

Supplementary Material

Supplementary Table 1. Proteins harboring FN3 domains that are located outside the PFNA operon in human gut-derived strains of different species of bifidobacteria.

| Name of the protein harboring FN3 domains | Species of <i>Bifidobacterium</i> | Protein sequence identifier in NCBI | Location of FN3 domain |
|-------------------------------------------|-----------------------------------------|-------------------------------------|------------------------|
| glycosyl hydrolase family 3 | <i>B. longum</i> subsp. <i>longum</i> | AIW44861 | C-terminus, 70 aa |
| | | AIW44860 | C-terminus, 71 aa |
| glycosyl hydrolase family 3 | <i>B. longum</i> subsp. <i>infantis</i> | EEQ54738 | Central region, 75 aa |
| | | EEQ55439 | C-terminus, 71 aa |
| glycosyl hydrolase family 3 | <i>B. adolescentis</i> | BAF40392 | Central region, 75 aa |
| | | BAF40391 | Central region, 79 aa |
| glycosyl hydrolase family 3 | <i>B. animalis</i> subsp. <i>lactis</i> | OQM47627 | C-terminus, 71 aa |
| glycosyl hydrolase family 3 | <i>B. bifidum</i> ** | SPU27118 | C-terminus, 71 aa |
| glycosyl hydrolase family 3 | <i>B. breve</i> *** | OQM47627 | C-terminus, 71 aa |
| glycosyl hydrolase family 3 | <i>B. angulatum</i> | AMK57051 | C-terminus, 70 aa |
| | | AMK57045 | C-terminus, 71 aa |
| glycosyl hydrolase family 3 | <i>B. dentium</i> | ADB10806* | C-terminus, 71 aa |
| glycosyl hydrolase family 3 | <i>B. pseudocatenulatum</i> | EEG71159 | C-terminus, 71 aa |

*Apart from the protein ADB10806, there was another 11 proteins belonging to the GH family 3 in the genomes of all *B. dentium* strains: ADB08901, ADB10769, ADB10268, ADB10809, ADB10807, ADB09184, ADB10263, ADB10249, ADB08889, ADB10708, ADB10313.

**Most *B. bifidum* strains do not contain proteins harboring FN3 domains outside the PFNA operon; a single protein of the glycosyl hydrolase family 3 was found in the genome of the strain NCTC10471.

***In *B. breve*, proteins of the glycosyl hydrolase family 3 harboring FN3 domains were present mainly in strains isolated from the gut microbiota of infants.

Supplementary Table 2. Proteins harboring FN3 domains in human gut-derived bacteria of the families *Lactobacillaceae*, *Bacteroidaceae* and *Clostridioides*.

| Name of the protein harboring FN3 domains | Species | Protein sequence identifier in NCBI | FN3 domain Location of FN3 domain | Cytokine receptor motif |
|-------------------------------------------|----------------------------------|-------------------------------------|-----------------------------------|-------------------------|
| glycosyl hydrolase family 31 | <i>Lactobacillus acidophilus</i> | KRK27797 | C-terminus, 78 aa | |
| | <i>Lactobacillus crispatus</i> | QWW29158 | C-terminus, 78 aa | |
| | <i>Lactobacillus gasseri</i> | ABJ59525 | C-terminus, 68 aa | |
| | <i>Lactobacillus helveticus</i> | GFP12355 | C-terminus, 78 aa | VSXWS |
| | <i>Lactobacillus paragasseri</i> | EFJ69427 | C-terminus, 68 aa | |
| | <i>Lactobacillus ultunensis</i> | QQP28688 | C-terminus, 78 aa | |
| | <i>Bacteroides fragilis</i> | AKA51120 | C-terminus, 73 aa | YSXWA |
| glycosyl hydrolase family 43 | <i>Bacteroides fragilis</i> | KXU49890 | C-terminus, 86 aa | VSXPS |
| | <i>Bacteroides ovatus</i> | QUT79180 | C-terminus, 85 aa | RSXLS |
| | <i>Bacteroides uniformis</i> | KAB4109992 | C-terminus, 87 aa | VSXRS |
| glycosyl hydrolase family 18 | <i>Clostridium perfringens</i> | KAB4109856 | C-terminus, 80 aa | ENXRS |
| | | AOY53222 | C-terminus, 85 aa | EIXKR |
| glycosyl hydrolase family 20b | | AOY53993 | C-terminus, 85 aa | EVXSK |
| | | AOY54134 | C-terminus, 85 aa | EIXSK |
| alpha-N- | | AOY53263 | C-terminus, 85 aa | ELXSK |

