



HAL
open science

A metagenomic beta-glucuronidase uncovers a core adaptive function of the human intestinal microbiome

Karine K. Gloux, Olivier O. Berteau, Hanane H. El Oumami, Fabienne F. Beguet-Crespel, Marion M. Leclerc, Joel J. Dore

► To cite this version:

Karine K. Gloux, Olivier O. Berteau, Hanane H. El Oumami, Fabienne F. Beguet-Crespel, Marion M. Leclerc, et al.. A metagenomic beta-glucuronidase uncovers a core adaptive function of the human intestinal microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, 2011, 108, pp.4539 - 4546. 10.1073/pnas.1000066107 . hal-01001013

HAL Id: hal-01001013

<https://hal.science/hal-01001013>

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A metagenomic β -glucuronidase uncovers a core adaptive function of the human intestinal microbiome

Karine Gloux, Olivier Berteau, Hanane El Oumami, Fabienne Béguet, Marion Leclerc, and Joël Doré¹

Institut National de la Recherche Agronomique, Unité Mixte de Recherche 1319 Micalis, F-78352 Jouy en Josas, France

Edited by Todd R. Klaenhammer, North Carolina State University, Raleigh, NC, and approved June 1, 2010 (received for review February 4, 2010)

In the human gastrointestinal tract, bacterial β -D-glucuronidases (BG; E.C. 3.2.1.31) are involved both in xenobiotic metabolism and in some of the beneficial effects of dietary compounds. Despite their biological significance, investigations are hampered by the fact that only a few BGs have so far been studied. A functional metagenomic approach was therefore performed on intestinal metagenomic libraries using chromogenic glucuronides as probes. Using this strategy, 19 positive metagenomic clones were identified but only one exhibited strong β -D-glucuronidase activity when subcloned into an expression vector. The cloned gene encoded a β -D-glucuronidase (called H11G11-BG) that had distant amino acid sequence homologies and an additional C terminus domain compared with known β -D-glucuronidases. Fifteen homologs were identified in public bacterial genome databases (38–57% identity with H11G11-BG) in the Firmicutes phylum. The genomes identified derived from strains from Ruminococcaceae, Lachnospiraceae, and Clostridiaceae. The genetic context diversity, with closely related symporters and gene duplication, argued for functional diversity and contribution to adaptive mechanisms. In contrast to the previously known β -D-glucuronidases, this previously undescribed type was present in the published microbiome of each healthy adult/child investigated ($n = 11$) and was specific to the human gut ecosystem. In conclusion, our functional metagenomic approach revealed a class of BGs that may be part of a functional core specifically evolved to adapt to the human gut environment with major health implications. We propose consensus motifs for this unique Firmicutes β -D-glucuronidase subfamily and for the glycosyl hydrolase family 2.

functional core | intestinal microbiota | functional metagenomics | glycosyl hydrolase I Firmicutes

Bacterial β -D-glucuronidases (BG; E.C. 3.2.1.31) are part of the extensive human intestinal microbiome and are involved in the metabolism and bioavailability of food and drug compounds in the human body. They catalyze the hydrolysis of exogenous β -glucuronides naturally occurring in diet and drugs as well as endogenous β -glucuronides produced in the liver by glucuronosyltransferases, a major xenobiotic detoxication pathway. Indeed, numerous compounds including metabolites, vitamins, steroid hormones, xenobiotics, and drugs are excreted via the bile and the digestive tract after conversion to a more hydrophilic glucuronidated form. Nevertheless, secondary deglucuronidation, primarily due to intestinal bacteria (1), promotes recycling of the aglycone forms through the enterohepatic cycle, which prevents their removal from the human body. This phenomenon is particularly well described in the case of xenobiotics (2, 3) and is suggested for circulating hormones (4, 5). Despite their detrimental effects, bacterial BGs are also thought to have beneficial effects, notably on the bioavailability of active metabolites derived from dietary compounds, including lignans, flavonoids, sphingolipids, and glycyrrhizin (6–8).

Finally, observations argue in favor of the involvement of bacterial BGs in several pathologies. Indeed, a lower BG activity was detected in the feces of Crohn's disease patients compared to healthy subjects (9). Conversely, a high BG activity is recognized as a prognosis marker for colon cancer (10). Furthermore, the *Clostridium leptum* group (cluster IV) or Lachnospiraceae (cluster

XIVa), major reservoirs of BG-positive bacteria (11), are dysbiosis markers in Crohn's disease (12–15) and in intestinal carcinogenesis (16). Nevertheless, the relative contribution of bacterial strains to the global intestinal BG activity remains unclear, and the molecular bases are essentially unknown.

In the human digestive tract, bacterial BGs are known to be distributed among the Enterobacteriaceae family in some Firmicutes genera (*Lactobacillus*, *Streptococcus*, *Clostridium*, *Ruminococcus*, *Roseburia*, and *Faecalibacterium*) and in a specific species of Actinobacteria (*Bifidobacterium dentium*) (11, 17, 18). On the basis of their sequence, BGs are highly homologous to β -galactosidases and only a few of them have been investigated and clearly annotated as β -D-glucuronidases in sequence databases. The most notorious prokaryotic BGs are those encoded by *uidA* homologs (also named *gusA*) from *Escherichia coli*, *Lactobacillus gasseri*, and *Ruminococcus gnavus* (17, 19, 20).

In the present study, we report the identification of a BG from intestinal bacteria by a functional metagenomic approach. The identified enzyme is unrelated to the previous ones and expands the number of bacterial enzymes responsible for glucuronidase activity. Furthermore, its sequence and distribution among bacterial genomes and metagenomes highlights the relevance and the diversity of BGs within the intestinal microbiome.

Results

High Frequency and Diversity of BG Activities in Metagenomic Libraries. A strategy was developed to assay BG activity from metagenomic inserts in spite of the constitutive activity of the *E. coli* strain used to build the metagenomic libraries (*E. coli* DH10B) (Fig. 1). The method consisted of a two-step screen of clones containing large metagenomic inserts. Using *p*-nitrophenyl β -D-glucuronide as chromogenic substrate (PNP-G deglucuronidation test), the clones' activity was first compared to the basal activity of the receiving strain. Fosmids born by positive clones were further transferred into an *E. coli* strain deprived of BG activity (*E. coli* L90 $\Delta uidA$). The clones were then grown with 5-bromo-4-chloro-3-indolyl- β -D-glucuronide in the medium (X-GlcA deglucuronidation test) to confirm BG activity.

With this strategy, two libraries from human gut microbiota were screened: a metagenomic library of 4,608 clones combining

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Microbes and Health," held November 2–3, 2009, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. The complete program and audio files of most presentations are available on the NAS Web site at http://www.nasonline.org/SACKLER_Microbes_and_Health.

Author contributions: K.G., O.B., and J.D. designed research; K.G. performed research; K.G., H.E.o., and F.B. contributed new reagents/analytic tools; K.G., O.B., and J.D. analyzed data; and K.G., O.B., M.L., and J.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: BG sequences from C7D2 and H11G11 metagenomic inserts have been deposited at EMBL/GenBank/DDBJ databases under accession nos. FN666674 and FN666673, respectively.

¹To whom correspondence should be addressed. E-mail: joel.dore@jouy.inra.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1000066107/-DCSupplemental.

