydrolase 95 N-terminal (IPR027414) (Table S2). The observed ambiguous prediction on the catalytic architecture could be due to the lack of updating and mismatch annotations. Nevertheless, a common evolutionary origin for GH65 and GH95 families, among others, with conservation of their putative catalytic amino acid residues was noticed and likely influenced the *in silico* results [18]. Nevertheless, and contrarily to GH65 family, the only GH95-BifA representative fucosidase recombinantly produced and characterized (AfcA from *B. bifidum* JCM1254) showed a great activity against Fuc-α1,2 Gal linkages, hydrolyzing mainly 2'-Fucosyllactose and lacto-N-fucopentaose I [17, 34].

On the other hand, while NCBI CDD database detected 2 YjdB overlapping domains (accession number cl35007), whose functions are still uncharacterized but in turn containing Ig-like domain, InterPro database noticed Ig-like_Bact and Bacterial Ig-like group 2 (BIG2) domains instead (accession number IPR022038 and IPR003343 respectively) (Table S2). Despite this coincidence, only the position of one domain practically matches in

 doi:10.20944/preprints202106.0736.v1

both databases (YjdB and BIG2) (Table S2). In addition, Interpro identifies Ig-like_Bact near to N-terminal unlike NCBI CDD and probably GH95-BifA sequences could exhibit up to 3 accessory domains.

It should be noted that, although the function of BIG2 domain has not been unraveled, it have been hypothesized to participate in facilitating the protrusion of the AfcA catalytic GH95 domain from the cell surface to allow its extracellular activity and degrade the fucosyl residues present on glycoconjugates of enterocytes [17]. This fact could lead to define AfcA as a bifidobacterial tool for protecting the host health through modifying α1,2 fucosylated Lewis antigen receptors b and y, recognized by gut pathogens such as *Helicobacter pylori* [35], and Norovirus [36]. Taking into account the conserved domains, GH95 fucosidases from *B. imperatoris* and *B. sanguini* could be close to be clustered within the GH95-BifA (Table S2). The extracellular character of *B. imperatoris* and *B. sanguini* fucosidases could even be affirmed since signal peptides and transmembrane helices are found, although they have not yet been characterized. Indeed, cladogram phylogenetic analysis revealed that both fucosidases actually exhibit more similarities with GH95-BifA (Figure 3).

Beyond GH95-BifA, there are a large number of intracellular GH95 fucosidases from *Bifidobacterium breve* and *B. logum* subsp. *infantis* strains *in silico* categorized by showing a glycosil hydrolase 65 N-terminal domain (cl22392; NCBI CDD). They share the catalytic domain with GH95-BifA without exhibiting accessory BIG2 (Table S2). Nevertheless, Interpro database managed to identify a catalytic domain of greater length than in the GH95-BifA sequences denominated Alpha_L_Fuco family (IPR016518). The presence of this domain could be key for *B. breve* and *B. logum* subsp. *infantis* GH95 fucosidases to show phylogenetic differences with GH95-BifA as shown by the PCA and cladogram analyzes (Figure 3), and therefore clustered in GH95-BifB.

Unfortunately, no *B. breve* GH95-BifB fucosidases have been yet characterized although the described hydrolytic activity of *B. breve* on Fuc-α1,2 Gal linkages supports the presence of a functional GH95 fucosidase [37]. Blon_2335 from *B. longum* subsp. in*fantis* is the only representative of GH95-BifB that has been characterized [21]. In that study, Blon_2335 showed a strong preference for Fuc-α1,2 linkages (2'-FL, LNFP-I), although partially cleaved Fuc-α1,3 linkages (3-FL) unlike to AfcA from *B. bifidum* [21]. Because AfcA structural exploration revealed its catalytic reaction as a α1,2 fucosidase [18], and since both AfcA and Blon_2335 fucosidases show catalytic architecture differences, further studies concerning crystallization of Blon_2335 are needed in order to elucidate its ability for hydrolyzing both Fuc-α1,2 and Fuc-α1,3 linkages. Structure elucidation also could explain the substantial differences between the GH95-BifB fucosidases from *B. breve* and *B. longum* subsp. in*fantis*, also observed in PCA and cladogram (Figure 3), despite presenting the same conserved architecture (Table S2).