| | | | | |
|--------------------------------------------------------------------------------------|-----------------------------------------|------------|---------------------------------|---------------------|
| acetylglucosaminidase | | | | |
| sialidase | | AOY52943 | C-terminus, 85 aa | ELXAK |
| calcium-binding adhesion protein | <i>Clostridioides difficile</i> | QWR70506 | Central region, 89 aa | TSXMS |
| phage tail protein | <i>Ligilactobacillus salivarius</i> | RGM22630 | N-terminus, 61 aa | ESXKS |
| porin-like fold protein | <i>Bacteroides ovatus</i> | QUT79494 | N-terminus, 79 aa | DSXYA |
| PQQ-binding-like beta- propeller repeat protein | | QUT82397 | 2 FN3 domains - 63 and 80 aa | B FN3(2) - NSXAV |
| xanthan lyase | <i>Bacteroides uniformis</i> | KAB4110324 | C-terminus, 87 aa | ESXPS |
| | | KAB4109902 | C-terminus, 90 aa | ESXPS |
| fibronectin type III domain-containing protein (ferrous iron transporter A) | <i>Lactobacillus acidophilus</i> | KRK28980 | C-terminus, 74 aa | EGXWS |
| | <i>Lactobacillus helveticus</i> | QAU31920 | C-terminus, 74 aa | EGXWS |
| | <i>Lactobacillus crispatus</i> | QWW29210 | C-terminus, 74 aa | TGXWS |
| | | QWW28111 | C-terminus, 96 aa | VGXWS |
| fibronectin type III domain-containing protein | <i>Phocaeicola dorei</i> | EEB23863 | C-terminus, 92 aa | ESXPS |
| | <i>Bacteroides fragilis</i> | KXU48500 | C-terminus, 92 aa | ESXPS |
| | | KXU42275 | C-terminus, 90 aa | ESXPS |
| | | KXU45955 | C-terminus, 75 aa | FSXYA |
| | <i>Bacteroides ovatus</i> | QUT82347 | C-terminus, 92 aa | ESXPS |

Supplementary Table 3. The occurrence of proteins containing FN3 domains lacking MCRs in bacteria of the families *Lactobacillaceae*, *Bacteroidaceae* and *Clostridioides*.

| Name of the protein harboring FN3 domains | Species | Protein sequence identifier in NCBI | Location of FN3 domain | |
|-------------------------------------------|--------------------------------------|-------------------------------------|------------------------|----------------|
| peptidase S8 | <i>Lactobacillus acidophilus</i> | KRK28186 | Central region | |
| | <i>Lactocaseibacillus casei</i> | KLI75633 | C-terminus | |
| | <i>Lactocaseibacillus paracasei</i> | QDP35932 | Central region | |
| | | QDP37060 | | |
| | | QDP35931 | | |
| | | GFP12550 | Central region | |
| | <i>Lactobacillus helveticus</i> | GFP12480 | | |
| | | GFP12998 | | |
| | | TAR98505 | C-terminus | |
| | <i>Lactiplantibacillus plantarum</i> | KFC34860 | C-terminus | |
| KFC37136 | | Central region | | |
| <i>Ligilactobacillus ruminis</i> | TGJ61417 | Central region | | |
| | <i>Lactobacillus ultunensis</i> | QQP27945 | Central region | |
| | | glycosyl hydrolase family 3 | KLI75030 | Central region |
| glycosyl hydrolase family 20 | <i>Phocaeicola dorei</i> | | EEB26793 | Central region |
| | EEB26209 | | Central region | |
| | EEB26053 | | Central region | |
| | EEB24908 | | Central region | |
| | EEB24073 | | Central region | |
| | EEB25570 | | Central region | |
| glycosyl hydrolase family 20 | <i>Bacteroides fragilis</i> | | KXU48352 | Central region |
| | KXU46540 | | Central region | |
| | KXU42591 | | Central region | |