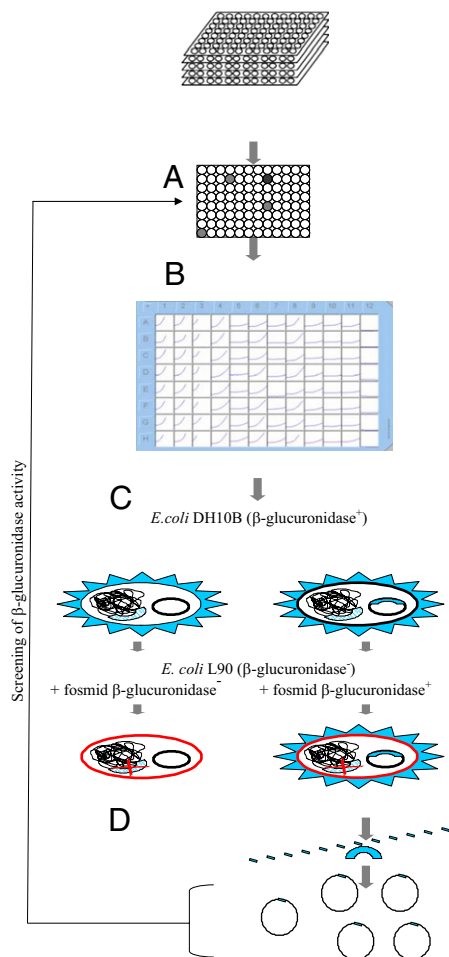


Fig. 1. Screening of BG activity from metagenomic libraries. The metagenomic clones originated from different libraries derived from healthy and Crohn's disease subjects' feces and from a distal ileum biopsy. Metagenomic libraries were screened for over-expressing clones (compared with the *uidA*⁺ *E. coli* receiving strain) using PNP-G as substrate with a two-step method: a visual detection and a quantitative validation. Fosmid activity was further assessed by transfer into a BG-negative strain (*E. coli* L90 $\Delta uidA$). Subclones (3-kbp fragments inserted into pcDNA 2.1 plasmids) were further screened for PNP-G deglucuronidation, and candidate genes were cloned into an expression vector (pRSFDuet-1) to definitively confirm BG activity. Random screening of 4,608 metagenomic clones from fecal and ileal libraries (*E. coli* DH10B as receiving strain) involved the following: (A) a qualitative screen with precocity of β -glucuronidase activity detection that yielded 1.79% of clones over-expressing the activity; (B) a quantitative validation of clones over-expressing the activity (*E. coli* DH10B deprived of insert as control) that yielded 0.73% of positive clones; (C) a selection of fosmids generating β -glucuronidase activity (following transformation of a β -glucuronidase-negative strain) yielding 0.41% of positive clones among all clones tested; and (D) the identification of the genes involved (subcloning, sequencing, and bioinformatic analysis).

uncharacterized inserts (*Materials and Methods*) and a sublibrary of 1,536 clones previously selected for the presence of 16S rRNA genes (14). An initial visual screen was performed and revealed that 1.79% of clones over-expressed BG activity in the uncharacterized library. The quantitative validation step confirmed an over-expression for 40% of them (Fig. 1B), and transfer of the corresponding metagenomic inserts into the $\Delta uidA$ strain allowed the determination of a hit rate of 0.41% of positive fosmids. Including the second screen performed on the 16S rRNA gene sublibrary, 19 positive fosmids were identified of a total of 6,144 metagenomic clones. BG activity in the positive clones ranged

from 0.02 to 1.30 U (units) with, as expected, a lower global level of BG activity when metagenomic inserts were expressed in the $\Delta uidA$ *E. coli* strain (Fig. 2).

A Unique Class of BGs in the Firmicutes Phylum. An initial attempt was made to sequence the potential BGs present in the 19 BG-positive clones on the basis of degenerate primers. The sequence of the primers used derived from the sequences of known prokaryotic BG genes (17, 19). As this approach proved to be unsuccessful, the BG-positive fosmids were subcloned and subsequently sequenced. The 40-kbp inserts were fully sequenced, and six candidate genes were identified as potentially encoding BG activity. The predicted proteins encoded by these genes were the following: β -galactosidase, antimicrobial peptide ABC transport system, serine/threonine kinase, sporulation protein, 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, and aspartyl-tRNA synthetase (Table S1). The corresponding genes were subcloned into an expression plasmid (pRSF plasmid) and further transformed into *E. coli* BL21(DE3) or *E. coli* L90 T7⁺ (*uidA*⁺ and $\Delta uidA$ strains, respectively).

Only the gene encoding the putative β -galactosidase was able to promote BG over-expression in the *uidA*⁺ strain (5.3 versus 0.4 U for the control deprived of insert) and to induce a BG activity in the $\Delta uidA$ strain (4.9 U). This gene originated from a metagenomic insert (H11G11 insert) in which BG activity was in the range of the 19 fosmids (Fig. 2). This insert was from a pool of fecal samples from healthy individuals. Remarkably, the identified glycosidase had low homology with known BGs (<33% identity, Fig. 3) and signatures that diverged from the two consensus sequences of glycosyl hydrolase family 2 (82 and 93% similarity with Prosite patterns PS00719 and PS00608, respectively; Table S2). This glycosidase also had low homology with putative BGs recently identified in the gut dominant species *Roseburia intestinalis* and *Bacteroides ovatus* using degenerate PCR primers (27% and 33% identity, respectively) (11). However, the two glutamate residues involved in the catalytic mechanism of other members of family 2 BGs (21) were conserved in this BG, and a search in the Conserved Domains Database (22) revealed the highest matches to the TIM barrel domain of glycosyl hydrolase family 2 (pfam02836) and BGs (PRK10150) and lower matches to β -galactosidases (COG3250, PRK09525, PRK10340). An additional C-terminal domain compared to known BGs was identified, but no putative function could be proposed (expected value 0.01).

A search of homologous proteins in the National Center for Biotechnology Information (NCBI) nr database revealed a set of 15 H11G11-BG homologs with more than 50% coverage and similarity (Fig. 4). On the basis of amino acid sequence alignment,

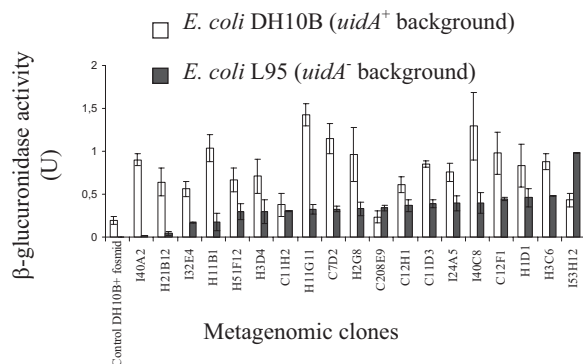


Fig. 2. BG activity of positive metagenomic fosmids in *E. coli* DH10B and *E. coli* L90 $\Delta uidA$. Activity was assayed using PNP-G as substrate and expressed in units (U) of BG activity ($\Delta OD_{405 \text{ nm}} \cdot \text{min}^{-1} \cdot \text{mg}^{-1}$). Means and SD values are determined from experiments performed in quadruplicates for each clone.

