Functional Metagenomics Reveals Novel Pathways of Prebiotic Breakdown by Human Gut Bacteria

Davide A. Cecchini1,2, Elisabeth Laville1,2, Sandrine Laguerre1,2, Patrick Robe3, Marion Leclerc4, Joël Doré4, Bernard Henrißat5, Magali Remaud-Siméon1,2, Pierre Monsan1,2, Gabrielle Potocki-Véronèse1,2*

1 Université de Toulouse, Institut National des Sciences Appliquées, Université Paul Sabatier, Institut National Polytechnique, Ingénierie des Systèmes Biologiques et des Procédés, Toulouse, France, 2 UMR5504, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, Centre National de la Recherche Scientifique, Institut National de la Recherche Agronomique, Toulouse, France, 3 LibraGenSA, Toulouse, France, 4 Institut National de la Recherche Agronomique, Micalis, UMR1319, Jouy en Josas Cedex, France, 5 Architecture et Fonction des Macromolécules Biologiques, UMR6098, Centre National de la Recherche Scientifique, Universités Aix-Marseille I & II, Marseille, France

Abstract

The human intestine hosts a complex bacterial community that plays a major role in nutrition and in maintaining human health. A functional metagenomic approach was used to explore the prebiotic breakdown potential of human gut bacteria, including non-cultivated ones. Two metagenomic libraries, constructed from ileum mucosa and fecal microbiota, were screened for hydrolytic activities on the prebiotic carbohydrates inulin, fructo-oligosaccharides, xylo-oligosaccharides, galacto-oligosaccharides and lactulose. The DNA inserts of 17 clones, selected from the 167 hits that were identified, were pyrosequenced in-depth, yielding in total 407, 420 bp of metagenomic DNA. From these sequences, we discovered novel prebiotic degradation pathways containing carbohydrate transporters and hydrolysing enzymes, for which we provided the first experimental proof of function. Twenty of these proteins are encoded by genes that are also present in the gut metagenome of at least 100 subjects, whatever are their ages or their geographical origin. The sequence taxonomic assignment indicated that still unknown bacteria, for which neither culture conditions nor genome sequence are available, possess the enzymatic machinery to hydrolyse the prebiotic carbohydrates tested. The results expand the vision on how prebiotics are metabolized along the intestine, and open new perspectives for the design of functional foods.

Introduction

The human gut hosts a complex microorganism community that is dominated by bacterial phyla belonging to Firmicutes, Bacteroidetes and Actinobacteria. The bacterial density increases along the digestive tract to reach 10^7 and 10^11 bacteria mg^-1 of intestinal content in the distal parts of the small intestine (ileum) and of the large intestine (colon), respectively [1–4]. Collectively, gut bacterial genomes contain 150-fold more genes than the human genome equivalent [5,6,7], and code for many metabolic functions that make the gut microbiota a “virtual organ” which provides the host with missing metabolic capabilities that are crucial for maintaining its health and welfare [8,9]. In particular, a huge diversity of enzymes involved in carbohydrate transport and metabolism are encoded by the gut metagenome, as revealed by Gill et al. [10]. Indeed, the intestinal microbiota contributes to host nutrition by harvesting energy from dietary fibers, these carbohydrates that are not digested in the upper part of the digestive tract because the human genome does not encode adequate carbohydrate active enzymes (CAZymes) [11]. Fibers are fermented by gut bacteria to produce organic acids (e.g., lactate, pyruvate, succinate), gases (e.g., H2, H2S, CO2 and CH4), and also short chain fatty acids (SCFA) (e.g., acetate, propionate and butyrate), these compounds contributing to the metabolic balance of the host [12]. Non-digestible carbohydrates include the natural constituents of vegetables, cereals, leguminous seeds and fruits, but also prebiotic oligosaccharides and polysaccharides [13]. Prebiotics are defined as food ingredients that fit to the three following criteria: 1) resistance to gastric acidity, to hydrolysis by mammalian enzymes, and to gastrointestinal absorption; 2) fermentation by intestinal microbiota; and 3) selective stimulation of the growth and/or activity of those intestinal bacteria that contribute to health and well-being [14], in particular bifidobacteria. For this last reason, the world market for prebiotics has grown rapidly in the last three decades [15], focusing on the production of compounds with established prebiotic effects (inulin, fructo-oligosaccharides,
galacto-oligosaccharides and lactulose), as well as development and commercialization of other dietary carbohydrates, like resistant starch [16], xylo-oligosaccharides [17], gluco-oligosaccharides [18,19], polydextrose [20], lactosucrose [21], pectin-derived [22] or soybean oligosaccharides [23].

However, the health benefits of prebiotics are controversial. Indeed, the causal relationship between increasing the number of bifidobacteria in the intestine and improving health status has not been established [24]. There is thus an urgent need to re-evaluate the impact of prebiotics on the intestinal microbiota as a whole, and not only on few cultivated Bifidobacterium and Lactobacillus species of whose growth is stimulated by prebiotics [25,26], but which do not represent the ecosystem complexity. Several studies showed that other genera can be stimulated by prebiotics, directly or through cross-feeding of prebiotic breakdown products, such as Bacteroides [27,28], Faecalibacterium [29], Roseburia, Ruminococcus and Eubacterium [30–32].