2.3. Bifidobacterial GH151 fucosidases

GH151 enzymes form the smallest group of fucosidases (Table S3) and although there are still doubts about their fucosidase activity, *B. longum* subsp. *infantis* ATCC 15697 counts with a GH151 enzyme (Blon_0346) that exhibits probed Fuc-α1,2 Gal activity [21]. Interestingly bifidobacterial GH151 fucosidases are quite divergent from the fucosidases classified in other GH families [21] and all of them belong to *B. longum* subsp. *infantis* species although shown little differences in their sequences (Figure 4). While no signal peptide or transmembrane helices were observed, CDD architecture analyses revealed AmyAc_family superfamily and A4_beta-galactosidase_middle_domain, although some sequences are also identify as containing GanA superfamily domain as well (Table S3).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2021 doi:10.20944/preprints202106.0736.v1

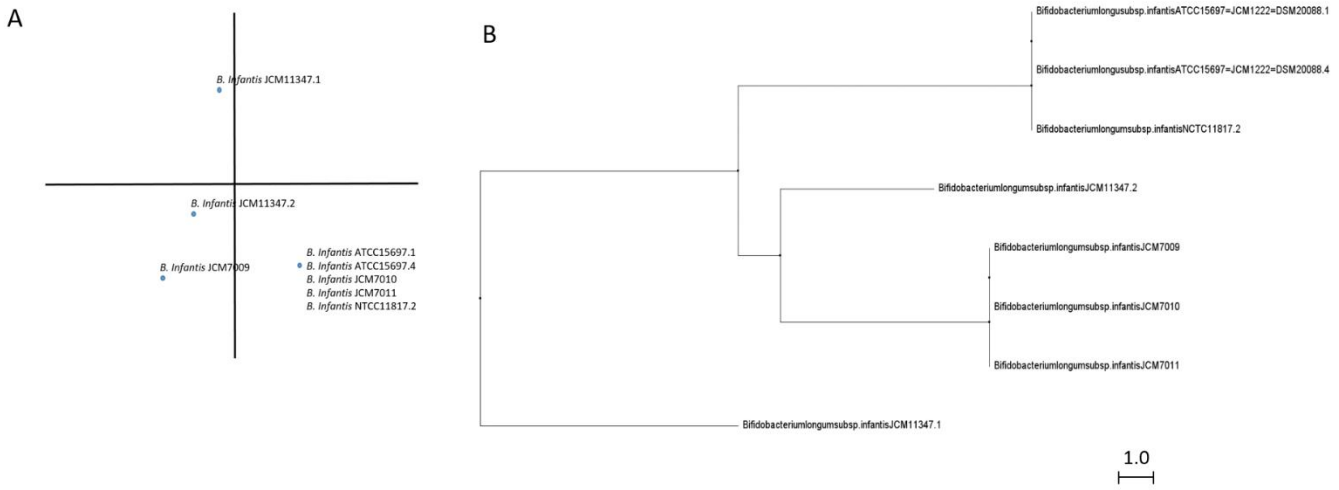

Figure 4. Phylogenetic analysis of bifidobacterial GH151 fucosidases. PCA (A) and cladogram tree (B) distributions of bifidobacterial GH151 complete fucosidase sequences listed in CAZy released from Jalview 2.11.1.4. software using the neighbor-joining method.

GH151 enzymes probably have domains closest to GH29-BifC fucosidases identified by containing conserved AmyAc superfamily domain and likely the ability to hydrolyze α glycosidic linkages [32]. However, because GH151 accessory domains shown (Table S3), they could be considered as potential nonspecific beta galactosidase enzymes with the capacity to hydrolyze Fuc-α1,2 Gal linkages as occurs with Blon_0346. Nevertheless, further studies in order to elucidate their subjacent activity, substrate specificity and conformational structure are needed to understand their role in the hydrolysis of fucosylated carbohydrates.

3. Discussion

Breast milk, beyond its nutritional function, provides the necessary pillars for the initial establishment of the gut microbiota in newborns. In this regard, FHMOs and FHMGs stand out for their ability to stimulate the growth of bifidobacteria [8, 12], which in turn produce SCFAs acetate, formate, lactate, pyruvate [13], stimulating the immune system [14], and serving as an energy source for colonocytes [38].

Although only a few bifidobacterial species have been studied extensively at both cellular and genomics level for their ability to utilize fucosylated carbohydrates such as *B. bifidum* and *B. longum* subsp. *infantis* [22, 23, 24], their success in colonizing the gut is due to the different strain-dependent metabolic abilities developed for the use of both FHMOs and FHMGs [25]. Therefore, fucosidases play a key role in the bifidobacterial gut establishment. Concerning to that, *B. bifidum* strains show two extracellular fucosidases belonging to GH29 and GH95 families. Both fucosidases cover the hydrolysis of Fuc-α1,3Glu; Fuc-α1,3/4GlcNAc and Fuc-α1,2 Gal linkages [17, 18, 28, 34]. Since *B. bifidum* prefer the utilization of lactose [29], 2'-fucosyllactose could be its target substrates releasing to the environment lactose and fucose, which is also liberated from blood Lewis a, b, x, and y antigens [28]. For all the above, *B. bifidum* fucosidases could be considered altruist and essential for microbial gut establishment through promoting bifidobacterial mutualism and carbohydrate syntrophy in the infant gut [39]. Given that bifidobacteria are able to metabolize lactose and species such as *B. longum* subsp. *infantis* or *B. breve* can metabolize fucose, their growth is improved under the presence of fucosidases from *B. bifidum*. Thus, Gotoh et al. (2018) suggested that extracellular fucosidases