H11G11-BG
 gi 12802352 *Lactobacillus gasseri*
 gi 584839 *E. coli*
 gi 34581788 *Ruminococcus gnavus*

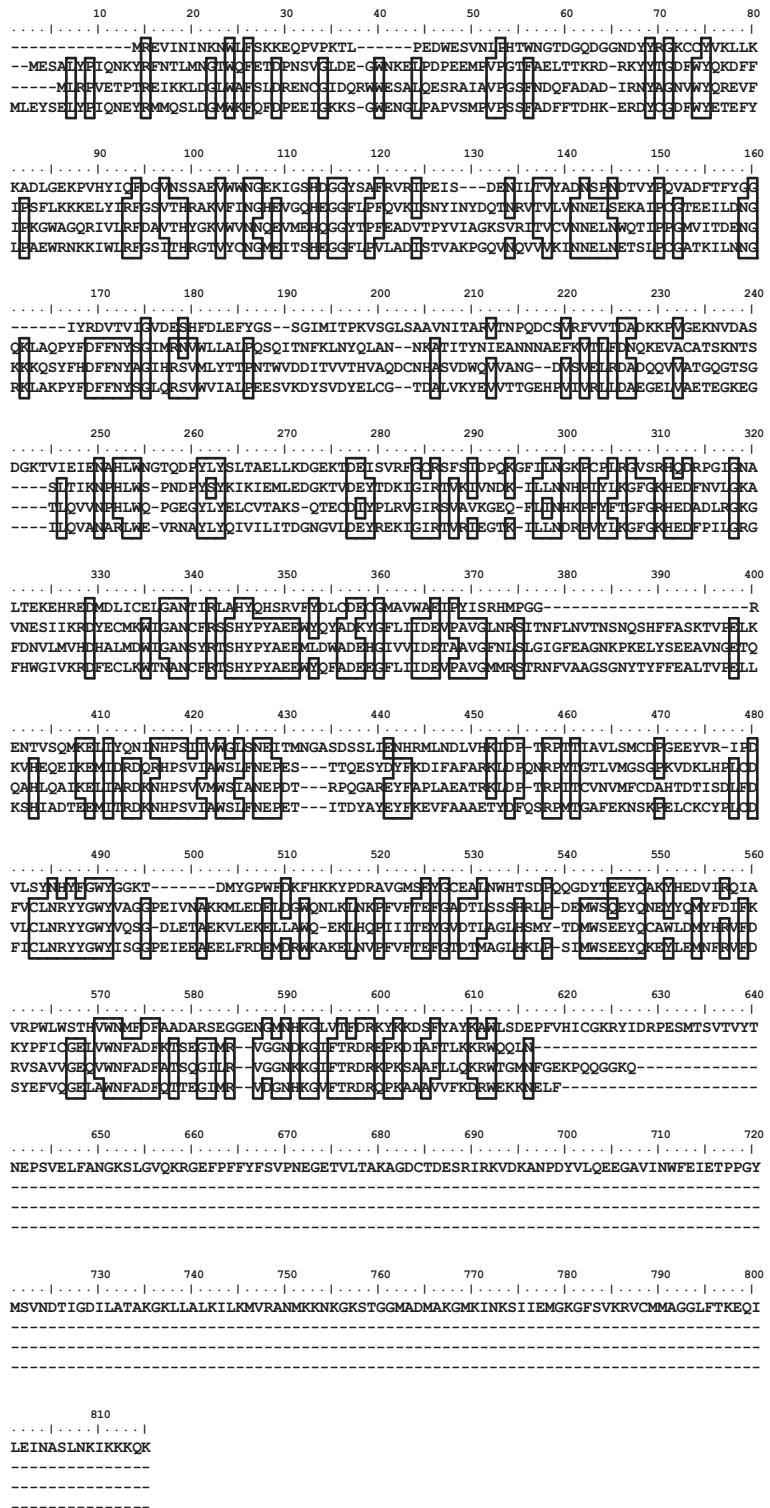


Fig. 3. Amino acid sequence alignment of the unique BG (H11G11-BG) and known BGs from the gut microbiota. Alignment was performed using the ClustalW multiple sequence alignment program. Framed amino acids were conserved in at least three sequences.

all homologs constituted a biologically relevant group (COFFEE score: from 89 to 95) with 38–57% identity with H11G11-BG. These proteins were all identified in prokaryotic genomes under sequencing and were annotated as “uncharacterized protein” except the one from *Paenibacillus* sp. JDR-2, which was annotated as a β -galactosidase. The H11G11-BG homologs were identified only

in gut bacteria from the Firmicutes phylum including *Bacteroides capillosus*, recently reassigned to Gram-positive bacteria (23). The only exception was *Shuttleworthia satelles* DSM 14600, which is a Firmicutes isolated from the oral cavity (24). The closest neighbors were functionally uncharacterized genes in gut Bacteroidales and represented a likely subgroup of Bacteroidetes BGs (Fig. 1B).

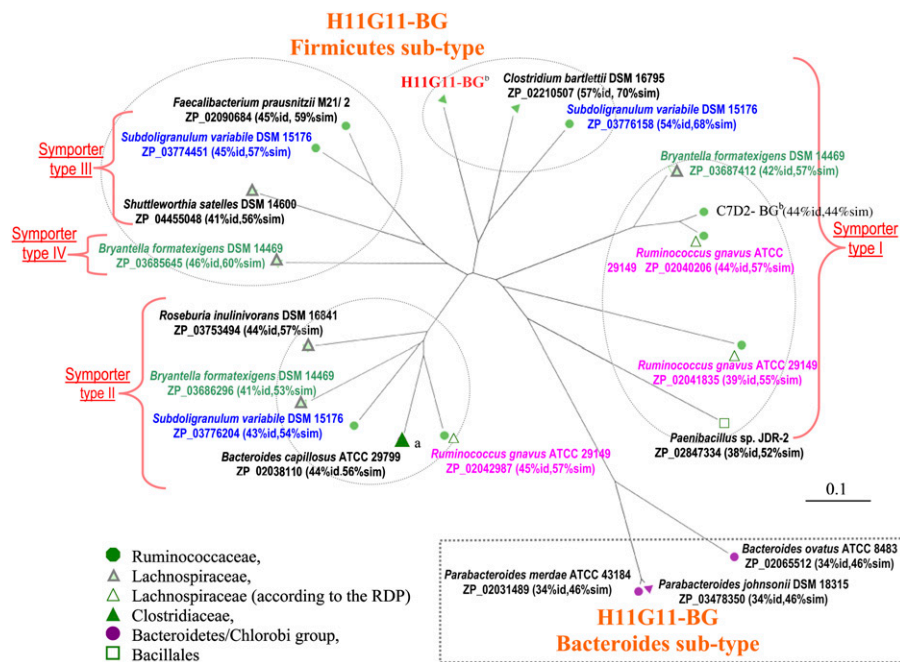


Fig. 4. Identification of a unique class of β -D-glucuronidase, phylogenetic affiliation, and association with potential symporters. Detailed tree view of the H11G11BG-like homologs identified in GenBank genomic databases (Fig. S1). The unique BGs (>50% of both coverage and similarity with H11G11-BG) have been identified only in Firmicutes species, but a potential BG subgroup has also been identified in Bacteroidetes constituting an independent subclass of BGs. Strains known to generate a BG activity are *F. prausnitzii* M21/2, *R. gnavus* ATCC 29149, and *B. capillosus* ATCC 29799. Strains demonstrated as BG positive in this study are *S. variabile* DSM 15176, *B. formatexigens* DSM 14469, *C. bartlettii* DSM 16795, *B. ovatus* ATCC 8483, *P. merdae* ATCC 43184, and *P. johnsonii* DSM 18315. ^aClassification according to Carlier et al. (23). ^bPhylogenetic assignment of metagenomic inserts are as described in *Materials and Methods* (H11G11 insert: 41% *Clostridium* genus. C7D2 insert: 81% *R. gnavus* species.).

