Molecular Microbiology (2016) 102(4), 579–592 ■
doi:10.1111/mmi.13480
First published online 14 September 2016

Functional characterization of a gene locus from an uncultured gut *Bacteroides* conferring xylo-oligosaccharides utilization to *Escherichia coli*

Alexandra S. Tauzin,[1,2] Elisabeth Laville,[1] Yao Xiao,[3] Sébastien Nouaille,[1] Pascal Le Bourgeois,[1] Stéphanie Heux,[1] Jean-Charles Portais,[1] Pierre Monsan,[2] Eric C. Martens,[3] Gabrielle Potocki-Veronese[1] and Florence Bordes[1]*

[1]*LISBP, CNRS, INRA, INSAT, Université de Toulouse, Toulouse, France.*
[2]*TWB, INRA, Ramonville Saint-Agne, France.*
[3]*Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI, USA.*

Summary

In prominent gut *Bacteroides* strains, sophisticated strategies have been evolved to achieve the complete degradation of dietary polysaccharides such as xylan, which is one of the major components of the plant cell wall. Polysaccharide Utilization Loci (PULs) consist of gene clusters encoding different proteins with a vast arsenal of functions, including carbohydrate binding, transport and hydrolysis. Transport is often attributed to TonB-dependent transporters, although major facilitator superfamily (MFS) transporters have also been identified in some PULs. However, until now, few of these transporters have been biochemically characterized. Here, we targeted a PUL-like system from an uncultivated *Bacteroides* species that is highly prevalent in the human gut metagenome. It encodes three glycoside-hydrolases specific for xylo-oligosaccharides, a SusC/SusD tandem homolog and a MFS transporter. We combined PUL rational engineering, metabolic and transcriptional analysis in *Escherichia coli* to functionally characterize this genomic locus. We demonstrated that the SusC and the MFS transporters are specific for internalization of linear xylo-oligosaccharides of polymerization degree up to 3 and 4 respectively. These results were strengthened by the study of growth dynamics and transcriptional analyses in response to XOS induction of the PUL in the native strain, *Bacteroides vulgatus*.

Accepted 8 August, 2016. *For correspondence. E-mail bordes@insa-toulouse.fr; Tel. +33 5 61 55 94 39; Fax +33 5 61 55 94 00.

Introduction

Due to scarcity of genes coding for complex polysaccharide-degrading enzymes (the so-called Carbohydrate Active enZymes, or CAZymes), humans depend on the symbiotic microorganisms within their digestive tract to breakdown dietary glycans that are recalcitrant to digestion in the upper parts of the gut. These glycans are mainly plant cell wall components, consisting of a cellulose scaffold cross-linked with hemicelluloses and pectins. The structural complexity and diversity make the complete degradation of these glycans a complex issue.

To face this complexity, bacteria from various genera have developed sophisticated systems involving batteries of CAZymes and carbohydrate transporters, encoded by genes co-localized on specific loci. In *Bacteroides* strains, which are the most prominent glycan degraders in the intestine, Polysaccharide Utilization Loci (PULs) encode all the proteins involved in sensing, binding, transport and hydrolysis, that are required to achieve the complete breakdown and uptake of glycans (Hehemann *et al.*, 2010; Larsbrink *et al.*, 2014; Rogowski *et al.*, 2015; Cuskin *et al.*, 2015). In the archetypal PUL system specific to starch utilization (SUS) from *Bacteroides thetaiotaomicron*, a TonB-dependent transporter (SusC) works in synergy with binding proteins (SusD, SusE and SusF) to internalize the oligosaccharides derived from the hydrolysis of starch by the cell surface α-amylase (SusG). The TonB-dependent transporter in complex with ExbB and ExbD proteins allows the transport of macromolecules across the outer membrane of Gram negative bacteria via energy derived from the proton motive force (for review see ref. Ferguson and Deisenhofer, 2002; Schauer *et al.*, 2008).

Xylan is a major component of plant cell walls and is highly abundant in cereal-derived human foods. In the human gut, most of the xylanolytic bacteria were identified among the *Bacteroides* genus (Dodd *et al*., 2011; Martens *et al*., 2011). To date, only two xylan PULs from *Bacteroides ovatus* (PUL-XylS and PUL-XylL) were thoroughly studied but their characterization focused exclusively on glycoside hydrolases and carbohydrate binding proteins (Rogowski *et al*., 2015). Interestingly, the PUL-XylL exhibits two SusC-like transporters while the PUL-XylS possesses a SusC and a major facilitator superfamily (MFS) transporter. MFS, which is located in the inner membrane in *Bacteroides* species, is a secondary transporter of small molecules, including carbohydrates, in response to electrochemical potentials (for review see ref. Yan, 2015). Few PULs have been identified harbouring both a SusC/D transport system and a MFS transporter, such as the glycosaminoglycan and the N-glycan PULs of *Bacteroides thetaiotaomicron* or the sialic acid cluster of *Bacteroides fragilis* (Martens *et al*., 2008; Stafford *et al*., 2012; Phansopa *et al*., 2014). Nevertheless, the specificity of each of these proteins in carbohydrate harvesting has not been deeply studied.

The characterization of transporters in native strains indeed faces several bottlenecks (i) the deletion of targeted gene might not be sufficient to confirm its functionality due to functional redundancy insured by other native proteins; (ii) in a PUL-like system, the activation of the system requires sensing of a specific glycan in periplasm, which is usually different from the internalized oligosaccharides obtained by extracellular hydrolysis, and (iii) despite the huge efforts dedicated to bacterial genetics, genome engineering remains a challenge for numerous species, and is even impossible for uncultivated ones.

During the last decade, significant efforts have been put into functional genomics and metagenomics (Turnbaugh *et al*., 2007; Hess *et al*., 2011; Nielsen *et al*., 2014) in order to elucidate the main functionalities of microbiomes. Functional metagenomic is a powerful tool to decipher the diversity of functions present within the uncultured gut bacteria fraction, which represents up to 70% of the human gut microbiota. From activity-based screening approaches emerge large metagenomic DNA fragments (25–40 kb) containing full multigenic clusters such as PULs that contain putative transporters (Tasse *et al*., 2010). However, the diversity of carbohydrate transporter specificities remains largely under-explored.

In this context, we decided to extend the characterization of carbohydrate transporters to those harboured by uncultured gut bacteria. We thus studied the recombinant expression and functional capabilities of a PUL issued from a highly prevalent uncultured *Bacteroides* strain, involved in the metabolism of xylo-oligosaccharides (XOS). This assembly of genes was identified from a fecal metagenomic library screened for prebiotic degradation (Cecchini *et al*., 2013). Here, by combining a transcriptomic analysis of each gene of the metagenomic insert in *Escherichia coli* with the biochemical characterization of glycan hydrolysis and transport specificities, we showed that this highly conserved PUL-like system possesses a complete functional arsenal for XOS metabolism in the *E. coli* recombinant host. It is composed of two transporters, one of them working in synergy with a carbohydrate binding protein and a battery of PUL-associated glycoside hydrolases (GH) allowing XOS hydrolysis into xylose, which is further metabolized by the cells. These results were strengthened by the study of growth dynamics and transcriptional analyses in response to XOS induction of the PUL in the native strain, *Bacteroides vulgatus*.

Results and discussion

Sequence analysis reveals a PUL involved in XOS utilization

Previously, Cecchini *et al*. (2013) identified the metagenomic clone F5, which was able to hydrolyze XOS up to a degree of polymerization of 6 (DP6) (Cecchini *et al*., 2013). The metagenomic DNA insert, sizing 39093 bp, was assigned to *Bacteroides vulgatus* strain. Over 93% of the F5 sequence showed 99% sequence identity with a part of the *B. vulgatus* ATCC 8482 genome (Fig. 1A). Its functional annotation revealed a PUL containing genes encoding a truncated glycoside hydrolase of family 43 (GH43_t), a hybrid two-component system (HTCS), a TonB-dependent porin (SusC), a binding protein (SusD), two members of the glycoside hydrolase family 43 (GH43A and GH43B), a member of the glycoside hydrolase family 10 (GH10), a MFS transporter and a member of the glycoside hydrolase family 16 (GH16) (Cecchini *et al*., 2013).

By comparison with the genome of *B. vulgatus* ATCC 8482, this metagenomic locus was interrupted within the *gh43* gene upstream of the HTCS. This interruption might imply that at least the *gh43* gene and likely the other PUL genes upstream have been truncated during the library construction process.