 doi:10.20944/preprints202106.0736.v1

from *B. bifidum* could be crucial during the development of a bifidobacteria-rich microbiota in the breast-fed infant gut, by providing fucosylated conjugates degradants [34]. On the other hand, *B. bifidum* fucosidases contribute to the protection of host through the modification of lewis antigens [28].

Regarding the catalytic domains of the *B. bifidum* fucosidases, it should be noted that GH29-BifA present orthologous fucosidases in other bifidobacterial species clustered in GH29-BifB/D, and they probably all have a common phylogenetic lineage (Figure 2). However, this statement has only been functionally corroborated through the characterization of the enzymes AfcB (GH29-BifA) and Blon_2336 (GH29-BifB) due to lack of results of GH29-BifD fucosidases.

Conversely, GH95-BifA fucosidases as well as those grouped in GH95-BifB, and according to CDD database observations (Table S2), could phylogenetically descend from either an evolutionary specialization or unspecification of glycosidases clustered in GH65. Indeed, this *in silico* observation agrees with the crystallization results obtained of the structure AcFA from *B. bifidum* [18]. According to that, both GH65 and GH95 enzymes share an α/α 6 barrel fold with inverting mechanism and glutamate^{566} as catalytic proton donor. Moreover, Nagae et al. (2007) compared the structures between families GH65 and GH95 revealing conservation of the general acid residues, but catalytic acid / base aspartate^{766}, which is shifted in AfcA [18]. That shifting was also found in the rest of bifidobacterial GH95 fucosidases (data not shown) and agreeing to above mentioned authors, reaction mechanism of bifidobacterial GH95 fucosidases differ from those of the GH65 family [18].

The other species widely studied for its fucosidase activity is *B. longum* subsp. *infantis*. Actually, it is the only species of bifidobacteria that exhibits GH29, GH95 and GH151 fucosidases that have been recombinantly purified and characterized [21]. Those fucosidases allow *B. longum* subsp. *infantis* to use a wide range of substrates, hydrolyzing Fuc-α1,3Glu; Fu-cα1,2/3Gal; Fuc-α1,3/4/6GlcNAc linkages [21, 33]. As previously commented, *B. longum* subsp. *infantis* GH29-BifB fucosidases are orthologous with those classified in GH29-BifA. However, this species also shows GH29 duplicated fucosidases, clustered in the GH29-BifC, with different architecture and paralogs from those of GH29-BifB (Figure 3). Taking in to account the fucosidase duplication and in agreement with You et al. (2019), *B. longum* subsp. *infantis* GH29-BifC fucosidases could have evolved from a different glycosyl hydrolase [31]. According to CDD database observations (Table S1) and because their predicted structure is composed by an β/α 6 barrel fold with retaining mechanism and glutamate as catalytic proton donor, GH29-BifC fucosidases from *B. longum* subsp. *infantis* could descend from GH13 glycosidases (α-amylases).

GH29-BifC fucosidases, similar to GH95-BifB, which is probably phylogenetically originated from GH65 family as described above, need to have further explored their structural crystallization in order to elucidate their origins and evolution pathway. In addition, GH29-BifC fucosidases show similarities with metazoan fucosidases according to InterPro database (Table S1) including aspartate^{224} and glutamate^{270} residues (data not shown), which play the role of the catalytic nucleophile and catalytic acid / base respectively in metazoan fucosidases [26].

Finally, GH151 fucosidases are exclusively present in *B. longum* subsp. *infantis*. This fact could suggests a fourth pathway of fucosidases phylogenetic evolution in that species closely related to GH29-BifC fucosidases since they present a N-terminal α amylase catalytic domain. In addition, Blon_0346 was originally classified as a member of GH29 family due to their fucosidase activity despite low similarity [21]. However, GH151 enzymes may be the result of a branch in the evolution of GH29-BifC fucosidases since they show

 doi:10.20944/preprints202106.0736.v1

a GH42 beta galactosidase trimerization architecture instead of conserved features of metazoan fucosidases.

4. Conclusions

This is the first study that explores phylogenetically the three families of the bifidobacterial fucosidases GH29, GH95 and GH151 through their conserved architecture, showing that *B. bifidum* and *B. longum* subsp. *infantis* reveal 2 and 4 different phylogenetic lineages respectively, belonging to different fucosidases families. On the other hand, given the differences in the catalytic architecture observed in this work, the bifidobacterial fucosidases belonging to the GH29 and GH95 families could be subclassified into 4 and 2 groups respectively.