The unique Firmicutes BGs had amino acid sequences distantly related to known BGs, even from the ones previously characterized in this phylum (Table S2 and Fig. S1), and had specific uncharacterized and additional domains (Fig. S2). Most of the identified genes were restricted to the Clostridiales order, including members of the Ruminococcaceae, Lachnospiraceae, and Clostridiaceae families (Fig. 4). Three strains were previously known to exhibit BG activity (i.e., *Faecalibacterium prausnitzii* M21/2, *R. gnavus* ATCC 29149, and *Bacteroides capillosus* ATCC 29799). Seven strains of Firmicutes and Bacteroidetes containing H11G11-BG homologs in their genomes were tested for their ability to hydrolyze PNP-G (i.e., *B. ovatus* ATCC 8483, *Bryantella formatexigens* DSM 14469, *Clostridium bartlettii* DSM 16795, *Roseburia inulinivorans* DSM 16841, *Subdoligranulum variabile* DSM 15176, *Parabacteroides johnsonii* DSM 18315, and *Parabacteroides merdae* ATCC 43184). Except for *R. inulinivorans* DSM 16841 (BG activity <0.025 U), all strains tested exhibited high BG activity. The search of known *uidA* homologs in their genomes and in the three Firmicutes strains with known BG activity led to different situations (Table S3). For most of the genomes tested (7/10), the homologs found were more closely related to the H11G11 protein sequence than to *UidA*. The presence of the two classes of BG (*uidA* and H11G11-like) within the same genome remained possible for *F. prausnitzii* M21/2 and *B. formatexigens* DSM 14469, and even a third class was suggested in *B. ovatus* ATCC 8483. Nevertheless, these data suggest that in most cases H11G11-BG was responsible for the observed BG activity. Furthermore, paralogs were found in the genome of *S. variabile* DSM 15176, *B. formatexigens* DSM 14469, and *R. gnavus* ATCC 29149.

Because the two existing Prosite consensus signatures (PS00719 and PS00608) for glycosyl hydrolase family 2 could not include these unique BGs (Table S2), we propose different consensus motifs and a signature for the unique Firmicutes BG group (Fig. 5

A–B) and more broadly for the glycosyl hydrolase family 2 (Fig. 5C). Interestingly, a clear distinction inside the unique BG family was observed. Indeed, the invariant histidine/tyrosine/glutamine motif was conserved in the Bacteroidetes subgroup but was replaced by a histidine/tyrosine/proline motif in the Firmicutes subgroup (Fig. 5C and Fig. S3).

Is Functional Diversity of Unique BGs Based on Associated Symporters?

Insert sequence alignments and genetic environment exploration of 15 of the 17 Firmicutes BGs identified revealed that the conserved genetic environments close to the BG genes were restricted to symporters. Fifteen BGs were associated with putative symporters that, by sequence homology, clustered into four types (Fig. S4). The symporter homologies correlate well to the BG subclasses (Fig. 4). Despite the diversity of the genetic organizations observed (direct vicinity, intermediate sequences, lengths) (Figs. S4 and S5), most of these putative symporters had best hits for conserved domains corresponding to Na^+ /melibiose symporter and related transporters (COG2211) or the sugar transporter Gph (TIGR00792), indicating that these BGs are involved in the metabolism of extracellular glucuronides.

The Unique BG Within the Phylogenetic Core. Among healthy adults the human intestinal microbiota phylogenetic core consists of a restricted set of dominant species (25). To determine the presence of the H11G11-BG-positive species within this phylogenetic core, sequence alignments between the 16S rRNA of H11G11-BG-positive strains and the 66 prevalent operational taxonomic units (OTUs) (25) were performed. Almost half of the BG-positive strains (i.e., *F. prausnitzii* M21/2, *R. inulinivorans* DSM 16841, *S. variabile* DSM 15176) belonged to the phylogenetic core (>99% 16S rRNA identity). Because the phylogenetic core represents only a very few highly prevalent and dominant

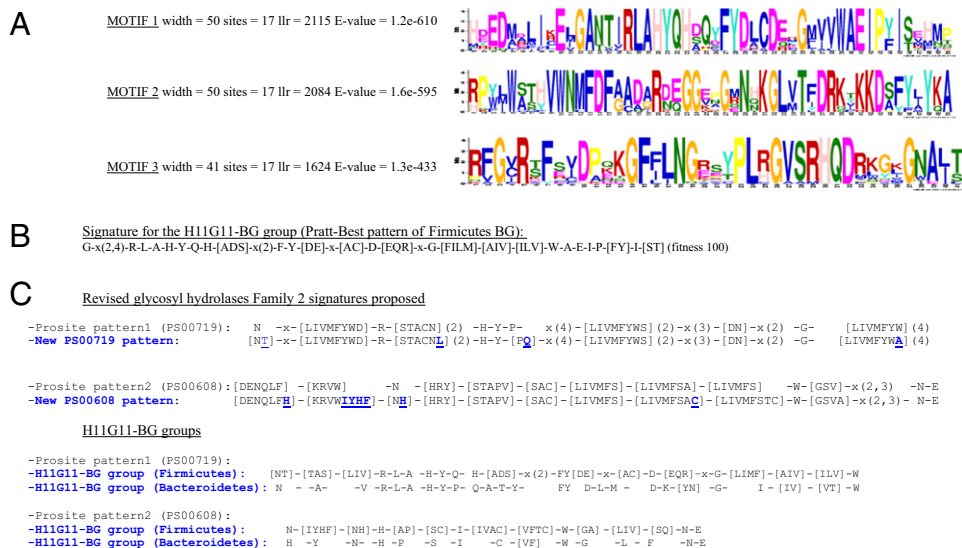


Fig. 5. Conserved motifs within the H11G11-BG group. (A) Conserved motifs within the 17 proteins identified in this study (>50% coverage and similarity with H11G11-BG). (B) Signatures proposed for the unique Firmicutes BG. (C) Previously undescribed glycosyl hydrolase family 2 patterns, including the 17 unique Firmicutes BG identified in this study and the three Bacteroidetes BG. Amino acids added compared to the actual patterns are in blue.

OTUs, this result underlined the importance of this BG function in the human gut.

A BG Highly Prevalent in and Specific to the Human Gut. The frequency of the unique Firmicutes or Bacteroidetes BGs reached an average of $9.5 \pm 3.0 \times 10^{-8}$ hit/bp within the human gut metagenomes (Fig. 6). At least one homolog can therefore be found every 10^7 bp, which is approximately equivalent to 10^4 bacterial genes. Previously known BG genes (*uidA* homologs from Clostridiales, Lactobacillales, and *E. coli* K12) were at least three

times less frequently represented in the human gut metagenomes. Remarkably, the H11G11-BGs subgroup was almost not represented or was totally absent in microbial metagenomes arising from other environments whereas the *uidA* BGs subgroup was more systematically found in soil ecosystems (Waseca County Farm Soil Metagenome ID 13699).