However, very little data are available regarding the enzymes that are directly and specifically involved in prebiotic metabolism by gut bacteria. Only a few studies based on sequence functional annotation [33,34], transcriptomics [35–37], proteomics [38] or recombinant enzyme expression and in vitro characterization [26,39–42] reported on the identification of genes and proteins involved in inulin, fructo-, xylo- and galacto-oligosaccharide metabolism by some cultivated Bifidobacterium, Lactobacillus Roseburia, Faecalibacterium and Bacteroides strains. Nevertheless, these strains constitute only a minor fraction of the gut microbial diversity.

In this paper, we present a screening-based functional metagenomics study to identify genes encoding key prebiotic degrading enzymes from both cultivated and non-cultivated gut bacteria. Both the ileum mucosa and fecal microbiota were targeted, to highlight the specific potential of prebiotic metabolism in two gut compartments and locations, which, to our knowledge, has never been studied. Our strategy was first based on the design of functional screens to guide the in-depth sequencing of metagenomic fragments encoding enzymes that hydrolyse five compounds sold as prebiotics for human consumption. Then, the occurrence in the human gut metagenome of the fished genes, obtained from two individuals, was studied in the light of the latest metagenomic data released from the Human Microbiome Project, to identify the genes that are shared by numerous individuals, whatever are their age or their geographical origin [43]. The results provided new insights on metabolism of inulin, fructo-oligosaccharides, lactulose, galacto-oligosaccharides, and xylo-oligosaccharides.

**Results and Discussion**

**Screening the human gut metagenome for prebiotic-degrading activities**

Twenty thousand clones of each of the two E. coli fosmid libraries constructed from fecal and ileum mucosa microbiota (named F and I libraries respectively), covering in total 1.4 Gb of metagenomic sequence, were screened for clone ability to degrade prebiotic compounds of various structures (Table 1). E. coli was chosen as the recombinant host because it was previously shown to allow expression of genes belonging to bacteria that are distantly related from a taxonomical point of view, like Bifidobacteria [44].

First, a high-throughput primary search for clones growing on xylo-oligosaccharides, lactulose and fructo-oligosaccharides was performed using solid selective media containing the oligosaccharides as sole carbon source. Because the expression host EPI100 E. coli strain used to construct the library was unable to metabolize galactose, another strategy presented below has been applied to search for galacto-oligosaccharide degrading clones. The 120, 000 growth assays that were carried out allowed the identification of 11 and 49 hit clones from the F and I libraries, respectively. Hit yields varied between 0.2 and 1.35%, depending on the library and on the oligosaccharide structure (Table 1).

In a second step, we validated and discriminated the prebiotic oligosaccharide hydrolysis ability of selected clones. For this purpose we performed HPAEC-PAD analysis of the products obtained after incubation of the cytoplasmic extracts with the prebiotic compound. All the clones growing on media containing xylo-oligosaccharides, fructo-oligosaccharides and lactulose were able to release mono-saccharides from these substrates (Table 2, Figure S1). In addition, this analysis showed that the clones selected by using fructo-oligosaccharides were also able to degrade inulin. Regarding the search for galacto-oligosaccharide degrading clones, the 14 I library clones found on lactulose were tested for galacto-oligosaccharide hydrolysis. Eleven clones were able to degrade oligosaccharides that were present in the commercial galacto-oligosaccharide preparation. The 107 clones which were identified in a previous study by screening 136, 000 clones from the F library for AZCL-galactan breakdown [44] were tested for galacto-oligosaccharide hydrolysis. All of them were able to degrade galacto-oligosaccharides (Table 1). Depending on the HPAEC-PAD profiles of their prebiotic degradation products (amount and polymerisation degree of the hydrolysis products) the hit clones were classified into 19 clusters. The 17 clones that were representative of clusters containing the most efficient clones were chosen for subsequent pyrosequencing of metagenomic DNA insert (Table 2).

This study provides experimental evidence of the prebiotic degrading potential of non-cultivated gut bacteria, in particular those from ileum. This hydrolysis ability correlates with sequence-based metagenomic results which revealed that functions related to uptake and metabolism of relatively simple carbohydrates (mono- and di-saccharides) are well represented in the small intestine microbiota (6, 7). However, it also proves that ileum bacteria are also able to breakdown longer oligosaccharides (for instance fructo-oligosaccharides of polymerisation degree 5 or higher, like those present in inulin). Numerous prebiotics contain the same glycoside residues and the same glycosidic linkages as plant cell wall polysaccharides (cellulose, hemicellulose, pectin, galactan) or polysaccharides involved in plant energy storage (starch and inulin). Indeed, they are industrially produced by partial hydrolysis of these polysaccharides that also constitute natural dietary fibres and the main carbon sources for gut colonizing bacteria. Unlike the other prebiotics tested in this study, lactulose is not produced from plants; this is an artificial oligosaccharide that is chemically or enzymatically synthesized from lactose [45]. Lactulose is a structural isomer of lactose, a disaccharide that is present in most mammals’ milk, including human [46] and bovine milks [47], and of galacto-disaccharides that derives from the enzymatic hydrolysis of the plant cell wall polysaccharide galactan. We thus propose that at least a part of the metabolic pathways of the natural dietary components lactose and galacto-oligosaccharides of gut bacteria is also responsible for the breakdown of the synthetic prebiotic lactulose.