Taking into account that the observations described in this work were obtained *in silico* and supported by current characterization results from some *B. bifidum* and *B. longum* subsp. *infantis* fucosidases, further studies regarding structural characterization and physicochemical properties of more of fucosidases identified by computational analysis are needed in order to validate the novel classification of bifidobacterial fucosidases here proposed.

Concerning to *B. longum* subsp. *infantis* fucosidases evolved from different GH families as GH29-BifC, GH95-BifB and GH151 and given that their conserved architecture presents vestiges of ancestral glycosidases GH13, GH65 and GH42 respectively, as well as *B. Bifidum* GH95-BifA fucosidases phylogenetically descendant of GH65, deepening substrate spectrum analyses could determine their underlying roles in this species. In this context, and since some fucosidases have been used to transfucosylate carbohydrates or glycoconjugates, the application of these evolved and hypothetically nonspecific *B. longum* subsp. *infantis* fucosidases mentioned above can open a new perspective towards the synthesis of novel fucosylated conjugates by using different substrates beyond the lactose for synthetizing 2'-fucosyllactose. This vision is oriented towards the supply those novel fucosylated conjugates to adults in combination with fucosidase producer bifidobacteria in order to maintain a healthy microbiota or to reestablish it from dysbiosis states as described previously [40, 41]. In this regard, it would be important to elucidate phylogenetically as well as structurally and physicochemically the fucosidases of many other gut microorganism genus as for instance *Lactobacillus, Bacteroides,* and *Akkermansia* with the aim to reveal the whole gut fucosidases interaction.

5. Materials and Methods

5.1. Identification and selection of fucosidase sequences

Complete bifidobacterial fucosidase protein sequences belonging to GH29, GH95 and GH151 families were retrieved from CAZy database [19]. Fucosidase sequences were used as probes in PSI-BLAST searches [42] against the NCBI [43], Swiss-Prot [44] and Ensembl [45] protein databases.

5.2. Protein sequence, alignment and phylogenetic analysis of α-L-fucosidases

Fucosidase sequences were analyzed using SignalP-5.0 [46], with default options to predict signal peptide sequences; SOSUI [47] and HMMTOP [48] with default parameters for the prediction of transmembrane helices. NCBI Conserved Domains Database (CDD) [49] and InterPro databases (EMBL_EBI) [50] were used to predict the domain architecture. Inferred fucosidase amino acid sequences were aligned using Clustal Omega web

version [51]. All sequences belonging to the same GH families were considered in phylogenetic analyses. Neighbor-joining method cladogram and PCA analyses were performed using the program Jalview 2.11.1.4. [52].

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, **Table S1:** Architecture and main features of bifidobacterial GH29 fucosidase sequences. **Table S2:** Architecture and main features of bifidobacterial GH95 fucosidase sequences. **Table S3:** Architecture and main features of bifidobacterial GH151 fucosidase sequences.

Abbreviations: FHMG, Fucosylated human milk glycoprotein; FHMO, Fucosylated human milk oligosaccharide; Fuc, Fucose; Gal, Galactose; GH Glycosyl hydrolase; GlcNAc, N-acetylglucosamine; Glu, Glucose; HMG, human milk glycoprotein; HMO, human milk oligosaccharide; LNFP, Lacto-N-Fucopentaose; pNP-fucose, *p*-nitrophenyl-α-L-fucopyranoside; SCFAs, Short-chain fatty acids.

Author Contributions: Conceptualization J.A.C.; methodology J.A.C. and J.M.L.; validation and formal analysis J.A.C. and A.P.; investigation J.A.C, A.P., J.M.L., S.L., and A.R.B.; writing-original draft preparation J.A.C and J.M.L.; writing-review and editing J.A.C, and A.P.; visualization and supervision J.A.C. and J.L.A.; project administration, J.A.C.; funding acquisition, J.A.C. All authors have read and agreed to the published version of the manuscript

Funding: This research was funded by Spanish Ministry of Science and Innovation - Ramón y Cajal program, grant number RYC2019-026368-I.