To determine the distribution among the human population, homologs of H11G11-BGs were searched in gut metagenomic databases from different healthy individuals (26) (Fig. 7). Homologs of the unique Firmicutes or Bacteroidetes BGs

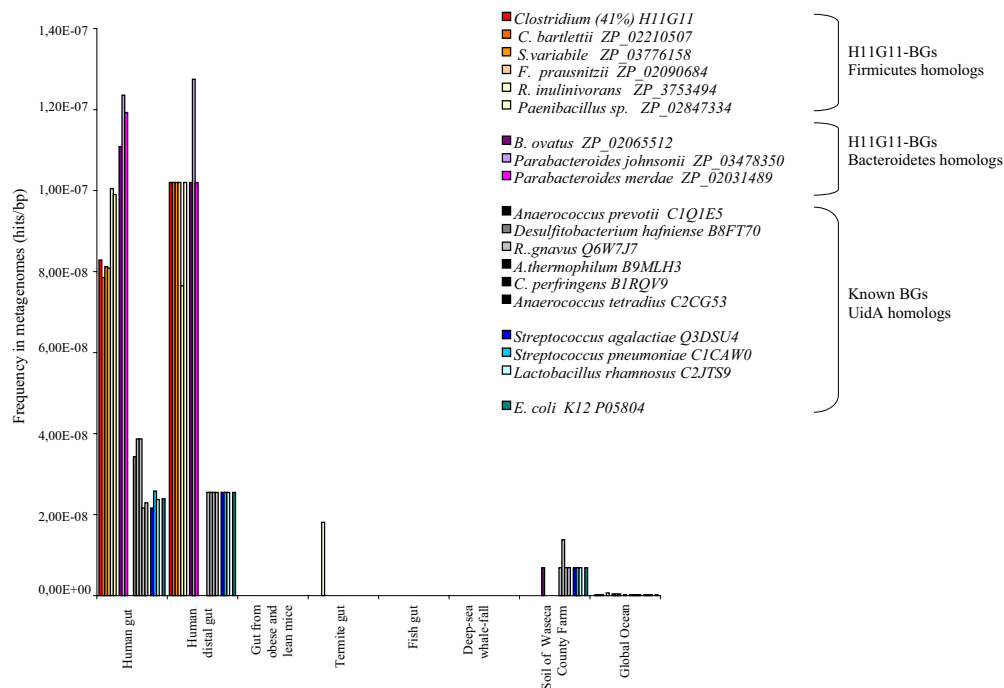


Fig. 6. Frequency of the unique β -glucuronidases and known β -glucuronidases in environmental and human gut metagenomes. Homologs of unique and known BG proteins were searched in NCBI metagenomic project datasets (*Materials and Methods*). The hit threshold was at least 50% similarity with 50% sequence coverage. Results of frequencies were expressed as hits per base pair to correct the different sizes of the metagenomic datasets investigated.

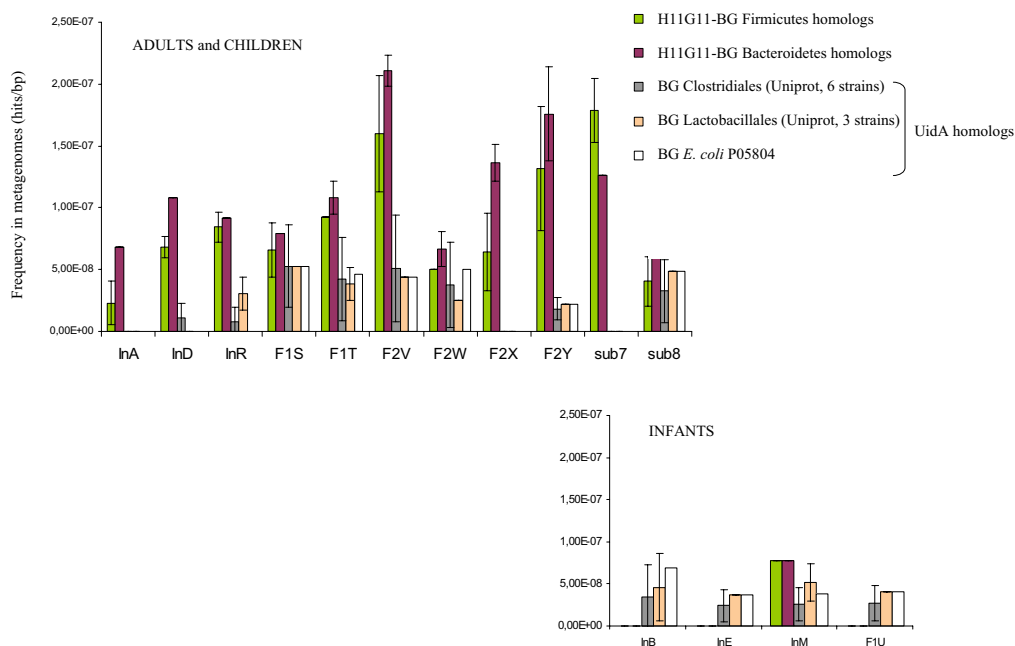


Fig. 7. Distribution of the unique and known BGs among adult, child, and infant gut metagenomes. Homologs of unique and known BG proteins were searched in the two projects Human Gut Metagenome (13 healthy individuals) (ID 28117) and Human Distal Gut Biome (ID16729) as described in Fig. 6. Results are presented per BG group (i.e., Firmicutes-BGs, Bacteroidetes-BGs, UidA homologs) and expressed as means of hits per base pair for each group (details per BG protein, Fig. S6).

were identified in all 11 adults and children but only in one infant among the four explored (Fig. 7 and Fig. S6). In contrast, *uidA* BGs were less abundant among the “adults/children” group (6/11, 8/11, and 7/11 for *E. coli*, Clostridiales, and Lactobacillales BGs, respectively) but were present in all infants. Furthermore, in adults and children, the H11G11-BG frequencies were always higher than those of the *gusA/uidA* BGs ($P \leq 0.01$). They ranged from 2.3×10^{-8} to 2.2×10^{-7} hit/bp, and the frequency in each individual was not correlated to the previously determined level of COG3250 (β -galactosidase/ β -glucuronidase) (26).

Discussion

Using a functional metagenomic approach, we revealed a high frequency of BG activity (0.41%) in the human gut microbiota. In the high range of hits usually obtained during functional screens of glycosyl hydrolases from environmental or gut metagenomic libraries (from 0.00002% to 0.8%) (27), this result is coherent with the particular enrichment of the gut microbiome with β -galactosidase/ β -glucuronidase family (28). This result probably reflects the abundance and the diversity of glucuronide compounds that reach the intestine and the critical role of BGs in several bacterial metabolic pathways. However, this high score could also be due to the secondary deglucuronidation potentialities of proteins that exhibit no obvious sequence homologies with BGs as suggested in this study. We were unable to find evidence of BG activity among these cloned genes. Further investigations of these genes in their genetic context will be required to definitely demonstrate their functionality.