**Identification of gene clusters assigned to prebiotic catabolism**

Pyrosequencing of metagenomic DNA from the 17 selected clones was performed with a high sequencing depth (between 45X and 171X), ensuring reliable sequence assembly. For each clone, one contig sizing between 13, 000 and 39, 000 bp was obtained,
except for clones 9 and 10 which generated two contigs each. Partial or complete sequence redundancy was observed for 4 clones from the I library and 6 clones from the F library, respectively (Table 3). As most of these redundant sequences were not assignable from a taxonomic point of view (based on sequence homology analysis with sequenced genomes), it was impossible to determine if their bacterial origin corresponds to over-represented species in the samples used for library construction.

In total, 213,780 bp (containing 172 ORFs) and 193,640 bp (containing 158 ORFs) of metagenomic DNA were analysed for F and I libraries, respectively. On average, 26% of the detected ORFs were assigned to the G COG cluster, corresponding to proteins that are predicted to be involved in Carbohydrate Transport and Metabolism (Tables 3 and S1). The dominance of G COG cluster, which was not observed in the randomly sequenced human gut metagenome [5,48,49], reflects as previously described [44], the power of screening-based strategies to guide the sequencing of DNA fragments that are rich in genes of interest.

### Table 1. Structure of the carbohydrate compounds used for primary and secondary screening and number of hit clones isolated from the fecal (F) and the ileum mucosa (I) libraries.

<table>
<thead>
<tr>
<th>compounds</th>
<th>Screening steps*</th>
<th>Structure</th>
<th>Number of hit clones (Hit yield)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>F library</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>I library</td>
</tr>
<tr>
<td>XOS*</td>
<td>1 and 2</td>
<td>Xyl-[β(1,4)-Xyl]n (1 ≤ n ≤ 7)</td>
<td>4 (0.20%a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>27 (1.35%)a</td>
</tr>
<tr>
<td>FOS*</td>
<td>1 and 2</td>
<td>Glc-[α(1,2)-β(1,2)-Fru]n (1 ≤ n ≤ 4)</td>
<td>7 (0.35%a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8 (0.4%a)</td>
</tr>
<tr>
<td>Inulin</td>
<td>2</td>
<td>Glc-[α(1,2)-β(1,2)-Fru]n (10 ≤ n)</td>
<td>7 (0.35%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8 (0.4%)</td>
</tr>
<tr>
<td>AZCL-galactan</td>
<td>1</td>
<td>[β(1,4)-Gal]n (n = 100)</td>
<td>107 (0.79%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NTd</td>
</tr>
<tr>
<td>GOS*</td>
<td>2</td>
<td>[β(1,4)-Gal]n-[β(1,4)-Gal]n-[β(1,4)-Gl] (0 ≤ n ≤ 14)</td>
<td>107 (0.79%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11 (0.55%)</td>
</tr>
<tr>
<td>Lactulose</td>
<td>1 and 2</td>
<td>Gal-[β(1,4)-Fru]n</td>
<td>NTd</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>14 (0.7%)</td>
</tr>
</tbody>
</table>

* clones, xylo-oligosaccharides; FOS, fructo-oligosaccharides; GOS, galacto-oligosaccharides.

*Clones positive on fructo-oligosaccharides were further assayed for inulin degrading activity.

*Clones positive on lactulose (I library) and those identified by screening the entire F library (136,000 clones) on AZCL-galactan were further assayed on galacto-oligosaccharides.

NTd: not tested.

*Screening steps: 1, high-throughput primary screen; 2, HPAEC-PAD based secondary screen.

doi:10.1371/journal.pone.0072766.t001

### Table 2. Clone classification according to the amount and degree of polymerisation of prebiotic hydrolysis products, determined by HPAEC-PAD analysis.