Within fully and partially sequenced genomes of gut bacteria (Joint Genome Institute, Markowitz *et al*., 2012), such PUL organization is closely conserved throughout *B. vulgatus* strains and their phylogenetic related neighbors such as *B. dorei*, *B. sartorii* and *B. massiliensis* (Fig. 1A). While the PULs from *B. sartorii* and *B. massiliensis* are not listed in the Polysaccharide-Utilization Loci Database (PULDB), the PULs from *B. vulgatus* and *B. dorei* are listed as predicted with some length differences (Terrapon *et al*., 2015). The closest

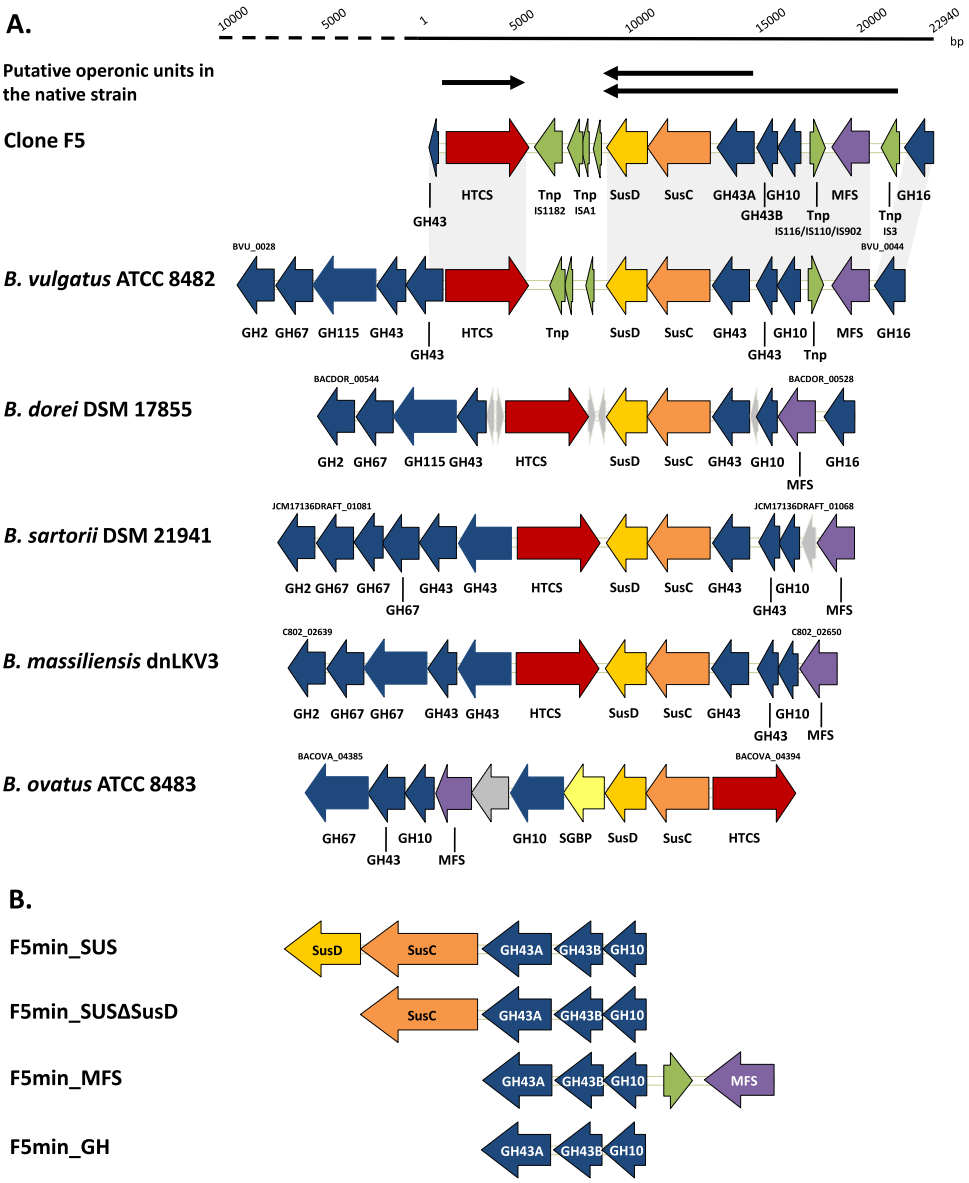

Fig. 1. Representation of the PUL-like system.
A. Organization of the XOS utilization locus based on selected annotated *Bacteroides* genomes. Genes encoding known and predicted functionalities are colour-coded: glycoside hydrolase (GH) with family number in blue; hybrid two component system (HTCS) in red; SusC in orange; SusD in yellow; surface glycan binding protein (SGBP or SusE-positioned) in light yellow; transporter of the major facilitator superfamily (MFS) in purple; transposase (Tnp) in green and unknown in grey. Synteny (corresponding to 99% identity at the DNA level) between the sequence of clone F5 and the genome locus of *B. vulgatus* ATCC 8482 are shown by grey bars. Black arrows represent putative transcription units in the *Bacteroides* natural host, according to the consensus promoter sequence of *Bacteroides* strain.
B. The reduced constructs of F5 used in the present work.

PUL characterized so far, in terms of functionality, is the small xylan PUL (PUL-XylS) from *B. ovatus* ATCC 8483 (Rogowski *et al.*, 2015). This PUL encodes an HTCS, a tandem SusC/D, a surface glycan binding protein (SGBP), two GH10s, a MFS transporter, a GH43 and a GH67 (Fig. 1A). It is induced by wheat arabinoxylan, glucuronoxylan and linear XOS (Martens *et al.*, 2008; Rogowski *et al.*, 2015). Immediately downstream of the SusD-like, PULs usually encode a SGBP contributing to the additional binding of the substrates (Cameron *et al.*, 2012; Rogowski *et al.*, 2015) which is absent in the F5 PUL.

In addition to the arranged SusC- and D proteins to potentially bind and transport glycans, the clone F5 exhibited a gene encoding a MFS transporter. In *E. coli*, the sialic acid uptake is due to a specific MFS transporter (NanT) (Vimrt and Troy, 1985) while in other bacteria the sialic-acid-targeting PULs display MFS transporters that are sometimes associated with the SusC/D transport system (NanO/U) such as in *Tannerella forsythia* and *B. fragilis* (Roy *et al.*, 2010; Stafford *et al.*, 2012; Phansopa *et al.*, 2014). As introduced above, such an association has also been observed in other *Bacteroides* PULs and was demonstrated as being part of the operon. Examples include the glycosaminoglycan and the N-glycan PULs of *B. thetaiotaomicron* and, more recently, the PUL-XylS of *B. ovatus* (Martens *et al.*, 2008; Rogowski *et al.*, 2015).

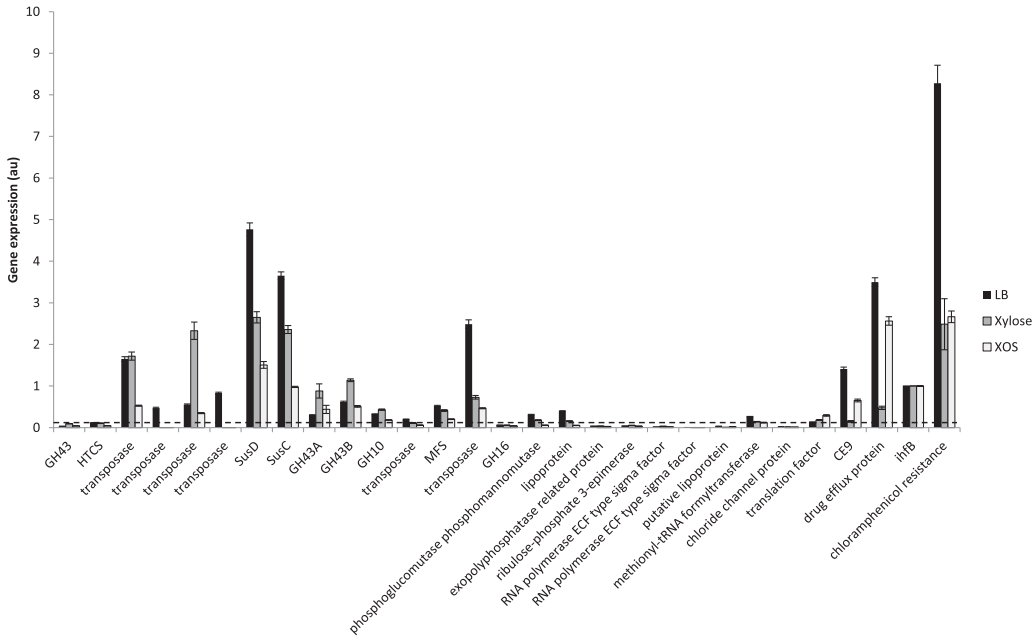

Fig. 2. Gene expression analysis of the clone F5 on LB (black), xylose (grey) or XOS (white) growth conditions. The mean of the biological triplicates is represented with +/− the standard error of the mean. Hybrid two-component system (HTCS), major facilitator superfamily transporter (MFS), integration host factor β-subunit (ihfB). Dashlined represents an arbitrary threshold of expression.