Our functional screen led to the identification of a unique β -D-glucuronidase from the human gut microbiota that possesses several key features, including an additional C terminus domain and distant sequence homologies with known BGs. Mining of genome sequence databases demonstrated that this unique BG is specifically present into several Firmicutes species of the gastrointestinal tract or the oral cavity including *C. bartlettii*, *S. variable*, *B. formatexigens*, *R. gnavus*, *Penibacillus* sp., *R. inulinivorans*, *S. satelles*, and *F. prausnitzii*. This group includes bacterial strains

already reported to exhibit BG activity (i.e., *F. prausnitzii* M21/2, *R. gnavus* ATCC 29149, and *B. capillosus* ATCC 29799) (11, 17, 29) and strains that we demonstrated here as BG positive (i.e., *S. variable* DSM 15176, *B. formatexigens* DSM 14469, *C. bartlettii* DSM 16795). In the same way, BG-positive strains from the Bacteroidetes phylum were evidenced (i.e., *B. ovatus* ATCC 8483, *P. merdae* ATCC 43184, and *P. johnsonii* DSM 18315). Thus this class of BGs, which is present in many dominant Firmicutes and Bacteroidetes species, might represent a major deglucuronidation pathway in the human gut. The unique BGs exhibit conserved domains that diverged from the Prosite consensus sequences (PS00719 and PS00608) of glycosyl hydrolase (GHase) family 2. We thus proposed consensus signatures to include this unique class of BGs. Interestingly, the invariant “HYP” amino acid motif found in the PS00719 signature was maintained in Bacteroidetes BGs but specifically replaced by a “HYQ” motif in Firmicutes BGs. These Firmicutes BGs may therefore have originated from a common ancestor and evolved in the context of the intestinal environment. Considering the importance of GHase family 2, which is the third most abundant family of GHases in the Human gut (27), the functional significance of these sequence specificities deserves further investigation.

Our observations support an important physiological role for these unique BGs. Indeed, in contrast to previously identified BGs (UidA/GusA homologs), these BGs are found with high frequency and high specificity in the human gut microbiome and are well represented and distributed among adults and children but absent in most of the very young infants tested despite the relative known abundance of Firmicutes in their microbiota (26, 30). The absence of correlation between frequencies of H11G11-BG homologs and the galactosidase/glucuronidase COG family within human gut microbiomes demonstrates the necessity of improving this family classification and annotation. Furthermore, among the bacterial species harboring this enzyme family, three Firmicutes species belong to the restricted phylogenetic core of the human intestinal microbiota (25), suggesting an ecological drive that ensures the presence of the activity in spite of micro-

biota variability among humans. This unique class of BGs might therefore play a critical role in the establishment of a definitive and stabilized adult microbiota (31). Results also revealed duplication of unique BG-encoding genes in several strains and association with diverse permeases. These organizations are biologically highly relevant because they are important in generating genetic variability and in facilitating adaptive evolution and functional diversity. Interestingly, most of the species harboring the unique BG were from the *C. leptum* group (cluster IV) or Lachnospiraceae (cluster XIVa), both subject to microbiota disturbances in the physiopathological contexts of Crohn's disease (12, 32–34), ulcerative colitis (35), familial Mediterranean fever (36), and intestinal carcinogenesis in Apc (Min) mice (16). Recent studies revealed a functional link between xenobiotic neutralization, steroid metabolism, and inflammation via mutual repression of xenobiotic and hormone nuclear receptors and the NF- κ B-signalling pathways in the intestine (37–39). Determination of potential xenobiotics or hormone substrates for BGs may be critical to determining their physiological function(s) and potential contribution to the modulation of eukaryotic regulatory networks.

In conclusion, the functional metagenomic approach reported here allows the identification of a unique BG type, which is predominant in the gut microbiota of human adults and children. This BG family, identified in several Firmicutes, may be part of the functional core of the human microbiome recently proposed for the GIT ecosystem (40) and may represent an adaptive evolution. Further investigations should address the impact of functional BGs in individuals' gut physiology and their involvement in the crosstalk with host cells, especially with respect to immune homeostasis.

Materials and Methods

Metagenomic Libraries. Fosmid metagenomic libraries were constructed as previously described using the EpiFOS Library production kit (Epicentre Technologies) and *E. coli* DH10B as a host (14). Three fosmid metagenomic libraries were used in this study; namely the "Healthy," "Crohn," and "Ileum" libraries. The first two libraries were previously constructed from bacterial DNA extracted from pools of feces of six healthy individuals and six inflammatory bowel disease patients, respectively (14), and the third from a colorectal cancer patient biopsy obtained from the healthy distal part of the ileum. From these we used a set of 4,600 clones that contained random genomic bacterial DNA and a set of 1,536 clones selected for the presence of 16S rRNA genes by DNA hybridization (14).

Screening of BG-Positive Fosmid Clones. Because the *E. coli* strain used to build the metagenomic libraries (*E. coli* DH10B) constitutively expressed a β -glucuronidase encoded by the *uidA* gene, an initial visual screening was performed to detect metagenomic clones over-expressing BG activity. Metagenomic clones were grown overnight in 2YT medium with chloramphenicol ($12.5 \mu\text{g} \cdot \text{ml}^{-1}$), and the activity was measured on permeabilized cells after toluene treatment (10%, 10 min at 4 °C). BG activity was measured by adding 2 mM of PNP-G (Sigma-Aldrich) (41). Positive clones were submitted to a second quantitative screen using a kinetic measurement with *E. coli* DH10B transformed with an empty fosmid as control. Eight replicates were performed for each clone. Activity and bacterial density (*E. coli* dry weight) were estimated by measuring $\Delta\text{OD}_{405\text{nm}}$ and $\text{OD}_{600\text{nm}}$, respectively. The BG activity was expressed in units of $\Delta\text{OD}_{405\text{nm}} \cdot \text{min}^{-1} \cdot \text{mg}^{-1}$ of dry weight.

As *E. coli* DH10B possesses a β -glucuronidase activity encoded by the *uidA* gene, the *uidA E. coli* L90 strain was used to unambiguously identify fosmids expressing β -glucuronidase activity. This strain was derived from the TG1 *uidA::Kan'* strain that was previously constructed by gene homologous recombination (3). The kanamycin resistance cassette was removed using FLP recombinase borne by the pCP20 plasmid. Purified fosmids (NucleoSpin 96 Flash, extraction kit, Marchery Nagel) were then electrotransferred (25 μF , 200 Ω , and 2.5 Kv) in the *E. coli* L90 strain. Transformants were grown on LB agar containing the appropriate antibiotic with 5-bromo-4-chloro-3-indolyl- β -D-glucuronide (25 $\mu\text{g} \cdot \text{ml}^{-1}$) (X-gluc, VWR) to reveal BG activity.

Fosmids Subcloning and Sequence Analysis. Metagenomic fosmids were purified using the NucleoSpin 96 Flash extraction kit (Macherey-Nagel). Insert size was

estimated by HindIII digestion followed by pulsed-field gel electrophoresis. Fosmid DNA was subjected to mechanical shearing, and 3-kbp fragments were inserted into the high-copy plasmid pCDNA2.1. A total of 384 subclones originating from each metagenomic BG-positive clone were screened for BG activity, and positive subclones were further sequenced (MWG Biotech). A homolog search was made using the Blastp algorithm available on the NCBI web site (<http://www.ncbi.nlm.nih.gov/BLAST>). Domain architectures, potential functions, and associated patterns were analyzed using the InterProScan (<http://www.ebi.ac.uk/InterProScan/>), MyHits (<http://www.isb-sib.ch/>), PROSITE (<http://www.expasy.org/prosite/>), SMART (<http://smart.embl-heidelberg.de/>), PFAM (<http://www.sanger.ac.uk/Software/Pfam/>), Motif Scan (http://myhits.isb-sib.ch/cgi-bin/motif_scan), or SUPERFAMILY (<http://supfam.org/SUPERFAMILY/>) softwares.