<table>
<thead>
<tr>
<th>Origin of the metagenomic library</th>
<th>Prebiotics</th>
<th>Cluster</th>
<th>Hydrolysis products</th>
<th>Number of positive clones</th>
<th>Identifiers of sequenced clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feces</td>
<td>FOS</td>
<td>1 Glc, Fru</td>
<td>Glc, Fru</td>
<td>4</td>
<td>F1-F15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Glc, Fru, Glc-[α(1,2)-Fru]</td>
<td></td>
<td>2</td>
<td>F2-F16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 Low amounts of Fru, Glc-[α(1,2)-β(1,2)-Fru]n (1 ≤ n ≤ 3)</td>
<td>1</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inulin</td>
<td>1 Glc, Fru, Glc-[α(1,2)-β(1,2)-Fru]n (n ≤ 10)</td>
<td>4</td>
<td>F2-F15-F16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Low amounts of Glc, Fru and Glc-[α(1,2)-β(1,2)-Fru]n (n ≤ 10)</td>
<td>2</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 Glc, Fru</td>
<td>Glc-[β(1,4)-Xyl]n</td>
<td>1</td>
<td>F1</td>
</tr>
<tr>
<td></td>
<td>XOS</td>
<td>1 Xyl</td>
<td>Xyl, Xyl-[β(1,4)-Xyl]</td>
<td>3</td>
<td>F3-F5-F17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Xyl and Xyl-[β(1,4)-Xyl]</td>
<td></td>
<td>1</td>
<td>F4</td>
</tr>
<tr>
<td></td>
<td>GOS</td>
<td>1 Gal, Glc</td>
<td>Gal, Glc</td>
<td>74</td>
<td>F6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 [β(1,4)-Gal]n (4 ≤ m ≤ 7), [β(1,4)-Gal]n-[β(1,4)-Gal]n-[β(1,4)-Gl] (2 ≤ n ≤ 5)</td>
<td>33</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>ileum mucosa</td>
<td>FOS</td>
<td>1 Glc, Fru</td>
<td>Glc-[β(1,4)-Xyl]</td>
<td>6</td>
<td>I10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Glc, Fru, Glc-[α(1,2)-β(1,2)-Fru]n (2 ≤ n ≤ 4)</td>
<td>2</td>
<td>I9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>XOS</td>
<td>1 Xyl</td>
<td>Xyl</td>
<td>18</td>
<td>I7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Xyl, Xyl-[β(1,4)-Xyl]</td>
<td></td>
<td>8</td>
<td>I8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 Low amount of Xyl</td>
<td></td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Lactulose</td>
<td>1 Gal, Fru</td>
<td>Gal and Fru</td>
<td>12</td>
<td>I11-I12-I13-I14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Low amount of Gal and Fru</td>
<td></td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>GOS</td>
<td>1 Gal, Glc, [β(1,4)-Gal]n-[β(1,4)-Gl] (1 ≤ n ≤ 14)</td>
<td>9</td>
<td>I11-I14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Gal, Glc, [β(1,4)-Gal]n-[β(1,4)-Gl] (1 ≤ n ≤ 3)</td>
<td>2</td>
<td>I13</td>
<td></td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0072766.t002
than 90% sequence identity with enzymes from (Figure 1, Table S1). Five of the discovered GHs share more families), surrounded by putative transcriptional regulators ABC transport family, to PTS systems or to SusD and SusC carbohydrate binding, metabolism and transport (belonging to the carbohydrate transport and catabolism highlighted of the gut bacteria of from which came the metagenomic inserts: taxonomical and functional assignation, and gene occurrence in the human gut metagenome.

Table 3. Sequence analysis of the metagenomic inserts isolated in this study: taxonomical and functional assignation, and gene occurrence.