Finally, five transposase sequences are present in the F5 sequence (Fig. 1A). Three are located between the *htcs*- and the s*usC*-like genes, one between the *gh10* and the *mfs* transporter genes and one between the *gh16* and the *mfs* transporter genes.

Metagenomic gene expression in E. coli

We (Tasse *et al.*, 2010) and others previously showed (Ferrer *et al.*, 2005; Wang *et al.*, 2012; Strachan *et al.*, 2014) that phenotype of fosmid/cosmid metagenomic clones is often related to the presence of several genes encoding enzymes with various, and often complementary activities. This is particularly true for clones harbouring PUL-like multigenic systems issued from Bacteroidetes. These clones encode synergistic CAZymes that are able to completely breakdown complex polysaccharidic structures (Tasse *et al.*, 2010). However, functional expression of such metagenomic genes in *E. coli*, which is still the predominantly used host for activity-based metagenomic, has never been experimentally investigated. Here, for the first time, the abilities of *E. coli* system to host and express a heterologous multigenic system that is involved in XOS metabolism from uncultured *Bacteroides* have been explored at the transcriptional level.

To further investigate the level of induction/expression of the 27 genes present on the F5 metagenomic insert, the transcriptional level of each open reading frame (ORF) in *E. coli* has been measured by quantitative RT-PCR (Fig. 2). In the LB medium, among the 27 genes that are present on the metagenomic DNA insert, only 10 were not expressed or expressed at very low level (including the truncated *gh43*, *htcs* and *gh16* encoding genes from the PUL cluster). The 17 others genes were either expressed at significant levels comparable to endogenous *E. coli* housekeeping gene (*ihfB*) or even at level close to the highly expressed fosmidic *cam* gene (chloramphenicol acetyltransferase) used for chloramphenicol selection. The strongest expression was detected for genes encoding SusD, SusC and a drug efflux protein, at a level over three-fold relative to the expression of *ihfB*. The genes coding for GH43A, GH43B, GH10 and MFS were expressed to a level of 1.7 to 3.5-folds lower than *ihfB*. In addition, more than half of the genes contained on the metagenomic insert were transcribed at various expression levels, demonstrating that the quantified gene expressions were due to the recognition of distinct promoter sequences by the recombinant host. A bioinformatics analysis of the full F5 sequence revealed 6 putative *Bacteroides* promoters, of which 3 are within the PUL. In the native *Bacteroides* strain, the F5 PUL could be expressed as an operon from *mfs* or *gh43A* to *susD* and regulated by the *htcs*, which is expressed separately (Fig. 1A). Nevertheless, the *Bacteroides* promoters cannot be responsible for heterologous expression in *E. coli*, since Mastropaolo *et al.* (2009) showed that *Bacteroides* promoters are not recognized by *E. coli* (Mastropaolo *et al.*, 2009). Using

BPROM program which predicts with an accuracy of 80% consensus σ^{70} promoter sequences for *E. coli*, about 184 putative promoters were identified within the metagenomic insert and 34 only within the PUL (Supporting Information Fig. S1). Recently, Lam and Charles (2015) suggested that metagenomic genes, especially those issued from *Bacteroides* species, are spuriously transcribed in *E. coli* thanks to the random presence of *E. coli rpoD*/σ^{70} promoter sequences on metagenomic DNA inserts, that would also be responsible for cloning bias in metagenomic libraries (Lam and Charles, 2015). They counted only around 10 promoter sequences/Mb of metagenomic DNA, but they focused exclusively on the most common consensus sequence TTGACA(-35) and TATAAT(-10). This specific promoter sequence was not detected within the F5 sequence. However, it is noteworthy that σ^{70} promoters vary in their sequence, the absence/presence of the -35 box and the length of the spacer between the -10 and -35 sequences (Shultzaberger *et al.*, 2007; Singh *et al.*, 2011). The BPROM tool, which allows the identification of degenerated promoter sequences, thus seems more pertinent to identify putative σ^{70} promoters, since the predictions are in good adequacy with the present transcriptomic results. In functional metagenomic, the recurrent bottleneck is to access the full potential of metagenomes and different strategies have been developed to overcome this limitation, mostly by improving cloning strategies and the screening host, i.e. *E.coli* (for perspective see ref. Lam *et al.*, 2015). However, our data point out that a significant proportion of metagenomic genes can be efficiently transcribed in *E. coli*, and that spurious transcription would be more advantageous than deleterious for heterologous expression of multiple genes.

Finally, comparison of the expression data obtained during growth on LB medium to results obtained using xylose- and XOS-grown cultures showed that the genes belonging to the PUL were not differentially expressed between all the conditions tested (Fig. 2), suggesting a lack of regulation in *E. coli*. It has been shown in *Bacteroidetes* that the ability to modulate the PUL expression depends on sensor-regulator systems such as the hybrid two-component systems in response to their targeted glycan (Bolam and Koropatkin, 2012). In our system and in all culture media, *htcs* gene was expressed at a very low level in *E. coli*. Thus we postulate that the PUL expression is not regulated either by the presence of XOS nor xylose in *E. coli*.

Functional potential of the hydrolases in E. coli

Previously, we demonstrated that the cell extracts of the clone F5 were able to hydrolyze XOS up to DP 6 (Supporting Information Fig. S1 from ref. Cecchini *et al.*, 2013). To characterize the catabolic potential of the CAZymes contained in the multiproteic system to breakdown hemicellulose, a more detailed screening of hydrolytic activities was carried out on a panel of polysaccharides, oligosaccharides and chromogenic substrates representing the different components of plant cell wall. Xylan and arabinoxylan were both hydrolyzed by the F5 cytoplasmic extracts with a preference for unbranched xylan which was cleaved more efficiently (Table 1). The cytoplasmic extracts have also been tested on synthetic substrates and showed activity on *p*NP-β-D-xylopyranose and *p*NP-α-L-arabinofuranose (Table 1). These detected xylanase, xylosidase and arabinofuranosidase activities are consistent with the known activities of CAZy families GH10 and GH43 identified in the PUL. Conversely, no activity was detected on β-glucan or xyloglucan, likely attributed to the GH16 enzyme, a result consistent with the absence of transcription of the *gh16* gene (Fig. 2). Considering the previous sequence annotation of the transcription units in *Bacteroides*, the gene expression analysis, and the activity results of the clone F5, the GH16 seems unlikely to be part of PUL.

To confirm that the enzymatic activities were due to the GH43A, GH43B and GH10 from the PUL, we generated a reduced construct named F5min_GH containing only the three *gh* genes (Fig. 1B). The enzymatic activities of the F5min_GH cell extracts were similar to those obtained for F5, implying that the enzymes responsible for these activities are encoded by the genes *gh43A*, *gh43B* and *gh10* (Table 1). The high number of genes to

Table 1. Activities of the cell extracts on complex polysaccharides and synthetic substrates.

	Clone	
Substrate	F5	F5min_GH
Complex polysaccharides		
Xylan	176.6 ± 4.2	329.8 ± 14.2
Arabinoxylan	+	n.d.
Arabinan	/	n.d.
Arabinogalactan	/	n.d.
β-glucan	/	n.d.
Xyloglucan	/	n.d.
Synthetic substrates		
*p*NP β-D-Xylopyranose	399.7 ± 10.0	609.3 ± 52.1
*p*NP α-D-Xylopyranose	/	n.d.
*p*NP α-L-Arabinofuranose	239.9 ± 10.2	353.7 ± 28.3
*p*NP β-L-Arabinopyranose	/	n.d.
*p*NP α-L-Arabinopyranose	/	n.d.

Activities were expressed in mU (with 1U = 1 μmol/min) per litre of culture. Mean of three biological replicates.
Abbreviations: + = residual activity after 24 h; / = no activity detected; n.d. = not determined.