Acknowledgments: J.A.C has a postdoctoral contract with the research program "Ramón y Cajal" (RYC2019-026368-I) and A.R.B is recipient of a Predoctoral contract (PRE2018-086293), both from Spanish Ministry of Science and Innovation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zivkovic, A.M.; German, J.B.; Lebrilla, C.B.; Mills, D.A. Human milk glycobiome and its impact on the infant gastrointestinal microbiota. *PNAS* **2011**, 108, 4653-4658.
2. Milani, C.; Mancabelli, L.; Lugli, G.A.; Duranti, S.; Turroni, F.; Ferrario, C.; Mangifesta, M.; Viappiani A.; Ferreti, P.; Corfer, V.; Tett, A.; Segata, N.; van Sinderen, D.; Ventura, M. Exploring vertical transmission of bifidobacteria from mother to child. *Appl. Environ. Microbiol*. **2015**, 81, 7078-7087.
3. Turroni, F.; Peano, C.; Pass, D.A.; Foroni, E.; Severgnini, M.; Claesson, M.J.; Kerr, C.; Hourihane, J.; Murray, D.; Fuligni, F.; Gueimonde, M.; Margolles, A.; De Bellis, G.; O'Toole, P.W.; Van Sinderen, D.; Marchesi, J.R.; Ventura, M. Diversity of bifidobacteria within the infant gut microbiota. *PLoS One* **2012**, 7,e36957.
4. O'Neill, I.; Schofield, Z; Hall, L.J. Exploring the role of the microbiota member *Bifidobacterium* in modulating immune-linked diseases. *Emerg. Top. Life Sci.* **2017**, 1, 333-349.
5. Wampach. L; Heintz-Buschart, A.; Fritz, J.V.; Ramiro-Garcia, J.; Habier, J.; Herold, M.; Nayaranasamy, S.; Kaysen, A.; Hogan, A.H.; Bindl, L.; Bottu, J.; Halder, R.; Sjöqvist, C.; May, P.; Andersson, A.F.; de Beaufort, C.; Wilmes, P. Birth mode determines earliest strainconferred gut microbiome functions and immunostimulatory potential. *Nat. Commun*. **2018**, 9, 1-14.
6. Wu, S.; Tao, N.; German, J.B.; Grimm, R.; Lebrilla, C.B. Development of an annotated library of neutral human milk oligosaccharides. *J. Proteome Res*. **2010**, 9, 4138-4151.
7. Bai, Y.; Tao, J.; Zhou, J.; Fan, Q.; Liu, M.; Hu, Y.; Xu, Y.; Zhang, L.; Yuan, J.; Li, W.; Ze, X.; Malard, P.; Guo, Z.; Uan. J.; Li, M. Fucosylated human milk oligosaccharides and N-glycans in the milk of Chinese mothers regulate the gut microbiome of their breast-fed infants during different lactations stages. *MSystems* **2018**, 3, e00206-18.
8. McGuire, M.K.; Meehan, C.L.; McGuire, M.A.; Williams, J.E.; Foster, J.; Sellen, D.W.; Prentice, A.M. What's normal? Oligosaccharide concentrations and profiles in milk produced by healthy women vary geographically. *Am. J. Clin. Nutr*. **2017**, 105, 1086-1100.
9. Huang, J.; Guerrero, A.; Parker, E.; Strum, J.S.; Smilowitz, J.T.; German, J.B.; Lebrilla, C.B. Site-specific glycosylation of secretory immunoglobulin A from human colostrum. *J. Proteome Res*. **2015**, 14, 1335-1349.
10. Li, W.; Yu, R.; Ma, B.; Yang, Y.; Jiao, X.; Liu, Y.; Cao, H.; Dong, W.; Liu, L.; Ma, K.; Fukuda, T.; Liu, Q.; Ma, T.; Wang, Z.; Gu, J.; Zhang, J.; Taniguchi, N. Core fucosylation of IgG B cell receptor is required for antigen recognition and antibody production. *J. Immunol.* **2015**, 194, 2596-2606.
11. Peterson, R.; Cheah, W.Y.; Grinyer, J.; Packer, N. Glycoconjugates in human milk: protecting infants from disease. *Glycobiology* **2013**, 23, 1425-1438.