<table>
<thead>
<tr>
<th>Sequenced Clone</th>
<th>Contig name and size</th>
<th>ORF number</th>
<th>Taxonomic assignation</th>
<th>Substrate</th>
<th>CAZy family</th>
<th>Number of ORFs assigned to COG</th>
<th>Occurrence analysis $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>1 $^a$ (33, 125bp)</td>
<td>23</td>
<td>Clostridiales $^b$</td>
<td>FOS &amp; Inulin</td>
<td>GH32 $^c$</td>
<td>3</td>
<td>21/8</td>
</tr>
<tr>
<td>F2</td>
<td>2 $^a$ (37, 022bp)</td>
<td>33</td>
<td>Bifidobacterium longum</td>
<td>FOS &amp; Inulin</td>
<td>GH32 $^c$</td>
<td>2</td>
<td>32/27</td>
</tr>
<tr>
<td>F3</td>
<td>3 $^a$ (35, 645bp)</td>
<td>26</td>
<td>Clostridiales $^b$</td>
<td>XOS</td>
<td>GH67, GH2, GH3 $^d$</td>
<td>6</td>
<td>24/24</td>
</tr>
<tr>
<td>F4</td>
<td>4 (31, 477bp)</td>
<td>34</td>
<td>Bifidobacterium adolescentis</td>
<td>XOS</td>
<td>GH113, GH77, GH43 $^e$, GH8 $^f$, GH120 $^g$</td>
<td>5</td>
<td>27/17</td>
</tr>
<tr>
<td>F5</td>
<td>5 (39, 093bp)</td>
<td>27</td>
<td>Bacteroides vulgatus</td>
<td>XOS</td>
<td>GH43 $^e$, GH43 $^e$, GH43 $^e$, GH10 $^h$, GH116, CE9</td>
<td>9</td>
<td>24/24</td>
</tr>
<tr>
<td>F6</td>
<td>6 (37, 418bp)</td>
<td>29</td>
<td>Bifidobacterium longum</td>
<td>GOS</td>
<td>GH13-CBM48, GH2 $^i$, GH42 $^j$</td>
<td>6</td>
<td>25/25</td>
</tr>
<tr>
<td>I7</td>
<td>7 $^a$ (37, 821bp)</td>
<td>18</td>
<td>Bacteroidales $^b$</td>
<td>XOS</td>
<td>GH115, GH10 $^i$, GH43 $^e$</td>
<td>5</td>
<td>17/0</td>
</tr>
<tr>
<td>I8</td>
<td>8 $^a$ (38, 518bp)</td>
<td>19</td>
<td>Bacteroidales $^b$</td>
<td>XOS</td>
<td>GH35, GH67, GH115, GH10 $^i$, GH43 $^e$</td>
<td>7</td>
<td>17/0</td>
</tr>
<tr>
<td>I9</td>
<td>9 $^a$ (14, 714bp)</td>
<td>12</td>
<td>Dorea longicatena</td>
<td>FOS</td>
<td>GH32 $^c$</td>
<td>3</td>
<td>10/3</td>
</tr>
<tr>
<td>I10</td>
<td>10 $^a$ (11, 460bp)</td>
<td>10</td>
<td>Eubacterium rectale</td>
<td>FOS</td>
<td>GH32 $^c$</td>
<td>4</td>
<td>9/9</td>
</tr>
<tr>
<td>I11</td>
<td>11 $^a$ (30, 347bp)</td>
<td>24</td>
<td>Streptococcus thermophilus</td>
<td>Lactulose &amp; GOS</td>
<td>GT2-GT8, GH2 $^c$</td>
<td>5</td>
<td>19/1</td>
</tr>
<tr>
<td>I12</td>
<td>12 $^a$ (13, 317bp)</td>
<td>12</td>
<td>Clostridiales $^b$</td>
<td>Lactulose</td>
<td>GH42 $^c$, GH2 $^c$</td>
<td>6</td>
<td>10/9</td>
</tr>
<tr>
<td>I13</td>
<td>13 $^a$ (32, 063bp)</td>
<td>30</td>
<td>Faecalibacterium prausnitzii</td>
<td>Lactulose &amp; GOS</td>
<td>GH2$^c$, GH77, GT3S</td>
<td>5</td>
<td>29/11</td>
</tr>
<tr>
<td>I14</td>
<td>14 $^a$ (26, 887bp)</td>
<td>22</td>
<td>Streptococcus thermophilus</td>
<td>Lactulose &amp; GOS</td>
<td>CBM41-CBM48, GH13, GH13, GH2 $^c$</td>
<td>8</td>
<td>16/8</td>
</tr>
</tbody>
</table>

$^a$Contig 7 is partially redundant with contig 8 (ORFs 1 to 14 of contig 7 are identical to ORFs 5 to 19 of contig 8), contig 11 is partially redundant with contig 14 (ORFs 9 to 12 and 18 to 24 of contig 11 are identical to ORFs 11 to 22 of contig 14). In addition, contig sequences of clones 15, 16 and 17 were totally covered by contig 1, 2 and 3 sequences, respectively.

$^b$MEGAN based assignation.

$^c$CAZy families known to degrade the glycosidic bonds contained in the prebiotic substrate.

$^d$Results of BLASTN comparison of the predicted ORFs with the metagenomic data sets issued from fecal sampling of 3 cohorts of 163, 139 and 110 subjects. The first number corresponds to the number of ORFs that present a sequence identity $\geq 90\%$ and an E-value $= 0$ with a gene detected at least in one subject of the three cohort. The second number corresponds to the number of ORFs that present a sequence identity $\geq 90\%$ and an E-value $= 0$ with a gene detected at least in 20 subjects.

Among these 330 ORFs, 38 were annotated as CAZyme encoding genes. Thirty-three of these predicted CAZymes contained a glycoside-hydrolase (GH) module (Tables 3, S1 and S2), others being esterases and glycosyl-transferases, which are not involved in prebiotic breakdown.

Moreover, on 11 contigs, the putative GHs are encoded by operon-like multigenic systems encoding proteins involved in carbohydrate binding, metabolism and transport (belonging to the ABC transport family, to PTS systems or to SusD and SusC families), surrounded by putative transcriptional regulators (Figure 1, Table S1). Five of the discovered GHs share more than 90% sequence identity with enzymes from Bifidobacterium species, which were already shown to be involved in prebiotic breakdown: one GH32 (contig F2) involved in fructo-oligosaccharide degradation [50–52], three GH43, GH8 and GH120 (contig F4), synergistically involved in the metabolism of xylo-oligosaccharides [42,53,54], and the last two predicted enzymes assigned to families GH2 and GH42 (contig F6), involved in the hydrolysis of galacto-oligosaccharides [55]. For all the other pathways of carbohydrate transport and catabolism highlighted here (encoded by contigs F1, F3, F5, F7, F8, F9, I10, I11, I14, I12, and I13), the present screening results constitute the first biochemical evidence of their ability to participate to prebiotic metabolism. Moreover, these results contribute to indicate functions of putative CAZyme sequences belonging to families which are often multispecific, and which contain only between 3 (for GH2) and 11% (for GH32) of characterized enzymes.