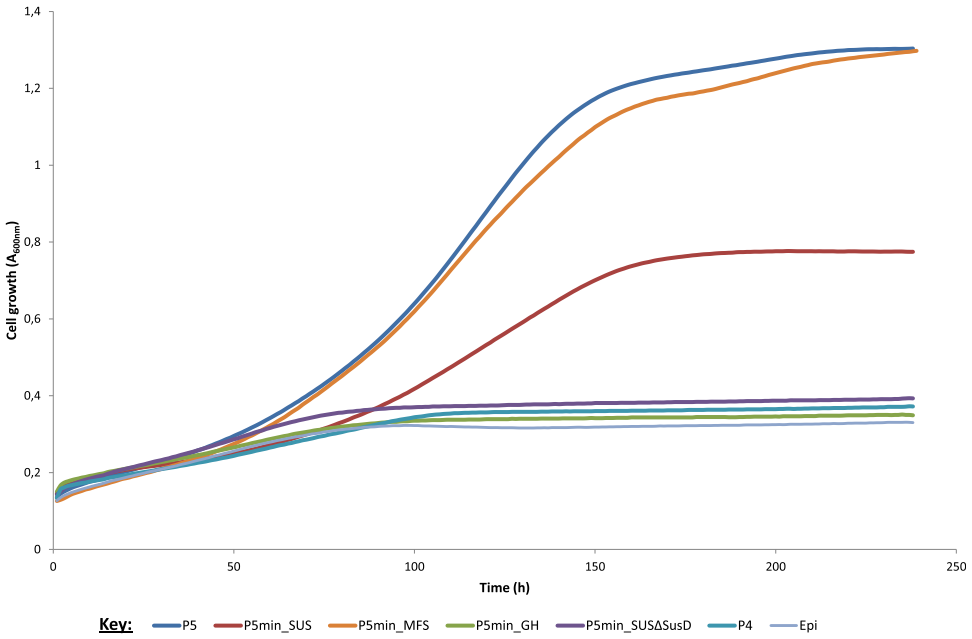

Fig. 3. Growth of the clones F5 (blue), F5min_SUS (red), F5min_MFS (orange), F5min_SUSΔSusD (purple), F5min_GH (green), F4 (cyan) and empty pCC1fos (Epi, grey) on a mixture of xylo-oligosaccharides at 0.5% (m/v). The data represent the average of at least biological triplicates.

be expressed in the clone F5 could explain the lower values in the observed activity compared to the clone F5min_GH. The enzymes are essential to hydrolyze XOS to xylose that is further metabolized by the strain. This result also confirmed that the GH16 is not responsible for any activity detected for F5 extracts.

The theoretical subcellular localization of the GHs produced by F5 was determined using LipoP 1.0 server (Juncker *et al.*, 2003). For GH43B, no peptide signal or N-terminal lipidation could be assigned, indicating a cytoplasmic location. GH43A and GH10 exhibited a putative signal peptidase II cleavage site and were predicted as N-terminally lipidated proteins indicating their potential attachment to the bacterial membrane. Bacterial lipoproteins are membrane proteins present both in Gram-negative and Gram-positive bacteria. In Bacteroidetes some lipoproteins, including the hydrolase (SusG, GH13) from SUS systems are known to be transported to the outer surface of the outer membrane (Shipman *et al.*, 1999). In *E. coli* the lipoproteins are anchored either to the inner or to the outer membrane, and oriented towards the periplasmic space (Tokuda and Matsuyama, 2004). The +2 rule postulates that the residue at the N-terminal second position is critical for the membrane specificity of lipoproteins in *E. coli* (for review see Okuda and Tokuda, 2011). An Asp in position +2 maintains the lipoprotein in the inner membrane. GH43A exhibits a Ser which is characteristic of outer membrane sorting signal (Yamaguchi *et al.*, 1988). GH10 possesses a Gly in position +2 which could imply an 'ambiguous' sorting signal (both inner and outer membrane facing periplasm) as observed for the periplasmic maltose-binding protein expressed in *E. coli* (Seydel *et al.*, 1999).

To examine the subcellular localization of the GHs in *E. coli*, enzymatic assays were performed on secreted, soluble intracellular, periplasmic and membrane protein fractions (Supporting Information Fig. S2). No secreted or soluble periplasmic activities were detected. Xylanase, arabinosidase and xylosidase activities were detected in both soluble intracellular and membrane fractions. These results are consistent with theoretical subcellular localization of the GHs that two out of the three should be attached to the membrane and one cytoplasmic.

XOS uptake in E. coli

To evidence and characterize the transport ability of the clone F5, its growth and that of different truncated variants have been monitored over 24 h in liquid minimal media (MM) supplemented with different xylose-containing glycans as sole carbon sources (Figs 3 and 4A, C and E). As control strain, we used a metagenomic clone (clone F4) able to hydrolyze a mixture of XOS into xylose and xylobiose (due the presence of GH8, GH43 and GH120) but without any transporter encoding gene within its metagenomic insert, as reported by Cecchini *et al.* (2013) (see Fig. S1 from ref. Cecchini *et al.*, 2013). Previously, the ability of the *E. coli* host strain to metabolize xylose as unique carbon source was confirmed.

When we compared the growth of F5 and F4 in MM containing a mixture of linear XOS of DP 2 to 6, only F5 could grow, even if both clones produce hydrolases able to hydrolyze XOS (Fig. 3). Similarly, F5min_GH that is

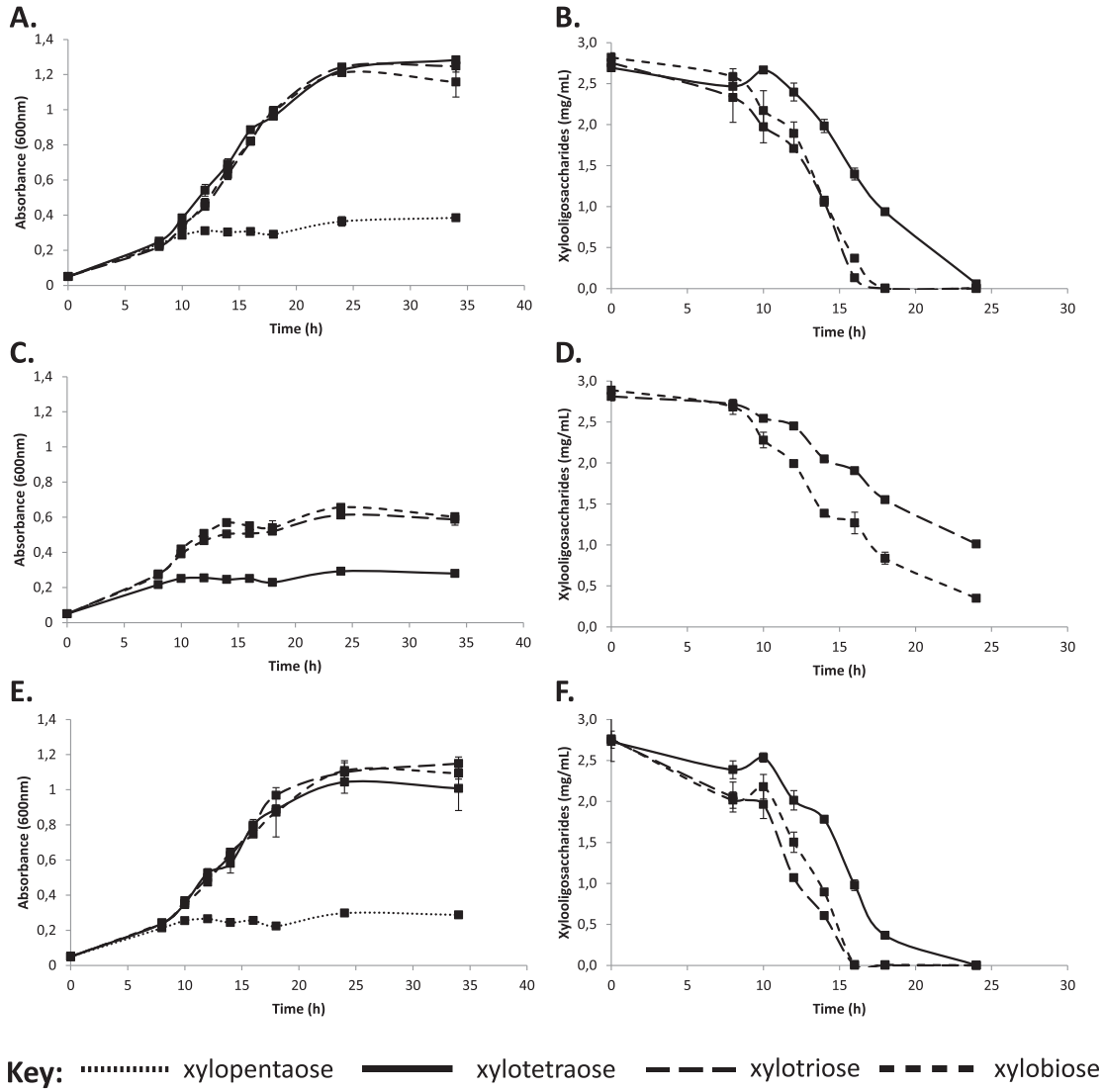

Fig. 4. Growth and xylo-oligosaccharides uptake of the clones F5, F5min_SUS and F5min_MFS. Growth curves of the clone F5 (A), F5min_SUS (C) and F5min_MFS (E) supplemented with xylopentaose (▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪), xylotetraose (▬▬▬▬▬▬), xylotriose (▬ ▬ ▬) and xylobiose (▬ ▬ ▬ ▬). HPAEC-PAD analysis of the culture supernatants of the clones F5 (B), F5min_SUS (D) and F5min_MFS (F) to measure uptake of xylotetraose (▬▬▬▬▬▬), xylotriose (▬ ▬ ▬) and xylobiose (▬ ▬ ▬ ▬). The squares indicate the sampling time points. The data represent the average of at least biological duplicates.