12. Korpela, K.; Salonen, A.; Hickman, B.; Kunz, C.; Sprenger, N.; Kukkonen, K.; Savilahti, E.; Kuitunen, M.; de Vos, W.M. Fucosylated oligosaccharides in mother's milk alleviate the effects of caesarean birth on infant gut microbiota. *Sci. Rep*. **2018**, 8, 1-7.
13. Zabel, B.E.; Gerdes, S.; Evans, K.C.; Nedveck, D.; Singles, S.K.; Volk, B.; Budinoff, C. Strain-specific strategies of 2′-fucosyllactose, 3-fucosyllactose, and difucosyllactose assimilation by *Bifidobacterium longum* subsp. *infantis* Bi-26 and ATCC 15697. *Sci. Rep*. **2020**, 10, 1-18.
14. Park, J.; Kim, M.; Kang, S.G.; Jannasch, A.H.; Cooper, B.; Patterson, J.; Kim, C.H. Short-chain fatty acids induce both effector and regulatory T cells by suppression of histone deacetylases and regulation of the mTOR–S6K pathway. *Mucosal immunol*. **2015**, 8, 80-93.
15. Grootaert, H.; Van Landuyt, L.; Hulpiau, P.; Callewaert, N. Functional exploration of the GH29 fucosidase family. *Glycobiology* **2020**, 30, 735-745.
16. Shaikh, F.A.; Van Bueren, A.L.; Davies, G.J.; Withers, S.G. Identifying the catalytic acid/base in GH29 α-L-Fucosidase subfamilies. *Biochemistry* **2013**, 52, 5857-5864.
17. Katayama, T.; Sakuma, A.; Kimura, T.; Makimura, Y.; Hiratake, J.; Sakata, K.; Yamanoi, T.; Kumagai, H.; Yamamoto, K. Molecular cloning and characterization of *Bifidobacterium bifidum* 1, 2-α-L-fucosidase (AfcA), a novel inverting glycosidase (glycoside hydrolase family 95). *J. Bacteriol*. **2004**, 186, 4885-4893.
18. Nagae, M.; Tsuchiya, A.; Katayama, T.; Yamamoto, K.; Wakatsuki, S.; Kato, R. Structural basis of the catalytic reaction mechanism of novel 1, 2-α-L-fucosidase from *Bifidobacterium bifidum. J. Bio. Chem*. **2007**, 282, 18497-18509.
19. Lombard, V.; Ramulu, H.G.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. **2014**, 42, 490-495.
20. Lezyk, M.; Jers, C.; Kjaerulff, L.; Gotfredsen, C.H.; Mikkelsen, M.D.; Mikkelsen, J.D. Novel α-L-fucosidases from a soil metagenome for production of fucosylated human milk oligosaccharides. *PLoS One* **2016**, 11, 1-18.
21. Sela, D.A.; Garrido, D.; Lerno, L.; Wu, S.; Tan, K.; Eom, H.J.; Joachimiak, A.; Lebrilla, C.B.; Mills, D.A. *Bifidobacterium longum* subsp. *infantis* ATCC 15697 α-fucosidases are active on fucosylated human milk oligosaccharides. *Appl. Environ. Microbiol*. **2012**, 78, 795-803.
22. Angeloni, S.; Ridet, J.L.; Kusy, N.; Gao, H.; Crevoisier, F.; Guinchard, S.; Kochhar, S.; Sigrist, H.; Sprenger, N. Glycoprofiling with micro-arrays of glycoconjugates and lectins. *Glycobiology* **2005**, 15, 31-41.
23. Turroni, F.; Peano, C.; Pass, D.A.; Foroni, E.; Severgnini, M.; Claesson, M.J.; Kerr, C.; Hourihane, J.; Murray, D.; Fuligni, F.; Gueimonde, M.; Margolles, A.; de Bellis, G.; O'Toole P.W.; van Sinderen, D.; Marchesi, J.R.; Ventura, M. (2012). Diversity of bifidobacteria within the infant gut microbiota. *PloS one* **2012**, 7, e36957.