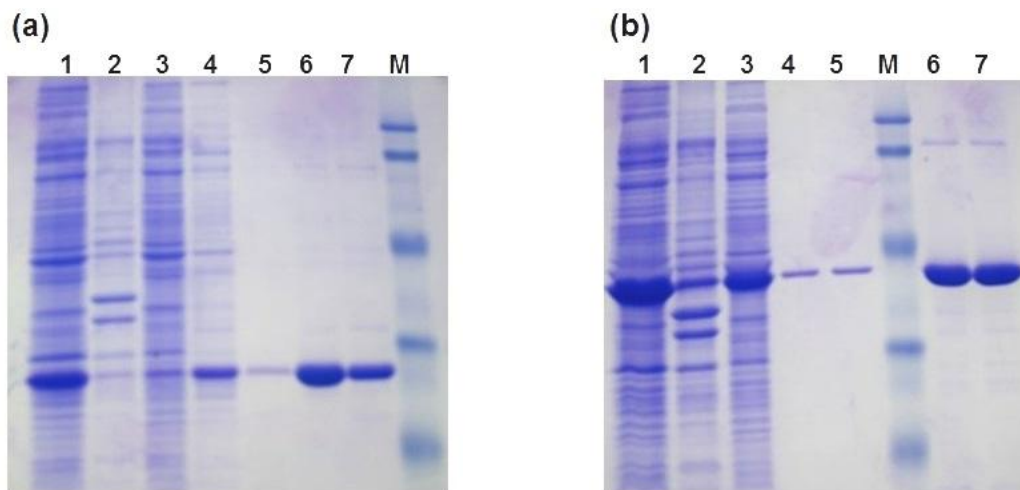
| | | | |
|------------------------------|-------------------------------------------------------|------------|----------------|
| glycosyl hydrolase family 20 | <i>Bacteroides ovatus</i> | QUT83024 | Central region |
| | | QUT82175 | Central region |
| | | QUT81276 | Central region |
| | | QUT80688 | Central region |
| | | QUT79064 | Central region |
| glycosyl hydrolase family 20 | <i>Bacteroides uniformis</i> | KAB4108785 | Central region |
| | | KAB4105426 | Central region |
| glycosyl hydrolase family 20 | <i>Phocaeicola vulgatus</i> | KAB6712571 | Central region |
| | | KAB6712466 | Central region |
| | | KAB6712388 | Central region |
| | | KAB6712252 | Central region |
| | | KAB6708280 | Central region |
| glycosyl hydrolase family 92 | <i>Bacteroides uniformis</i> | KAB4108790 | Central region |
| glycosyl hydrolase family 65 | <i>Clostridium perfringens</i> | AOY54643 | Central region |
| glycosyl hydrolase family 3 | <i>Clostridiales bacterium</i> (human gut metagenome) | PWL55319 | Central region |
| SGNH_hydrolase | | PWL49978 | Central region |

Supplementary Table 4. Sequence homology between FN3 domains of the identified proteins and FN3 domains of human proteins.

| Name of the protein harboring FN3 domains | Species | Protein sequence identifier in NCBI | Homologous % of homology | 3D Structures according to SAS Program Name of the structure | Number in PDB |
|-------------------------------------------|----------------------------------|-------------------------------------|--------------------------|-----------------------------------------------------------------------------------------------|---------------|
| glycosyl hydrolase family 31 | <i>L. acidophilus</i> | KRK27797 | 28,2 | Solution structure of the second fibronectin type III domain of human kiaa1568 protein | 1ujt |
| | <i>L. crispatus</i> | QWW29158 | | | |
| | <i>L. helveticus</i> | GFP12355 | | | |
| | <i>L. ultunensis</i> | QQP28688 | | | |
| | <i>L. gasseri</i> | ABJ59525 | 31,3 | | |
| | <i>L. paragasseri</i> | EFJ69427 | | | |
| glycosyl hydrolase family 43 | <i>B. fragilis</i> | AKA51120 | 28,8 | The solution structure of the second fibronectin type III domain of human neogenin | 1x5g |
| | <i>B. ovatus</i> | QUT79180 | | Solution structures of the fn3 domain of human tripartite motif protein 9 | 2db8 |
| | <i>B. uniformis</i> | KAB4109992 | 25,0 | | |
| glycosyl hydrolase family 18 | <i>Clostridium perfringens</i> * | KAB4109856 | 31,0 | The solution structure of the sixth fibronectin type III domain of human neogenin | 1x5k |
| | | AOY53222 | 31,2 | | |
| glycosyl hydrolase family 20b | | AOY53993 | 28,4 | Titin module a71 from human cardiac muscle, nmr, 50 structures | 1bpv |
| | | AOY54134 | 31,8 | | |
| alpha-N-acetylglucosaminidase | | AOY53263 | 28,1 | The solution structure of the second fibronectin type III domain of human netrin receptor dcc | 2ed8 |
| sialidase | <i>Clostridioides difficile</i> | AOY52943 | 23,9 | Solution structure of the second fibronectin type III domain of human neogenin | 1x5g |
| calcium-binding adhesion protein | | QWR70506 | 33,3 | | |
| phage tail protein | <i>L. salivarius</i> | RGM22630 | 42,9 | Solution structure of the fibronectin type-III domain of human fibronectin type III | 1x5x |