deleted of the genes encoding transporters did not grow on XOS (Fig. 3). Moreover, *E. coli* clone F5 displayed no ability to grow on xylan, although cell extracts were able to hydrolyze this substrate (Tables 1 and Supporting Information Table S2). These results confirm, as discussed above, that the hydrolases of F5 are not secreted in *E. coli*, and suggest that they are unlikely anchored to the outer surface of the outer membrane. But we cannot exclude the hypothesis that F5min_GH has a GH facing out and that the hydrolysis of the XOS in the media will be so slow that it does not support the growth on 24 h. It also demonstrated that functional transporters are required, in addition to functional XOS degrading GHs, to confer to recombinant *E. coli* the ability to grow on these oligosaccharides. Hence, we conclude that the clone F5 is able to metabolize XOS due to the XOS internalization mediated by one or several transporters, followed by their subsequent intracellular hydrolysis into xylose.

The specificity of the transporters has been characterized by using MM containing XOS with DP ranging from 2 to 5 as sole carbon source (Fig. 4A). While the clone F5 was able to grow in MM with DP 2 to 4, no growth has been detected in MM containing XOS larger than DP 4. Thus the internalization of xylose-containing glycans is possible up to DP 4. As both xylan and arabinoxylan could be hydrolyzed by F5 cell extracts, we tested arabino-xylo-oligosaccharides from 2 to 4 xylosyl residues branched

with 1 or 2 arabinosyl residues as sole carbon source. No growth has been observed after 24 h of incubation with these arabino-xylo-oligosaccharides (Supporting Information Table S2). Thus the oligosaccharide transport in F5 seems specific to unbranched XOS in *E. coli* host.

The PUL encodes two transport systems, a MFS and a SusC/D pair. MFS and SusC are known as membrane proteins and SusD is a lipidated protein predicted to be anchored to the outer membrane of *Bacteroides* species. To assess the transport ability and specificity of the potential transporters, new F5 variants were constructed (Fig. 1B). The first variant, named F5min_SUS, harbours the hydrolases and the arranged SusC/D homologs. The second variant, named F5min_MFS, harbours the hydrolases and the MFS transporter. The growth on a mixture of XOS (from DP 2 to DP 6) has been investigated. F5min_SUS was able to grow on linear XOS mixture but with a level of growth lower than F5 (Fig. 3). We thus confirmed that the SusC/D system protein is an active XOS transporter. However, this transporter was not sufficient to completely restore the growth of the *E. coli* strain harbouring the full F5 insert, suggesting the involvement of another transporter. On the XOS mixture, F5min_MFS showed a growth curve similar to F5 confirming the functionality of the MFS transporter (Fig. 3). To grow on XOS, F5 required at least one of its two functional transporters, the SusC or the MFS transporter.

We investigated further the function of the SusD by deleting the corresponding gene in the F5min_SUS clone. This construct was named F5min_SUSΔSusD (Fig. 1B). This deletion completely abolished the ability of *E. coli* to grow on XOS. As the *susD* gene being located downstream of *susC*, any transcriptional effect can be excluded, suggesting that the presence of SusD was essential for the functionality of the SusC transporter and for metabolism of XOS (Fig. 3). This result is in agreement with previous studies demonstrating that, in *Bacteroides* strains, ΔSusD mutants were unable to grow on their targeted poly- and oligosaccharides (Koropatkin *et al*., 2008; Cameron *et al*., 2014; Tauzin *et al*., 2016). The function of SusD is not restricted to its binding ability, its physical presence being essential for strain growth on its targeted glycan. SusD physical presence is sufficient, since supplementation of the ΔSusD strain with the SusD* variant, which is a SusD mutant unable to bind glycan, was enough to restore the growth of the bacteria (Cameron *et al*., 2014; Tauzin *et al*., 2016).

To characterize more precisely the transport specificity of each F5 transport system, namely SusC/D and MFS, we monitored the variant growth on individual linear XOS of various lengths (Fig. 4C and E). While F5min_SUS was able to grow on XOS up to DP 3, F5min_MFS was able to grow on XOS up to DP 4. To visualize the kinetic of internalization of XOS of different DPs (from DP 2 to DP 3 for F5min_SUS and from DP 2 to 4 for F5 and F5min_MFS), we monitored their disappearance from the culture supernatants using HPAEC-PAD analysis. F5 and F5min_MFS, grown on XOS of specified DP (2, 3 and 4), consumed each oligosaccharide to completion in 24 h (Fig. 4B and F). The rate of utilization was similar for xylobiose and xylotriose which were both consumed faster than xylotetraose (Fig. 4B and F). In contrast, F5min_SUS left residual xylobiose and xylotriose in the culture supernatant after 24 h, while the culture had already reached the stationary phase which remained at a final OD_{600nm} lower than with F5 and F5min_MFS (Fig. 4C and D).

To date, two other XOS MFS transporters, specific to linear XOS of either DP 3 or 6, have been reported in *Klebsiella* spp. (Qian *et al*., 2003; Shin *et al*., 2010). In addition, a TonB-dependent (SusC-like) transporter essential for growth on XOS has been identified in *Xanthomonas campestris* but its specificity has still not been investigated (Déjean *et al*., 2013). In contrast, no functional characterization of SusC and MFS transporters specific to XOS has been reported in *Bacteroides* so far, even if several xylan PULs have been described in gut bacteria. SusC (BACOVA_4393) and MFS (BACOVA_4388) contained in the XylS from *B. ovatus* shared only 37% and 30% identity with the SusC and the MFS, respectively, (Rogowski *et al*., 2015). These transporters are to date the best homologs to the F5 transporters in validated PULs targeting XOS, but again, their function has not been experimentally validated, neither was their transport specificity investigated.

Intriguingly, while no oligosaccharide was released throughout the XOS uptake by F5min_SUS, residual xylobiose and xylose were observed throughout the growth of F5 and F5min_MFS on xylobiose and xylotriose (Supporting Information Figs S3–S5). The released xylobiose and xylose were consumed thereafter as well as the XOS initially present in the culture supernatant. The observed exchanges of carbohydrates between the intra- and the extra-cellular compartments are in line with the characteristics of the transporters. The members of the MFS transporter family indeed transport the compounds depending on their concentration gradient while the TonB-dependent transporters depend on energy coupling. We hypothesized that the increase of xylose and xylobiose into the cells changed the gradient concentration leading to the release of these compounds outside the cells by the MFS transporter.

It is clear from the results presented on Figs 3 and 4 that in *E. coli*, the MFS transporter allows a higher growth rate than the SusC transporter. This is not due to the transcription levels because in *E. coli* grown on LB the MFS transporter is seven times lower expressed than SusC (Fig. 2). Three hypotheses may explain this phenomenon.