Which are the bacteria that possess these prebiotic catabolic machineries?

To get new insights into the relationships between bacteria taxonomy and their role in prebiotic catabolism, the origin of the metagenomic DNA sequences was predicted. All the contigs came from bacteria of which the genome has not yet been sequenced, as none of the contig sequences were perfectly identical to sequenced genomes. However, for the contigs F2, F4, F5, F6, I10, I11, I13 and I14, the high sequence similarity with reference genomes made it possible to identify at the species level the nearest homolog of the gut bacteria of from which came the metagenomic inserts (Table 3). A near perfect synteny was observed with the genomic sequences of Bifidobacterium longum NCC2705 (contigs F2 and F6), Bifidobacterium adolescentis ATCC 15703 (contig F4), Eubacterium rectale ATCC 33656 (contig I10a), Streptococcus thermophilus LMD9 and LMG 18311 (contigs I11 and I14) (Figure 1 and Table S1). Regarding the contig F5, synteny with Bacteroides vulgatus ATCC 8482 was interrupted in the region carrying the GHs proposed to be responsible for the screened activity. Synteny between contig I13 and the Faecalibacterium prausnitzii KLE1255 genome was also highly perturbed. The gut bacteria from which came the DNA inserts F2, F4, F5, F6, I10, I11, I13 and I14 are thus probably very
near, from a taxonomical point of view, of strains which were already described as being probiotics, carbohydrate fermenters, or, as being able to metabolize xylo-, fructo-, galacto-oligosaccharides or inulin [22,30,32,33,55–60]. However, as mentioned upper, apart from those of Bifidobacterium longum NCC2705, Bifidobacterium adolescentis ATCC 15703 and Streptococcus thermophilus LMD9, the enzymes involved in prebiotic degradation were not experimentally identified to date. The present metagenomic sequences are thus highly useful to pinpoint the corresponding gene machinery in the genome of cultivated strains. This is of prime importance for the design of innovative synbiotics, these functional food mixtures containing both a probiotic strain and the prebiotic compound that it is able to metabolize.

Regarding this aspect, clone I13, which was able to degrade lactulose and galacto-oligosaccharides, is of particular interest. Indeed, a cluster containing 17 putative genes (ORFs CCG34912 to CCG34928) were annotated as being involved in cobalamin (B12 vitamin) synthesis (Table S1). This vitamin plays a key role in the brain and in the nervous system [61]. Its absorption occurs by receptor-mediated endocytosis in the terminal ileum [62]. Cobalamin was already shown to be synthesized by bacteria present in the distal gut, in particular Lactobacillus reuteri strains that are marketed as probiotics and commonly used in the food industry [63]. Here, we showed that i) an ileum colonizing strain possesses both the machineries for lactulose/galacto-oligosaccharides hydrolysis and for putative cobalamin synthesis, ii) proteins which would be involved in prebiotic degradation and in vitamin synthesis share more than 90% identity with proteins of Faecalibacterium prausnitzii KLE1255. These results allow enlarging the panel of prebiotic compounds that could stimulate the growth of Faecalibacterium prausnitzii species that could beneficially impact health [29,64]. In particular, it would be interesting to test both the ability of Faecalibacterium prausnitzii KLE125 to produce vitamin B12 and to hydrolyse lactulose and galacto-oligosaccharides, to evaluate its interest as a probiotic strain in synbiotics.

No significant sequence homology with lactobacilli genomic sequences was found, even if these bacteria are reported, with bifidobacteria, to be the most common prebiotic targets. This is probably due to the lower abundance of Lactobacillus species in the human gut microbiota, compared to Bifidobacterium species [2,5], and to the fact that the gut metagenome diversity was not sufficiently covered by 40,000 clones of 30 to 40 kbp each to isolate such low abundant sequences. Therein, based on 16S

---

**Figure 1. Operon-like structures encoding prebiotic degrading enzymes.** The operon-like multigenic systems encoding proteins involved in carbohydrate binding, metabolism and transport surrounded by putative transcriptional regulators are shown. Genes are represented by arrows. Under the operon-like structures are reported the strains for which perfect synteny was identified between part of their genome and the present metagenomic sequences.

doi:10.1371/journal.pone.0072766.g001
Occurrence in the human gut metagenome of genes assigned to prebiotic breakdown pathways

Finally, we compared our metagenomic sequences, obtained from our 2 different subjects, with three different gene catalogues obtained from the random sequencing of fecal metagenomes, sampled from 162 subjects of MetaHIT cohorts [5,69]; 139 subjects of NIH-HMP cohort [70] and other 110 subjects from different cultures, ages and families [43]. Striking results were obtained. Indeed, 87% of the genes identified in this study were at least 90% identical to genes detected in the gut metagenomic sequences of at least one subject of these cohorts (62% of genes being 100% identical), and 53% were at least 90% identical to sequences obtained from at least 20 individuals (Tables 3 and S1).