β -Glucuronidase Candidates, Cloning, and Expression. Genes encoding putative BG were amplified by PCR with metagenomic fosmid DNA as templates and appropriate primers (Table S4). PCR products were purified by agarose gel electrophoresis and the Wizard SV Gel and PCR clean up system kit (Promega). PCR products were subcloned into pGEM-T plasmid (Promega) and fully sequenced. pGEM-T plasmids with exact insert sequence were digested by BamHI and PstI restriction enzymes (except for the putative serine/threonine protein kinase cloned with the EcoRI and PstI restriction site). Inserts were further purified by agarose gel electrophoresis and the Wizard SV Gel and PCR clean up system kit (Promega). Purified inserts were ligated into pRFDuet-1 plasmid (Novagen) previously digested with the same restriction enzymes. All cloned genes were sequenced to ensure that no errors were introduced. The plasmids obtained were transformed into both *E. coli* BL21(DE3) (Stratagene) and the L90 T7+ *E. coli* strain. The L90 T7+ *E. coli* strain was obtained by transforming the L90 *uidA E. coli* strain with the pAR1219 plasmid, which expresses the T7 polymerase under the control of an isopropyl 1-thio- β -D-galactopyranoside (IPTG)-inducible promoter. The resulting clones were analyzed for PNP-G deglucuronidation as described above, after gene expression induction with IPTG.

Sequence Analysis of the Unique BG Group and Associated Symporters.

Homologs of H11G11-BG were searched on May 2009 by Blastp method (limit expect *E*-value $\leq 2e-45$) on the NCBI nr database, including all non-redundant GenBank coding sequences. Radial tree of Blastp results was computed by the fast minimum evolution method (threshold of maximum sequence difference: 0.85). Definitive presentation and phylogenetic symbols were realized using Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 after conversion in Newick file format. Proteins showing at least 50% length coverage and 50% similarity of aligned sequences were retained to constitute the group of H11G11-BG homologs. Sequence similarities with the current conserved patterns of the glycosyl hydrolases family 2 (BG and galactosidases, PROSITE documentation PDOC00531, <http://www.expasy.org/prosite/>) were determined using PATTINPROT (<http://npsa-pbil.ibcp.fr/>). Conserved motifs were determined by using the motif-based sequence analysis tool MEME (http://meme.nbcr.net/meme4_1/). The patterns proposed were obtained by using the discover patterns tool PRATT (<http://www.expasy.ch/tools/pratt/>). Comparative analysis of codon usage was performed with known Firmicutes BGs and H11G11-like BG proteins identified by using the Blastp software.

The genetic environments of the H11G11 and C7D2 BGs were analyzed using GeneMark (Version 2.4) for Prokaryotes (http://opal.biology.gatech.edu/GeneMark/gmhmm2_prok.cgi). The different types of putative symporters and their identities/similarities within types were determined after homology search and ClustalW multiple alignment.

Phylogenetic Assignment. The phyla and family assignments of H11G11-BG homologs from genomic databases were from the taxonomy reports automatically generated by NCBI (completed with the classification of *R. gnavus* ATCC 29149 using the Ribosomal Database Project).

The C7D2 insert was isolated from the 16S rDNA sublibrary and thus harbored a 16S rRNA gene (NCBI accession no. AY850499) allowing phylogenetic assignment by the Ribosomal Database Project classifier (<http://rdp.cme.msu.edu/>). However, for C7D2 and H11G11 inserts, potential ORFs were determined using MetaGene (<http://metagene.cb.k.u-tokyo.ac.jp/>), and sequences were further phylogenetically assigned using the best blast hits (Blastn or Blastx software) on NCBI databases. The final phylogenetic assignment was expressed in percentage of insert length coverage.

Phylogenetic relationships with the human gut bacterial phylogenetic core (25) were analyzed in retrieving the 16S rRNA gene sequences of the H11G11-BG-containing strains and analyzing them with the RapidOTU pipeline (25) (0.02% identity cut-off, Kimura-2 parameter distance method). The percentage of identity between the 16S rRNA sequences of BG-positive strains

and their closest phylogenetic core homologs was determined using pairwise alignment with Jalview 2.4.0b2.

BG Activity Assay in Gut Bacteria. Strains encoding a H11G11 homolog in their genomes were grown for 24 h at 37 °C under anaerobic conditions using the Hungate method in appropriate media. The Cooked Meat Medium (Difco) supplemented with yeast extract (5 g/L, Sigma), KH₂PO₄ (5 g/L, Sigma), and cysteine (0.5 g/L, Sigma) was used for *C. bartlettii* DSM 16795. *B. ovatus* ATCC 8483 and *R. gnavus* ATCC 29149 were grown in LYHBH medium [brain–heart infusion medium with 0.5% yeast extract (Difco) and 5 mg/L hemin] supplemented with cysteine (0.25 mg/mL; Sigma). The same medium was supplemented with cellobiose (0.5 mg/mL; Sigma–Aldrich) and maltose (0.5 mg/mL; Sigma) for *S. variabile* DSM 15176 and *F. prausnitzii* M21/2. Soluble starch (0.5 mg/mL; Sigma) was added to this medium to grow *P. merdae* ATCC 43184 and *P. johnsonii* DSM 18315, and 0.4% of clarified rumen juice (42) was also necessary to grow *B. formatexigens* DSM 14469 and *R. inulinivorans* DSM 16841. PNP-G activity was tested as described above for *E. coli*.