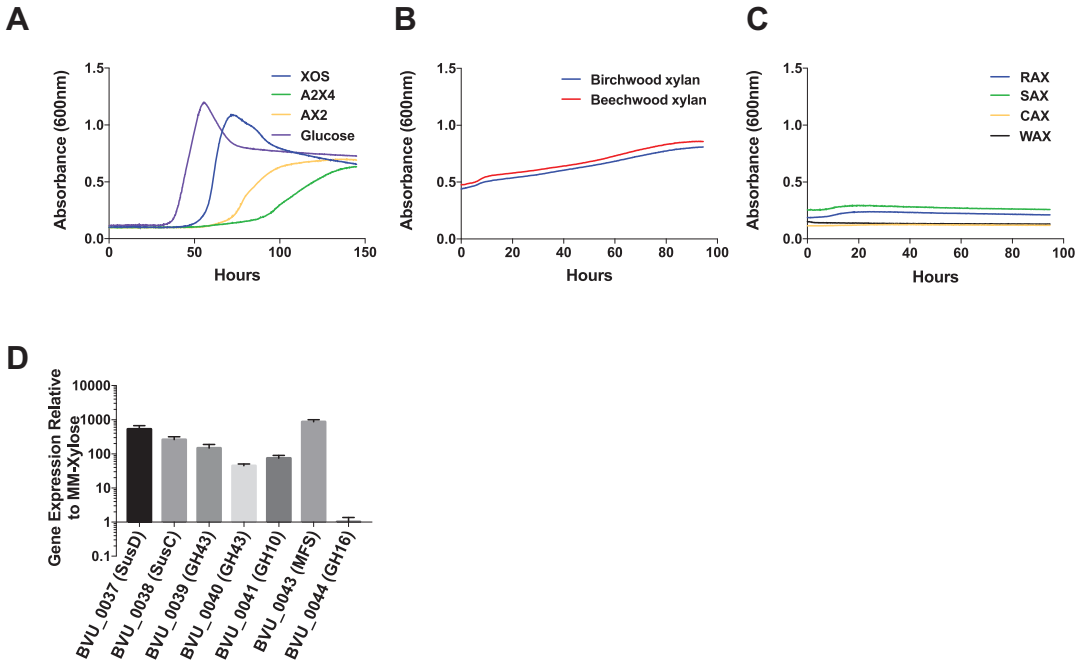

Fig. 5. Growth dynamics of *B. vulgatus* ATCC 8482 on xylo-oligosaccharides and various xylans; and transcriptional responses of *B. vulgatus* to xylo-oligosaccharides.
Growth curves of *B. vulgatus* (n=12) on xylo-oligosaccharides (A), two wood xylans (B), arabinoxylans from various sources (C). XOS, xylo-oligosaccharides; A2X4, 2^3,3^3-di-α-L-arabinofuranosyl-xylotetraose; AX2, 3^2-α-L-arabinofuranosyl-xylobiose; RAX, rice arabinoxylan; CAX, corn arabinoxylan; SAX, sorghum arabinoxylan; WAX, wheat arabinoxylan.
Transcription of *B. vulgatus* PUL *genes* in response to xylo-oligosaccharides (n = 3) (D).

First, it may be possible that the affinity for XOS is higher for the MFS transporter than for the SusC/D transporter. Another possibility would be that the SusC/D transport system would not be perfectly functional when expressed in *E. coli* compared to the native organism. The Ser residue at the N-terminal second position of SusD suggests a localization at the *E. coli* outer membrane (Yamaguchi *et al*., 1988; Okuda and Tokuda, 2011). This is in accordance with our results showing that SusD is necessary for SusC transport function. However, its orientation towards the outer membrane could not be confirmed. The last explanation could be related to the different characteristics of both transporters. The TBDT transporters indeed require energy derived from the proton motive force thanks to the interaction with the TonB-ExbB-ExbD complex (Noinaj *et al*., 2010) which is provided by *E. coli* host as described by Phansopa *et al*. (2014). Transport through SusC/D might thus be energy-consuming for the *E. coli* host. In contrast, transport through MFS is passive and just driven by gradient concentration (Yan, 2015). Whatever is the mechanism of XOS transport in the heterologous *E. coli* host, it cannot be extended to what happens in the native organism, as the cellular localization of the transporters could be different. As explained above, in *Bacteroides*, it is probable that the SusC/D transporter sits in the outer membrane, working in coordination with the MFS transporter located in the inner membrane. Based on the growth ability of the F5 clone and all of its variants, we assume that in *E. coli* the MFS protein would be located in the outer membrane, allowing the XOS internalization within the periplasmic space, where they would be hydrolyzed by the GHs bounds to the inner and/or outer membranes and oriented towards the periplasmic space. Nevertheless, further experiments will be necessary to demonstrate the exact localization of these transporters in the native strains and when produced in *E. coli*.

The PUL in B. vulgatus *is involved in XOS utilization*

We studied the biological function of the PUL in *B. vulgatus* as the metagenomic clone has high homology to genes in *B. vulgatus* ATCC 8482. Growth dynamics of *B. vulgatus* on XOS and a variety of xylans revealed that *B. vulgatus* was able to grow on the XOS and the two enzymatically digested products of wheat arabinoxylan, AX2 and A2X4. *B. vulgatus* growth on XOS is very robust, while on the branched oligosaccharides the growth is delayed and at a lower rate. In contrast to the growth dynamics on xylo- and arabino-xylo-oligosaccharides, *B. vulgatus* was unable to utilize complex heavily decorated xylans including RAX, SAX and CAX or the arabinoxylan WAX (Fig. 5A). Although the intrinsic turbidity of

Birchwood and Beechwood xylans resulted in elevated initial absorbance at 600 nm as compared to other xylans, the marginal growth rate suggested a poor and delayed utilization of these simple wood glucuronoxylans by *B. vulgatus*. Thus we hypothesize that gene clusters in *B. vulgatus* have evolved to target shorter xylose polymers, i.e. xylo-oligosaccharides, rather than xylans.

To investigate whether the utilization of XOS is contributed by the function of this PUL, we did transcriptional analyses of the PUL in response to XOS induction. The *SusC,D*-genes, two *gh43*, *gh10* and the *mfs* gene expressions are highly induced by XOS, suggesting that these genes of this PUL are responsive for XOS utilization. *SusC*- and *SusD*- like genes and *mfs* gene are most highly induced (more than 1000-fold) as compare to genes encoding the glycoside hydrolases, which are also induced at least 40-fold relative to glucose or xylose growing condition (Fig. 5B). Interestingly, no induction was observed with *gh16* gene, indicating a break in the gene cluster between MFS and GH16. Our results suggested that this PUL in *B. vulgatus* indeed involves XOS utilization. This finding further corroborated our functional analysis of the metagenomic clone in an *E. coli* recombinant host.

Conclusions

Metagenomic is a powerful tool to explore the gut bacteria diversity and specificity, as well as an extensive genetic source for discovering new functions. However, the rapid production of metagenomic data is vastly outpacing functional studies, which underscores the critical need for protein biochemical characterization and structural enzymology to inform bioinformatics and systems biology (André *et al*., 2014). As showed here, the direct study of metagenomic clones for characterization of new functions and/or new protein families could be a good strategy not only to save time, but also to study some complex mechanisms that require the synergistic action of different proteins. Previously, some multigenic systems issued of metagenomic libraries have been used to optimize *E. coli* abilities to produce ethanol or antifungal activity (Chung *et al*., 2008; Loaces *et al*., 2015), but no biochemical characterization of PULs heterologously expressed in *E. coli* has been published so far. Here, we demonstrated that *E. coli* is an interesting recombinant host for characterizing the individual components of PUL systems from *Bacteroides* strains. The present work constitutes the first experimental study of expression in *E. coli* of metagenomic multigenic cluster cloned in fosmids. Our results suggested that *E. coli* may be able to recognize its own promoter sequences within the metagenomic inserts, in particular DNA issued from Bacteroidetes (Lam and Charles, 2015). This spurious transcription can be highly advantageous to study the synergistic action of proteins encoded on a same metagenomic locus.

In the present work, we characterized a PUL-like system that confers *E. coli* the ability to metabolize XOS. Taking into account the possible inability of *E. coli* to produce extracellular or cell surface attached proteins, this new functionality requires the coordinated action of at least 2 activities: (i) functional transport to internalize oligosaccharides and (ii) oligosaccharide hydrolysis to release monomers that will be used for *E. coli* growth. The study of the transport system was based on growth screening and required the presence of both transport and hydrolytic functions.

To conclude, the present results pave the way for boosting the functional characterization of individual components of PULs, especially transporters issued from cultured and uncultured Bacteroidetes. The generic approach we developed could be extended to study other catabolic pathways that are crucial for host and dietary glycan harvesting by prominent gut bacteria, and even for metabolism of other bioactive compounds. Moreover, the construction and characterization of recombinant *E. coli* strains that are able to metabolize plant cell wall components opens the way to further metabolic engineering works to develop microbial cell factories dedicated to bio-sourced product synthesis.

Experimental procedures

Cloning

The metagenomic clones F5 (Genbank accession number HE717017) and F4 (control, HE717016) were obtained from metagenomic library issued of human fecal sample as previously described (Cecchini *et al*., 2013). Both are metagenomic fragments cloned into pCC1fos fosmid and transformed in EPI100 *E. coli* cells (Epicentre Technologies). The minimal variants of the clone F5 were constructed by using the In-Fusion HD Plus cloning kit (Clontech) following the manufacturer's instructions. The primers used in this study are listed in Supporting Information Table S1. Considering that the expression of the genes might be driven by promoter sequence potentially located in the sequence of the upstream gene, each variant was cloned by amplifying the CDS and the cognate ~600 bp to 1 kb upstream sequence.