Anyway, the 20 genes encoding the operon-like gene clusters were prebiotic degrading enzymes obtained from clones F1 (non assigned), F3 (non assigned), F5 (assigned to Bacteroides vulgatus) and I10 (assigned to Eubacterium rectale) had highest occurrences, whatever the targeted cohort. Indeed, they all are at least 90% identical to genes found in at least 100 different individuals. These results indicate that the present function-driven metagenomics study allowed the discovery of some of the major prebiotic metabolism machineries along the human digestive tract.

Concluding remarks

In this study, we were able, by using a functional metagenomics approach, to directly access key enzymes involved in prebiotic metabolism by gut bacteria, including non-cultivated ones. We provided the first experimental evidence of the prebiotic hydrolytic potential in the ileum, and showed that in this gut compartment, bacteria are not only able to breakdown short carbohydrates like di-saccharides, but also longer ones, thanks to both exo- and endo-acting glycoside-hydrolases (like for instance polysaccharide degraders classified in the GH10 family).

Our results are consistent with the known ability of bifidobacteria to hydrolyse prebiotics, and allowed to evidence the prebiotic metabolism potential of strains currently used as probiotics or assumed to have health-promoting properties, which could thus be used to design novel functional foods like symbiotics. However, we also showed the high occurrence in human gut of still unknown bacteria possessing the enzymatic machinery to breakdown these compounds. Their growth could thus also be selectively stimulated by diet supplementation with prebiotics, thus generating effects on human health that remain to be determined. Prebiotic compounds, that were previously thought to direct a very specific response on human gut microbiota, exhibit a broader spectrum of action regarding target species but also gastro intestinal tract location. However, metagenome profiling, comparative metatranscriptomics and metaproteomics studies would have to be performed to evaluate the involvement level of the proteins that were identified here in real physiological conditions, with and without diet supplementation with prebiotics.

Materials and Methods

Metagenomic library construction

Two different gut microbiota were sampled as previously described [44,71], in the frame of a study registered as a clinical trial (http://clinicaltrials.gov/, Calibrated Diets and Human Intestinal Microflora NCT00639561) and carried out at the Center for Clinical Investigation (CCI) of Grenoble University Hospital (France). The first one was a fecal sample of a healthy 30 years old male, who followed a vegetarian and fish-eating diet for twelve years before sampling. The distal ileum sample was obtained from a 51 years old male patient undergoing colonoscopy and surgery for lower colon cancer suspicion, after he has been submitted to a cleansing preparation. A segment of 2 cm² was obtained, immediately frozen and kept at −80°C until processing. Both individuals showed an average body mass index. They did not eat any functional food such as prebiotics or probiotics, nor did receive any antibiotics or other drugs during 6 months before sampling. The sampling protocols were approved by the local ethics committee “Comité de Protection des Personnes Sud Est V” (Ref 07-CHUG-21, composed of B.Habozit, J.L.Crakowski, J.Juge, J.Grunwald, E.Svhan and E.Fontaine) and informed written consent was obtained from the subjects before sampling.

Metagenomic DNA was extracted, and fragments ranging in size from 30 to 40 kb DNA were isolated and cloned into pCG1FOS fosmid (Epícentre Technologies). EPI100 E. coli cells were then transfected to obtain a 156,000 clone library from the fecal sample [44], and a 20,000 clone library from the ileum sample [71]. Recombinant clones were transferred to 384-well microtiter plates containing Luria-Bertani medium, supplemented with 12.5 mg/L chloramphenicol and 8% glycerol. After 22 h of growth at 37°C without any agitation, the plates were stored at −80°C.

Carbohydrate sources

Commercial prebiotic oligosaccharides and polysaccharides used for screening were the following (Table 1): xylo-oligosaccharides (Iro Taihe International, China), fructo-oligosaccharides

rRNA gene sequencing results, no Lactobacillus bacteria was found in the microbiota sample which was used to construct the F library [44], while Bifidobacterium bacteria represented 1% of this sample. No Lactobacillus bacteria was found in the ileum mucosa sample as well, but in this case, PCR based 16S rRNA gene taxonomic characterization was based only on 96 sequences, which is insufficient to be statistically reliable. One should thus not exclude that either cloning [65] or expression bias [66] may also which is insufficient to be statistically reliable. One should thus not exclude that either cloning [65] or expression bias [66] may also
(Actilight-FOS, Beigin Meiji, France), inulin (from dahlia tuber, Sigma, Germany), galacto-oligosaccharides (Vivarial-GOS, FrieslandFoods Domo, The Netherlands), lactulose (TEVA, France).