| | | | | | |
|-----------------------------------------------------------------------------|---------------------------|-----------------------------|----------|--------------------------------------------------------------------------------------------------|-------------------------------------------------|
| porin-like fold protein | <i>B. ovatus</i> | QUT79494 | 36,8 | domain containing protein 3 | A26A |
| PQQ-binding-like beta-propeller repeat protein | | QUT82397 | 26,1 | Collagen alpha-1(vii) chain | 6xay |
| xanthan lyase | <i>B. uniformis</i> | KAB4110324 | 30,4 | Structure of a fragment of human fibronectin containing the 10th, 11th and 12th type III domains | |
| | | KAB4109902 | 28,7 | Solution structure of the first fibronectin type iii domain of human kiaa0343 protein | 1uey |
| fibronectin type III domain-containing protein (ferrous iron transporter A) | <i>L. acidophilus</i> | KRK28980* | 26,9 | Solution structure of the second fibronectin type iii domain of human netrin receptor dcc | 2ed8 |
| | | QAU31920* | 30,3 | Collagen alpha-1(vii) chain | A26A |
| | <i>L. helveticus</i> | QWW29210 | 26,0 | Crystal structure of the tenth type iii cell adhesion module of human fibronectin | 1fna |
| | | QWW28111 | 30,6 | Solution structure of the second fibronectin type III domain of human kiaa1568 protein | 1ujt |
| fibronectin type III domain-containing protein | <i>Phocaeicola dorei</i> | EEB23863 | 34,6 | Solution structure of the first fibronectin type iii domain of human kiaa0343 protein | 1uey |
| | | <i>Bacteroides fragilis</i> | KXU48500 | 31,6 | Crystal structure of the titin a-band domain a3 |
| | KXU45955 | | 31,2 | Solution structures of the fn3 domain of human receptor-type tyrosine- protein phosphatase f | 2edy |
| | KXU42275 | | 27,5 | Solution structure of the first fibronectin type iii domain of human kiaa0343 protein | 1uey |
| | <i>Bacteroides ovatus</i> | QUT82347 | 30,9 | | |

KRK28980* and QAU31920* - these proteins also exhibit 31-33% sequence homology with the 3D structure of 6rpx Cytokine receptor-like factor 3 c-terminus residues (*Mus musculus*).



Supplementary Figure 1. Electrophoregram of stepwise of purification on Ni-NTA Superflow of proteins 2D FN3 (a) и CD FN3 (b). Lanes: 1 – lysate, 2 – pellet, 3 – flow-through, 4 – wash_1; 5 – wash_2; 6 – eluate_1; 7 – eluate_2; M – molecular weight protein marker SM0441 (Fermentas, Lithuania).