24. Bergström, A.; Skov, T.H.; Bahl, M.I.; Roager, H.M.; Christensen, L.B.; Ejlerskov, K.T.; Molgaard, C.; Michaelsen, K.F.; Licht, T.R. Establishment of intestinal microbiota during early life: a longitudinal, explorative study of a large cohort of Danish infants. *Appl. Environ. Microbiol*. **2014**, 80, 2889-2900.
25. Turroni, F.; Bottacini, F.; Foroni, E.; Mulder, I.; Kim, J.H.; Zomer, A.; Sánchez, B.; Bidossi, A.; Ferrarini, A.; Giubellini, V.; Delledonne, M.; Henrissat, B.; Coutinho, M.; Oggioni, M.; Fitzgerald, F.G.; Mills, D.; Margolles, A.; Kelly, D.; van Sinderen, D.; Ventura, M. Genome analysis of *Bifidobacterium bifidum* PRL2010 reveals metabolic pathways for host-derived glycan foraging. *PNAS* **2010**, 107, 19514-19519.
26. Intra, J.; Perotti, M.E.; Pavesi, G.; Horner, D. Comparative and phylogenetic analysis of α-l-fucosidase genes. *Gene* **2007**, 392, 34-46.
27. Abbott, D.W.; Eirin-Lopez, J.M.; Boraston, A.B. Insight into ligand diversity and novel biological roles for family 32 carbohydrate-binding modules. *Mol. Biol. Evol*. **2008**, 25, 155-167.
28. Ashida, H.; Miyake, A.; Kiyohara, M.; Wada, J.; Yoshida, E.; Kumagai, H.; Katayama, T.; Yamamoto, K. Two distinct α-L-fucosidases from *Bifidobacterium bifidum* are essential for the utilization of fucosylated milk oligosaccharides and glycoconjugates. *Glycobiology* **2009**, 19, 1010-1017.
29. Garrido, D.; Ruiz-Moyano, S.; Lemay, D.G.; Sela, D.A.; German, J.B.; Mills, D.A. Comparative transcriptomics reveals key differences in the response to milk oligosaccharides of infant gut-associated bifidobacteria. *Sci. Rep*. **2015**, 5, 1-18.
30. Asakuma, S.; Hatakeyama, E.; Urashima, T.; Yoshida, E.; Katayama, T.; Yamamoto, K.; Kumagai, H.; Ashida, H.; Hirose, J.; Kitaoka, M. Physiology of consumption of human milk oligosaccharides by infant gut-associated bifidobacteria. *J. Biol. Chem*. **2011**, 286, 34583-34592.
31. You, J.; Lin, S.; Jiang, T. Origins and evolution of the α-L-fucosidases: from bacteria to metazoans. *Front. Microbiol*. **2019**, 10, 1756.
32. Janeček, Š.; Svensson, B.; MacGregor, E.A. α-Amylase: an enzyme specificity found in various families of glycoside hydrolases. *Cell. Mol. Life Sci*. **2014**, 71, 1149-1170.
33. Ashida, H.; Fujimoto, T.; Kurihara, S.; Nakamura, M.; Komeno, M.; Huang, Y.; Katayama, T.; Kinoshita, T.; Takegawa, K. 1, 6-α-L-Fucosidases from *Bifidobacterium longum* subsp. *infantis* ATCC 15697 involved in the degradation of core-fucosylated N-glycan. *J. Appl. Glycosci*. **2020**, 67, 23-29.
34. Gotoh, A.; Katoh, T.; Sakanaka, M.; Ling, Y.; Yamada, C.; Asakuma, S.; Urashima, T.; Tomabechi, Y.; Katayama-Ikegami, A.; Kurihara, S.; Yamamoto, K.; Harata, G.; He, F.; Hirose, J.; Kitaoka, M.; Okuda, S.; Katayama, T. Sharing of human milk oligosaccharides degradants within Bifidobacterial communities in faecal cultures supplemented with *Bifidobacterium bifidum*. *Sci. Rep*. **2018**, 8, 1-14.
35. Moran, A.P. Relevance of fucosylation and Lewis antigen expression in the bacterial gastroduodenal pathogen *Helicobacter pylori*. *Carbohydr. Res*. **2008**, 343, 1952-1965.