XYO-oligosaccharide preparations were purified by preparative chromatography to remove xylose, as following: 250 ml of a 50% (w/v) carbohydrate solution was loaded in a preparative-column (bed volume 7.8l, diameter 8 cm, column length 100 cm) filled with an ionic exchange resin (Amberlite CR1320K). Oligosaccharides were eluted by using MilliQ H2O and the carbohydrate content of each fraction was analysed by high-performance anion-exchange chromatography coupled to pulsed amperometric detection (HPAEC-PAD) as described in the following section. Fractions containing carbohydrates of polymerisation degree ≥2 were pooled, and the resulting solution was freeze-dried with an SMH 15 freeze-dryer (Société Nouvelle Usifroid, France). The final purity of the oligosaccharide powder was determined by HPAEC-PAD.

**High throughput functional screening**

Twenty thousand clones from each library were gridded onto 22×22 cm trays containing minimum solid medium (Na2HPO4 6 g l⁻¹, KH2PO4 3 g l⁻¹, NH4Cl 1 g l⁻¹, NaCl 0.5 g l⁻¹, MgSO4 0.12 g l⁻¹, CaCl2·2H2O 0.015 g l⁻¹, FeSO4·7H2O 0.0042 g l⁻¹, thiamine hydrochloride 0.005 g l⁻¹, leucine 0.04 g l⁻¹, agar type E 15 g l⁻¹ and chloramphenicol 12.5 mg l⁻¹) supplemented with prebiotic oligosaccharides as sole carbon source, using a QPixII colony picker (Genetix), with a density of 2,304 clones per tray. Contigs showing a sequence length <1,000 bp and a sequencing depth <8 were removed. Remaining contigs were cleaned from the vector pCC1FOS sequence using Crossmatch (http://www.sanger.ac.uk/Software/CROSSMATCH).

Open reading frames (ORF) of at least 20 amino acids were predicted using MetaGene [73]. Annotated contig sequences were deposited in European Nucleotide Archive under accession numbers: HE663537 and HE717006 to HE717020 (Table S2).

**ORF analysis**

CAZyme encoding genes were identified by BLAST analysis of the predicted ORFs against the functional modules of glycoside hydrolases, polysaccharide lyases, carbohydrate esterases, carbohydrate-binding modules and glycosyltransferases included in the CAZy database (http://www.cazy.org) using a cut-off E-value of 1e⁻⁰⁸. Contigs showing a similarity with sequenced genomes, using BLASTX analysis (E-value <1e⁻⁰⁸, identity >35%, query length coverage ≥50%), ORFs were assigned to clusters of orthologous groups of proteins (COGs) using RPS-BLAST analysis against the CAZy database (E-values ≤1e⁻⁰⁸). Finally, gene occurrence in the human gut microbiota was determined by TLASTN comparisons of the predicted ORFs with the metagenomic data sets available, from fecal sampling of 162 subjects of the MetaHIT cohort [http://www.bork.embl.de/arumugam/Qin_et_al_2010.html; http://www.bork.embl.de/Docu/Arumugam_et_al_2011/data/genomes/; [5,69]]; of 139 U.S. subjects from the NIH-HMP cohort [http://www.hmpdacc.org/HMASM/; [70] and of the 110 subjects from different cultures, ages and families (http://www.metagenomics.anl.gov/projects/98; [43] (E-value = 0, identity ≥90% and ≥100%).

Contig taxonomic assignment was based on ORF sequence similarity with sequenced genomes, using BLASTX analysis against the non-redundant NCBI database (E-value ≤1e⁻⁰⁸, identity ≥90%, query length coverage ≥50%). Contigs were assigned to a class, genus or species only if at least 50% of the ORFs were assigned to the same organism. Contigs containing ORFs assigned to different classes were not assigned. The most probable common ancestor of the organism from which the non-assignable contigs came from was retrieved using MEGAN v3.2.1, [74], based on BLASTX analysis against the non-redundant NCBI.

**Supporting Information**

Figure S1 HPAEC-PAD analysis of reaction products resulting from prebiotic hydrolysis.

(DOC)
Table S1: Characteristics of metagenomic clones.

Table S2: Glycoside hydrolase families found in the study.

Acknowledgments

The high-throughput screening work was performed at the Laboratory for BioSystems & Process Engineering (Toulouse, France) with the ICEO automated facility. We thank Sophie Bozonnet for her assistance.

References


Author Contributions

Conceived and designed the experiments: GPV MRS PM JD. Performed the experiments: DAC EL PR SL. Analyzed the data: DAC EL BH ML. Wrote the paper: DAC EL GPV.

PLOS ONE | www.plosone.org 8 September 2013 | Volume 8 | Issue 9 | e72766
Prebiotic Metabolism by Gut Bacteria