| | | |
|----------|--------------------------------------------------------------------------------|-----|
| GT15 | PEPPLSPVAGEPQDSAVSLTWTPGAANGSPITNYKVSWTGAATGEKDCGAVTSCQVTGL | 60 |
| MCC10031 | PEPPLSPVAGEPQDSAVSLTWTPGAANGSPITNYKVSWTGAAT V TEKDCGAVTSCQVTGL | 60 |
| MCC10014 | PEPPLSPVAGEPQDSAVSLTWTPGAANGSPITNYKVSWTGAAT V TEKDCG T VTSCQVTGL | 60 |
| MCC10099 | PEPPLSPVAGEPQDSAVSLTWTPGAANGSPITNYKVSWTGAATGEKDCGAVTSCQVTGL | 60 |
| | *****.*****:***** | |
| GT15 | KNGKYSFTVAARNDVGWSKPSTAVEATPKVPSAPTDVTVTGGNKTAKVTWKAPSGDFS | 120 |
| MCC10031 | KNGKYSFTVAARNDVGWSKPSTAVEATPKVPSAPTDVTVTGGNKTAKVTWKAPSGDFS | 120 |
| MCC10014 | KNGKYSFTVAARNDVGWSKPSTAVEATPKVPSAPTDVTVTGGNKTAKVTWKAPSGDFS | 120 |
| MCC10099 | KNGKYSFTVAARNDVGWSKPSTAVEATPKVPSAPTDVTVTGGNKTAKV T WKAPSGDFS | 120 |
| | ***** | |
| GT15 | AVDNYSVTVTGAGAPQTKETGNTATELSFTFNNDISDGTAITATVKAHNKINWSAESAA | 180 |
| MCC10031 | AVDNYSVTVTGAGAPQTKETGNTATELSFTFNNDISDGTAITATVKAHNKINWSAESAA | 180 |
| MCC10014 | AVDNYSVTVTGAGAPQTKETGNTATELSFTFNNDISDGTAITATVKAHNKINWSAESAA | 180 |
| MCC10099 | AVDNYSVTVTGAGAPQTKETGNTATELSFTFNNDISDGTAITATVKAHNKINWSAESAA | 180 |
| | ***** | |
| GT15 | SPSEKIWGDPDAPNIALSNDTTVTAKVTLGNNRNAGCRRISLGGDVKDTIDCSDSGATF | 240 |
| MCC10031 | SPSEKIWGDPDAPNIALSNDTTVTAKVTLGNNRNAGCRRISLGGDVKDTIDCSDSGATF | 240 |
| MCC10014 | SPSEKIWGDPDAPNIALSNDTTVTAKVTLGNNRNAGCRRISLGGDVKDTIDCSDSGATF | 240 |
| MCC10099 | SPSEKIWGDPDAPNIALSNDTTVTAKVTLGNNRNAGCRRISLGGDVKDTIDCSDSGATF | 240 |
| | ***** | |
| GT15 | DINENDLNTREITVTATVVPQRDASSGTGKSSFVPQYKVPPTDVYTTGSGSTCTVHWTK | 300 |
| MCC10031 | DINENDLNTREITVTATVVPQRDASSGTGKSSFVPQYKVPPTDVYTTGSGSTCTVHWTK | 300 |
| MCC10014 | DINENDLNTREITVTATVVPQRDASSGTGKSSFVPQYKVPPTDVYTTGSGSTCTVHWTK | 300 |
| MCC10099 | DINENDLNTREITVTATVVPQRDASSGTGKSSFVPQYKVPPTDVYTTGSGSTCTVHWTK | 300 |
| | ***** | |
| GT15 | QGHAQSYSVTADGIDSGTATNRNSLDFRMSPWAKCNTASVQQVFNGATSDAVTGGPHNGA | 360 |
| MCC10031 | QGHAQSYSVTADGIDSGTATNRNSLDFRMSPWAKCNTASVQQVFNGATSDAVTGGPHNGA | 360 |
| MCC10014 | QGHAQSYSVTADGIDSGTATNRNSLDFRMSPWAKCNTASVQQVFNGATSDAVTGGPHNGA | 360 |
| MCC10099 | QGHAQSYSVTADGIDSGTATNRNSLDFRMSPWAKCNTASVQQVFNGATSDAVTGGPHNGA | 360 |
| | ***** | |
| GT15 | PYTYEKKAVINKPQELRWSDNAHVINVTGGSVDVYGKSATVQIIINGQPFTWNSGQPLDV | 420 |
| MCC10031 | PYTYEKKAVINKPQELRWSDNAHVINVTGGSVDVYGKSATVQIIINGQPFTWNSGQPLDV | 420 |
| MCC10014 | PYTYEKKAVINKPQELRWSDNAHVINVTGGSVDVYGKSATVQIIINGQPFTWNSG Q PLDV | 420 |
| MCC10099 | PYTYEKKAVINKPQELRWSDNAHVINVTGGSVDVYGKSATVQIIINGQPFTWNSG Q PLDV | 420 |
| | ***** | |
| GT15 | TDLVAADGNYAWSVKVTGGPGYSGLDNTADGGSVAGTRPDSSEGTKSLKSGTDVTTLA | 480 |
| MCC10031 | TDLVAADGNYAWSVKVTGGPGYSGLDNTADGGSVAGTRPDSSEGTKSLKSGTDVTTLA | 480 |
| MCC10014 | TDL T VAADGNYAWSVKVTGGPGYSGLDNTADGGSVAGTRPDSSEGTKSLKSGTDVTTLA | 480 |
| MCC10099 | TDL T VAADGNYAWSVKVTGGPGYSGLDNTADGGSVAGTRPDSSEGTKSLKSGTDVTTLA | 480 |
| | ***:***** | |
| GT15 | KNPWILGFGNTPISSTTIHKQ | 501 |
| MCC10031 | KNPWILGFGNTPISSTTIHKQ | 501 |
| MCC10014 | KNPWILGFGNTPISSTTIHKQ | 501 |
| MCC10099 | KNPWILGFGNTPISSTTIHKQ | 501 |
| | ***** | |

Supplementary Figure 2. Alignment of amino acid sequences of Δ FN3.1 proteins from *B. longum* subsp. *longum* GT15 (Group 1), MCC10031 (Group 2), MCC10014 (Group 3), and MCC10099 (Group 4). Amino acid substitutions are highlighted in black.