36. Huang, P.; Farkas, T.; Marionneau, S.; Zhong, W.; Ruvoën-Clouet, N.; Morrow, A.L.; Altaye, M.; Pickering, L.K.; Newburg, D.S.; LePendu, J.; Jiang, X. Noroviruses bind to human ABO, Lewis, and secretor histo-blood group antigens: identification of 4 distinct strain-specific patterns. *J. Infect. Dis*. **2003**, 188, 19-31.
37. Sakanaka, M.; Gotoh, A.; Yoshida, K.; Odamaki, T.; Koguchi, H.; Xiao, J. Z.; Kitaoka, M.; Katayama, T. Varied pathways of infant gut-associated *Bifidobacterium* to assimilate human milk oligosaccharides: Prevalence of the gene set and its correlation with bifidobacteria-rich microbiota formation. *Nutrients* **2020**, 12, 71.
38. Corrêa-Oliveira, R.; Fachi, J.L.; Vieira, A.; Sato, F.T.; Vinolo, M.A.R. Regulation of immune cell function by short-chain fatty acids. *Clin. Transl. Immunol*. **2016**, 5, e73.
39. Motherway, M.O.C.; O'Brien, F.; O'Driscoll, T.; Casey, P.G.; Shanahan, F.; van Sinderen, D. Carbohydrate syntrophy enhances the establishment of *Bifidobacterium breve* UCC2003 in the neonatal gut. *Sci. Rep*. **2018**, 8, 1-10.
40. Ryan, J.J.; Monteagudo-Mera, A.; Contractor, N.; Gibson, G.R. Impact of 2′-Fucosyllactose on Gut Microbiota Composition in Adults with Chronic Gastrointestinal Conditions: Batch Culture Fermentation Model and Pilot Clinical Trial Findings. *Nutrients* **2021**, 13, 938.
41. Elison, E.; Vigsnaes, L.K.; Krogsgaard, L.R.; Rasmussen, J.; Sørensen, N.; McConnell, B.; Hennet, T.; Sommer, M.O.A.; Bytzer, P. Oral supplementation of healthy adults with 2′-O-fucosyllactose and lacto-N-neotetraose is well tolerated and shifts the intestinal microbiota. *Brit. J. Nutr*. **2016**, 116, 1356-1368.
42. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **1997**, 25, 3389-3402.
43. Wheeler, D. L.; Barrett, T.; Benson, D.A.; Bryant, S.H.; Canese, K.; Chetvernin, V.; Church, D.M.; DiCuccio, M.; Edgar, R.; Federhen, S.; Feolo, M.; Geer, L.Y.; Helmberg, W.; Kapustin, Y.; Khovayko, O.; Landsman, D.; Lipman, D.J.; Madden, T.L.; Maglott, D.R.; Miller, V.; Ostell, J.; Pruitt, K.D.; Schuler, G.D.; Shumway, M.; Sequeira, E.; Sherry, S.T.; Sirotkin, K.; Souvorov, A.; Starchenko, G.; Tatusov, R.L.; Tatusova, T.A.; Wagner, L.; Yaschenko, E. Database resources of the national center for biotechnology information. *Nucleic Acids Res*. **2007**, 36, 13-21.
44. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.;Bairoch, A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. **2003**, 31, 3784-3788.
45. Yates, A.D.; Achuthan, P.; Akanni, W.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Azov, A.G.; Bennett, R.; Bhai, J.; Billis, K.; Boddu, S.; Marugán, J.C.; Cummins, C.; Davidson, C.; Dodiya, K.; Fatima, R.; Gall, A.; Garcia Giron C.; Gil, L.; Grego, T.; Haggerty, L.; Haskell, E.; Hourlier, T.; Izuogu, O.G.; Janacek, S.H.; Juettemann, T.; Kay, M.; Lavidas, I.; Le T.; Lemos, D.; Gonzalez Martinez, J.; Maurel, T.; McDowall, M.; McMahon, A.; Mohanan, S.; Moore, B.; Nuhn, M.; Cheh, D.N.; Parker, A.; Parton, A.; Patricio, M.; Sakthivel, M.P.; Abdul Salam, A.I.; Schmitt, B.M.; Schuilenburg, H.; Sheppard, D.; Sycheva, M.; Szuba, M.; Taylor, K.; Thormann, A.; Threadgold, G.; Vullo, A.; Walts, B.; Winterbottom, A.; Zadissa, A.; Chakiachvili, M.; Flint, B.; Frankish, A.; Hunt, S.E.; IIsley, G.; Kostadima, M.; Langridge, N.; Loveland, J.E.; Martin, F.G.; Morales, J.; Mudge, J.M.; Muffato, M.; Perry, E.; Ruffier, M.; Trevanion, S.J.; Cunningham, F.; Howe, K.L.; Zerbino, D.R; Flicek, P. Ensembl 2020. *Nucleic Acids Res*. **2020**, 48, 682-688.
46. Almagro Armenteros, J.J.; Tsirigos, K.D.; Sønderby, C.K.; Petersen, T.N.; Winther, O.; Brunak, S.; von Heijne G.; Nielsen, H. *Nat. Biotechnol*. **2019**, 37, 420-423.
47. Hirokawa, T.; Boon-Chieng, S.; Mitaku, S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **1998**, 14, 378-379.
48. Tusnady, G.E.; Simon, I. Topology of membrane proteins. *J. Chem. Inf. Comput. Sci*. **2001**, 41, 364-368.
49. Lu, S.; Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Marchler, G.H.; Song, J.S.; Thanki, N.; Yamashita, R.A.; Yang, M.; Zhang, D.; Zheng, C.; Lanczycki, C.J.; Marchler-Bauer, A. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. **2020**, 48, 265-268.
50. Blum, M.; Chang, H.Y.; Chuguransky, S.; Grego, T.; Kandasaamy, S.; Mitchell, A.; Nuka, G.; Paysan-Lafosse, T.; Qureshi, M.; Raj, S.; Richardson, L.; Salazar, G.A.; Williams, L.; Bork, P.; Bridge, A.; Gough, J.; Haft, D.H.; Letunic, I.; Marchler-Bauer, A.; Mi, H.; Natale, D.A.; Necci, M.; Orengo, C.A.; Pandurangan, A.P.; Rivoire, C.; Sigrist, C.J.A.; Sillitoe, I.; Thanki, N.; Thomas, P.D.; Tosatto, S.C.E.; Wu, C.H.; Bateman, A.; Finn, R. D. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res*. **2021**, 49, 344-354.
51. McWilliam, H.; Li, W.; Uludag, M.; Squizzato, S.; Park, Y.M.; Buso, N.; Cowley, A.P.; Lopez, R. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res*. **2013** 41, 597-600.
52. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview version 2-A Multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**; 25, 1189-1191.