- Rod TO, Midtvedt T (1977) Origin of intestinal beta-glucuronidase in germfree, monocontaminated and conventional rats. *Acta Pathol Microbiol Scand [B]* 85:271–276.
- Knasmuller S, et al. (2001) Impact of bacteria in dairy products and of the intestinal microflora on the genotoxic and carcinogenic effects of heterocyclic aromatic amines. *Mutat Res* 480–481:129–138.
- Humboldt C, et al. (2007) Beta-glucuronidase in human intestinal microbiota is necessary for the colonic genotoxicity of the food-borne carcinogen 2-amino-3-methylimidazo[4,5-f]quinoline in rats. *Carcinogenesis* 28:2419–2425.
- de Herder WW, Hazenberg MP, Pennock-Schroder AM, Hennemann G, Visser TJ (1986) Rapid and bacteria-dependent in vitro hydrolysis of iodothyronine-conjugates by intestinal contents of humans and rats. *Med Biol* 64:31–35.
- Graef V, Furuya E, Nishikaze O (1977) Hydrolysis of steroid glucuronides with beta-glucuronidase preparations from bovine liver, *Helix pomatia*, and *E. coli*. *Clin Chem* 23:532–535.
- Kim DH, et al. (1998) Intestinal bacterial metabolism of flavonoids and its relation to some biological activities. *Arch Pharm Res* 21:17–23.
- Kim DH, et al. (2000) Biotransformation of glycyrrhizin by human intestinal bacteria and its relation to biological activities. *Arch Pharm Res* 23:172–177.
- Schmelz EM, et al. (1999) Ceramide-beta-D-glucuronide: Synthesis, digestion, and suppression of early markers of colon carcinogenesis. *Cancer Res* 59:5768–5772.
- Carrette O, et al. (1995) Bacterial enzymes used for colon-specific drug delivery are decreased in active Crohn's disease. *Dig Dis Sci* 40:2641–2646.
- Geier MS, Butler RN, Howarth GS (2006) Probiotics, prebiotics and synbiotics: A role in chemoprevention for colorectal cancer? *Cancer Biol Ther* 5:1265–1269.
- Dabek M, McCrae SI, Stevens VJ, Duncan SH, Louis P (2008) Distribution of beta-glucuronidase and beta-glucuronidase activity and of beta-glucuronidase gene *gus* in human colonic bacteria. *FEMS Microbiol Ecol* 66:487–495.
- Baumgart M, et al. (2007) Culture independent analysis of ileal mucosa reveals a selective increase in invasive *Escherichia coli* of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. *ISME J* 1:403–418.
- Frank DN, et al. (2007) Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci USA* 104:13780–13785.
- Manichanh C, et al. (2006) Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 55:205–211.
- Scanlan PD, Shanahan F, O'Mahony C, Marchesi JR (2006) Culture-independent analyses of temporal variation of the dominant fecal microbiota and targeted bacterial subgroups in Crohn's disease. *J Clin Microbiol* 44:3980–3988.
- Mai V, Colbert LH, Perkins SN, Schatzkin A, Hursting SD (2007) Intestinal microbiota: A potential diet-responsive prevention target in ApcMin mice. *Mol Carcinog* 46:42–48.
- Beaud D, Tailliez P, Anba-Mondoloni J (2005) Genetic characterization of the beta-glucuronidase enzyme from a human intestinal bacterium, *Ruminococcus gnavus*. *Microbiology* 151:2323–2330.
- Roy D, Ward P (1992) Rapid detection of *Bifidobacterium dentium* by enzymatic hydrolysis of beta-glucuronide substrates. *J Food Prot* 55:291–295.
- Russell WM, Klaenhammer TR (2001) Identification and cloning of *gusA*, encoding a new beta-glucuronidase from *Lactobacillus gasserii* ADH. *Appl Environ Microbiol* 67:1253–1261.
- Blanco C (1987) Transcriptional and translational signals of the *uidA* gene in *Escherichia coli* K12. *Mol Gen Genet* 208:490–498.
- Salleh HM, et al. (2006) Cloning and characterization of *Thermotoga maritima* beta-glucuronidase. *Carbohydr Res* 341:49–59.

Comparative Metagenomic Analysis. Comparative metagenomic analysis was performed on May 2009 using the tBlastn software to search for H11G11-BG homologous proteins in the following NCBI metagenomic project datasets (<http://metasystems.riken.jp/metabiome/metagenome.php>): Human Gut Metagenome (13 healthy individuals) (ID 28117), Human Distal Gut Biome (ID16729), Combined Gut Metagenome from Obese and Lean Mice (ID17401), Termite Gut Metagenome (ID19107), Fish Metagenome (ID28955), Whale Fall Metagenome (ID13700), Microbial Mat Isolate (ID29795), Waseca County Farm Soil Metagenome (ID13699), and Global Ocean Sampling (ID19733). The same approach was performed with representative known or putative BGs (UniProtKB databases, <http://www.uniprot.org/>, June 2009). The similarity hit threshold was at least 50% similarity with 50% sequence coverage. To compare frequencies between databases, results are expressed as hits per base-pair unit.

ACKNOWLEDGMENTS. We thank Dr. J. Tap for helpful discussions, M. Serezat for technical assistance, and C. Bridonneau for advice about gut bacteria cultivation. This research was supported by the French Ministry of Research under project *GenoTube*, and insert sequences were performed by the French National Sequencing Center (Genoscope, Evry, France).

- Marchler-Bauer A, Bryant SH (2004) CD-Search: Protein domain annotations on the fly. *Nucleic Acids Res* 32(Web Server issue):W327–W331.
- Carlier JP, Bedora-Faure M, K'Ouas G, Alauzet C, Mory F (2010) Proposal to unify *Clostridium orbiscindens* (Winter et al. 1991) and *Eubacterium plautii* (Seguin 1928; Hofstad and Aasjord 1982) with description of *Flavonifractor plautii* gen. nov., comb. nov. and reassignment of *Bacteroides capillosus* to *Pseudoflavonifractor capillosus* gen. nov., comb. nov. *Int J Syst Evol Microbiol* 60:585–590.
- Downes J, Munson MA, Radford DR, Spratt DA, Wade WG (2002) *Shuttleworthia satelles* gen. nov., sp. nov., isolated from the human oral cavity. *Int J Syst Evol Microbiol* 52:1469–1475.
- Tap J, et al. (2009) Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol* 11:2574–2584.
- Kurokawa K, et al. (2007) Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* 14:169–181.
- Li LL, McCorkle SR, Monchy S, Taghavi S, van der Lelie D (2009) Bioprospecting metagenomes: Glycosyl hydrolases for converting biomass. *Biotechnol Biofuels* 2:10.
- Gill SR, et al. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312:1355–1359.
- Knivett VA, Shah HN, McKee AS, Hardie JM (1983) Numerical taxonomy of some non-saccharolytic and saccharolytic *Bacteroides* species. *J Appl Bacteriol* 55:71–80.
- Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO (2007) Development of the human infant intestinal microbiota. *PLoS Biol* 5:e177.
- Paliy O, Kenche H, Abernathy F, Michail S (2009) High-throughput quantitative analysis of the human intestinal microbiota with a phylogenetic microarray. *Appl Environ Microbiol* 75:3572–3579.
- Jansson J, et al. (2009) Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS One* 4:e6386.
- Sokol H, et al. (2009) Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. *Inflamm Bowel Dis* 15:1183–1189.
- Willing B, et al. (2009) Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. *Inflamm Bowel Dis* 15:653–660.
- Zhang M, et al. (2007) Structural shifts of mucosa-associated lactobacilli and *Clostridium leptum* subgroup in patients with ulcerative colitis. *J Clin Microbiol* 45:496–500.
- Khachatryan ZA, et al. (2008) Predominant role of host genetics in controlling the composition of gut microbiota. *PLoS One* 3:e3064.
- Wahli W (2008) A gut feeling of the PXR, PPAR and NF-kappaB connection. *J Intern Med* 263:613–619.
- Zhou C, et al. (2006) Mutual repression between steroid and xenobiotic receptor and NF-kappaB signaling pathways links xenobiotic metabolism and inflammation. *J Clin Invest* 116:2280–2289.
- Zhou C, Verma S, Blumberg B (2009) The steroid and xenobiotic receptor (SXR), beyond xenobiotic metabolism. *Nucl Recept Signal* 7:e001.
- Turnbaugh PJ, et al. (2009) A core gut microbiome in obese and lean twins. *Nature* 457:480–484.
- Bardonnet N, Blanco C (1992) *uidA*-antibiotic-resistance cassettes for insertion mutagenesis, gene fusions and genetic constructions. *FEMS Microbiol Lett* 72:243–247.
- Leedle JA, Hespell RB (1980) Differential carbohydrate media and anaerobic replica plating techniques in delineating carbohydrate-utilizing subgroups in rumen bacterial populations. *Appl Environ Microbiol* 39:709–719.