Growth study

Media. All *E. coli* cells were grown on minimal synthetic media ($Na_2HPO_4{\cdot}12H_2O$ 17.4 g l^{-1}, KH_2PO_4 3.03 g l^{-1}, NH_4Cl 2.04 g l^{-1}, NaCl 0.51 g l^{-1}, $MgSO_4$ 0.49 g l^{-1}, $CaCl_2$ 4.38 mg l^{-1}, $Na_2EDTA{\cdot}2H_2O$ 15 mg l^{-1}, $ZnSO_4{\cdot}7H_2O$ 4.5 mg l^{-1}, $CoCl_2{\cdot}6H_2O$ 0.3 mg l^{-1}, $MnCl_2{\cdot}4H_2O$ 1 mg l^{-1}, H_3BO_3 1 mg l^{-1}, $Na_2MoO_4{\cdot}2H_2O$ 0.4 mg l^{-1}, $FeSO_4{\cdot}7H_2O$

3 mg l^{-1}, $CuSO_4 \cdot 5H_2O$ 0.3 mg l^{-1}, thiamine 0.1 g l^{-1} and leucine 0.02 g l^{-1}) containing an appropriate carbon source and supplemented with 12.5 mg l^{-1} chloramphenicol. After a first growing step in LB medium supplemented with 12.5 mg l^{-1} chloramphenicol, an overnight culture in minimum synthetic media containing xylose was realized to inoculate 0.5 ml of minimum synthetic media containing xylose at an optical density (OD) at 600 nm of 0.05 into 48-well microplate. The growth was followed by measuring the OD_{600} over 24 h at 37°C using the FLUOStar Optima (BMG Labtech).

Bacteroide vulgatus ATCC 8482 strain was routinely grown in tryptone-yeast extract-glucose (TYG) medium (Holdeman *et al.*, 1977), type-1 minimal medium (Urs, Pudlo and Martens, unpublished data). Carbon sources were added to a final concentration of 5 mg ml^{-1} unless otherwise stated. Cultures were grown at 37°C in an anaerobic chamber (10% H_2, 5% CO_2, and 85% N_2; Coy Manufacturing, Grass Lake, MI).

To quantify growth dynamics of *B. vulgatus* on various carbon sources, the increase in culture absorbance (600 nm) in 200 μl cultures was measured every 10 min on an automated plate reader (Martens *et al.*, 2011). Growth dynamics showed the average of 12 replicates for each carbon source.

Growth substrates. Growth dynamics of *E. coli* and *B. vulgatus* were performed on minimal media supplemented with a variety of oligosaccharides and polysaccharides carbon sources. We used a mixture of XOS (WAKO and IOR-TAIHE, from DP 2 to DP 7). Individual XOS from DP 2 to 5 and arabino-xylo-oligosaccharides (3^2-α-L-arabinofuranosyl-xylobiose, AX2; 2^3-α-L-arabinofuranosyl-xylotriose, AX3; 2^3,3^3-di-α-L-arabinofuranosyl-xylotriose, A2X3; 2^3,3^3-α-L-arabinofuranosyl-xylotetraose, AX4; and 2^3,3^3-di-α-L-arabinofuranosyl-xylotetraose, A2X4) were purchased from Megazyme. Simple xylans, with sparsely decorated structures, were purchased from Sigma for beechwood xylan, from Sigma for birchwood glucuronoxylan and from Megazyme for wheat arabinoxylan (WAX). More complex heavily decorated glucuronoarabinoxylans (rice, RAX; sorghum, SAX, and corn, CAX) were kind gifts of Dr. Bruce Hamake (Purdue University).

Gene expression analyses

RNA extraction and retrotranscription in cDNA. From three independent cultures of F5 clone, total RNAs were extracted as previously described (Nouaille *et al.*, 2009). Briefly, 10 ml of exponentially growing cells ($OD_{600} = 1$) in LB medium were collected, centrifuged and the pellets were immediately frozen in liquid nitrogen. Cells were disrupted through high-speed shaking with stainless steel beads. Total RNAs were extracted using an RNeasy mini kit (Qiagen) following the manufacturer's instructions. RNAs were quantified using a NanoDrop™ and their quality was controlled using a Bioanalyzer RNA kit (Agilent Technologies).

The equivalent of 50 μg of RNA was subjected to DNAse treatment and purified with RNeasy Mini spin column (Qiagen). Then 5 μg of RNA were retrotranscribed using the SuperScript® II RT (ThermoFisher Scientific) according to the manufacturer's protocol and cDNA were purified using illustra™ MicroSpin™ G-25 columns (GE Healthcare).

For the transcriptional analysis in *B. vulgatus*, total RNA was extracted using RNeasy mini kit (Qiagen) from 5 ml of exponentially growing *B. vulgatus* culture in minimal medium containing 5 mg ml^{-1} of XOS. Contaminating DNA was removed with TURBO DNA-free™ Kit (Ambion). Reverse transcription was performed with 1 μg of RNA using Super Script® III Reverse Transcriptase (Thermo Fisher Scientific) using random primers (Invitrogen) according to manufacturer's instructions. cDNA quantification was performed with a Mastercycler® ep realplex (Eppendorf), using homemade SYBR® qPCR mix containing Hot-start Taq Polymerase (NEB) and 400 nM primers, except 62.5 nM primers for 16S rRNA, for 40 cycles of 95°C for 3 s, 52°C for 20 s, 68°C for 20 s, followed by a melting step to determine amplicon purity. All transcript levels were normalized based on 16S rRNA abundance. The expression of each gene in the PUL was expressed relative to the transcript level of glucose or xylose growing condition.

Primer design. The primers used for real-time quantitative PCR of each gene on the F5 metagenonic clone insert were designed with Bio-Rad Beacon Designer software to have lengths from 18 to 22 bases, GC contents of more than 50%, melting temperatures of about 60°C and to amplify PCR products between 83 and 148 bases long (Supporting Information Table S1).

The primers used for real-time quantitative PCR for each gene in *B. vulgatus* PUL were designed using Primer 3. These primers range from 18 to 24 bases, with GC contents between 40% and 60%. The melting temperatures lies around 60°C and the amplicon size range between 80 and 150 bases (Supporting Information Table S1).

High throughput real-time quantitative PCR. High throughput real-time quantitative PCR was carried out using the 48.48 dynamic array™ IFCs and the BioMark™ HD System (Fluidigm Corporation, CA, USA) following the manufacturer's protocol (Spurgeon *et al.*, 2008) and performed at GeT-PlaGe facilities (Castanet, France). Prior to RNA expression analysis, primer specificity and the absence of genomic DNA contamination in extracted total RNAs were checked.

In total 1044 data were collected from qPCR analyses combining 4 technical replicates (used at 3 different dilutions) issued from 3 biological samples and the 29 primer couples corresponding to the 27 genes of the metagenomic insert and the 2 additional control genes (*ihfB* and *cam*).

Data analysis. Relative mRNA expression means were calculated from the biological triplicates after initial raw data analysis accomplished with the Fluidigm real-time PCR analysis software v.4.1.2. The PCR efficiency was checked for each primer couple and was close to 100%. The comparative ΔΔCt method was used to calculate the change in transcripts levels with correction (Livak and Schmittgen, 2001). As the best alternative, the mean expression of 5 less expressed genes was used to determine which gene is

significantly expressed compared to the threshold we fixed at fivefold the value of the less expressed genes.

Two reference genes were used for data normalization between samples: the integration host factor β-subunit (*ihfB*) which is one of the commonly used reference gene in *E. coli* (Weglenska *et al.*, 1996) as its expression remained constant throughout growth, and the chloramphenicol resistance gene (*cam*) encoding for chloramphenicol acetyltransferase present on the recombinant vector and essential for antibiotic resistance.

Enzymatic assays

The cells were grown on LB supplemented with 12.5 mg l^{-1} chloramphenicol and inoculated with overnight culture at 0.05 of OD_{600}. When the OD_{600} reaches 1, the cells were harvested and the pellet was suspended in 50 mM potassium phosphate buffer pH 7.0 containing lysozyme (0.5 mg ml^{-1} final concentration) to reach an OD_{600} of 80. After incubation at 37°C for one hour, the suspension was frozen 15 min at −80°C and then defrosted. Then samples were centrifuged and the supernatant (cell extracts) was used to performed activity test. All reactions were carried out at 37°C in 50 mM potassium phosphate buffer pH 7.0.

The activity tests against complex polysaccharides (xylan, arabinoxylan, arabinan, arabinogalactan, β-glucan, xyloglucan) were measured using the 3,5-dinitrosalicylic acid reducing-sugar (DNS) assay. Reaction samples (250 µl of cell extract incubated with 5 mg ml^{-1} of specified substrate) were added to an equal volume of DNS reagent to terminate the reaction, and the colour was developed by boiling for 5 min. Enzymatic activities with various para-NitroPhenol (*p*NP) sugar derivatives were also realized. After incubation of 150 µl of cell extract with 1 mM *p*NP-glycosides (*p*NP-α-D-xylopyranose, *p*NP-β-D-xylopyranose, *p*NP-α-L-arabinofuranose, *p*NP-α-L-arabinopyranose and *p*NP-β-L-arabinopyranose), the reaction was stopped by raising the pH to 11.0 through the addition of an equal volume of 0.2M Na_2CO_3. The released of reducing-sugar (DNS assay) and *p*NP were measured in an Optima (TECAN) at A_{540nm} and A_{405nm} respectively. A standard curve was used to calculate product concentration.

Cellular localization of proteins

The control (pCC1fos empty) and F5min_GH clones were grown in 250 ml of LB at 37°C until OD_{600} reaches 0.9. The cells were collected by centrifugation at 4400 g for 10 min at 4°C. The supernatant was filtered (0.22 µm) and tested for secreted activity. The other protein fractions were obtained from the different treatments of the pellet as described by Larsbrink *et al.* (2011).

Briefly, the periplasmic proteins were collected using an osmotic shock. The cells were washed with 10 ml of 50 mM Tris-HCl (pH 7.7) and collected by centrifugation at 4400 g for 10 min at 4°C. The pellet was resuspended in 50 ml of 30 mM Tris-HCl, 20% (w/v) sucrose and 1 mM EDTA (pH 8.0), and the cells were incubated at room temperature for 10 min. The cells were then collected by centrifugation at 4400 g for 15 min at 4°C. The pellet was resuspended in ice-cold 5 mM $MgSO^4$, and the cells were incubated on ice for 10 min. The cells were collected by centrifugation at 14 000 g for 10 min at 4°C. The supernatant was retained and contained the periplasmic proteins. The pellet was resuspended in 50 mM sodium phosphate buffer (pH 7.4) and sonicated to lyse cells. The lysate was centrifuged at 5000 g for 10 min at 4°C. Using an ultracentrifuge, the supernatant was centrifuged at 100 000 g for 1 h at 4°C to recover the cytoplasmic proteins. The pellet of the lysate was resuspended in 100 mM sodium carbonate buffer (pH 9.0) and centrifuged at 100 000 g for 1 h at 4°C. The supernatant from this step contained the potential trapped soluble proteins and/or weakly membrane-associated proteins. The pellet, containing the membrane proteins, was resuspended in 50 mM sodium phosphate buffer (pH 7.4).

XOS uptake

To assay the XOS uptake, cells were grown in M9 medium supplemented with XOS of specified chain length or mixture of XOS at 37°C. Growth was monitored by measuring the A_{600nm}. During the growth, samples were collected at regular time point and centrifuged. The supernatants were filtered and conserved at −20°C. The amount of XOS present in the culture supernatants were analyzed by HPAEC-PAD on a Dionex ICS-3000 system (Dionex) equipped with a CarboPac PA100 column. The analyses were carried out at 30°C with a flow rate of 0.5 ml min^{-1} with the following multistep gradient: 0–30 min (0–60% B), 30–31 min (60–0% B) and 31–36 min (0% B). Solvents were 150 mM NaOH (eluent A) and 150 mM NaOH, 500 mM CH_3COONa (eluent B). To quantify the remaining concentration in the culture supernatant of XOS, the respective commercial oligosaccharides (Megazymes) were used as standards.

Bioinformatic analyses

Promoter consensus sequences used to identify promoters from *E. coli* (*rpoD*/σ^{70}) and *Bacteroides* (σ^{ABfr}) were ($TTGACA_{15\text{-}19}TATAAT$) and ($TTTG_{19\text{-}21}TA_2TTTG$), respectively, (Mastropaolo *et al.*, 2009). The BPROM program was used to identify the putative promoters in *E. coli* (Solovyev and Salamov, 2011).

LipoP and SignalP servers were used to determine the presence and location of lipoprotein and other protein signal peptide cleavage sites, respectively, (Juncker *et al.*, 2003; Petersen *et al.*, 2011).

Acknowledgements

This research was funded by the French National Center of Excellence Toulouse White Biotechnology. We cordially thank Amandine Deroite, Nathan Davidenko, Adrien Guibert, Clarisse Lozano and Nelly Monties for their technical assistance. The analytic work was carried out at the Laboratory for Bio-Systems & Process Engineering (Toulouse, France) with the equipment of the ICEO facility. MetaToul (Metabolomics & Fluxomics Facitilies, Toulouse, France, www.metatoul.fr) and

its staff members are gratefully acknowledged for technical support and access to microplate reader. MetaToul is part of the national infrastructure MetaboHUB-ANR-11-INBS-0010 (The French National infrastructure for metabolomics and fluxomics, www.metahub.fr). MetaToul is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund, SICOVAL, the Infrastructures en Biologie Sante et Agronomie (IBiSa, France), the Centre National de la Recherche Scientifique (CNRS) and the Institut National de la Recherche Agronomique (INRA). The work on Bacteroides was supported by the grant GM090080 from NIH.

Conflict of interest

Authors have no conflict of interest to declare.

References

André, I., Potocki-Véronèse, G., Barbe, S., Moulis, C., and Remaud-Siméon, M. (2014) CAZyme discovery and design for sweet dreams. *Curr Opin Chem Biol* **19**: 17–24.

Bolam, D.N., and Koropatkin, N.M. (2012) Glycan recognition by the Bacteroidetes Sus-like systems. *Curr Opin Struct Biol* **22**: 563–569.

Cameron, E.A., Maynard, M.A., Smith, C.J., Smith, T.J., Koropatkin, N.M., and Martens, E.C. (2012) Multidomain carbohydrate-binding proteins involved in Bacteroides thetaiotaomicron starch metabolism. *J Biol Chem* **287**: 34614–34625.

Cameron, E.A., Kwiatkowski, K.J., Lee, B.H., Hamaker, B.R., Koropatkin, N.M., and Martens, E.C. (2014) Multifunctional nutrient-binding proteins adapt human symbiotic bacteria for glycan competition in the gut by separately promoting enhanced sensing and catalysis. *MBio* **5**: e01441–e01414.

Cecchini, D.A., Laville, E., Laguerre, S., Robe, P., Leclerc, M., Doré, J., *et al*. (2013) Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. *PLoS One* **8**: e72766.

Chung, E.J., Lim, H.K., Kim, J., Choi, G.J., Park, E.J., Lee, M.H, *et al*. (2008) Forest soil metagenome gene cluster involved in antifungal activity expression in *Escherichia coli*. *Appl Environ Microbiol* **74**: 723–730.

Cuskin, F., Lowe, E.C., Temple, M.J., Zhu, Y., Cameron, E.A., Pudlo, N.A, *et al*. (2015) Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517**: 165–169.

Déjean, G., Blanvillain-Baufumé, S., Boulanger, A., Darrasse, A., Bernonville, T.D.D., Girard, A.L, *et al*. (2013) The xylan utilization system of the plant pathogen *Xanthomonas campestris* pv campestris controls epiphytic life and reveals common features with oligotrophic bacteria and animal gut symbionts. *New Phytol* **198**: 899–915.

Dodd, D., Mackie, R.I., and Cann, I.K.O. (2011) Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes. *Mol Microbiol* **79**: 292–304.

Ferguson, A.D., and Deisenhofer, J. (2002) TonB-dependent receptors-structural perspectives. *Biochim Biophys Acta* **1565**: 318–332.

Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Reyes-Duarte, D., Santos, V.A., *et al*. (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ Microbiol* **7**: 1996–2010.

Hehemann, J.H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G. (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**: 908–912.

Hess, M., Sczyrba, A., Egan, R., Kim, T.W., Chokhawala, H., Schroth, G., *et al*. (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331**: 463–467.

Holdeman, L.V., Cato, E.D., and Moore, W.E.C. (1977) *Anaerobe Laboratory Manual*, 4th ed. Blacksburg, VA: Virginia Polytechnic Institute and State University.

Juncker, A.S., Willenbrock, H., Heijne, G.V., Brunak, S., Nielsen, H., and Krogh, A. (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* **12**: 1652–1662.

Koropatkin, N.M., Martens, E.C., Gordon, J.I., and Smith, T.J. (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**: 1105–1115.

Lam, K.N., and Charles, T.C. (2015) Strong spurious transcription likely contributes to DNA insert bias in typical metagenomic clone libraries. *Microbiome* **3**: 22.

Lam, K.N., Cheng, J., Engel, K., Neufeld, J.D., and Charles, T.C. (2015) Current and future resources for functional metagenomics. *Front Microbiol* **6**: 1196.

Larsbrink, J., Izumi, A., Ibatullin, F.M., Nakhai, A., Gilbert, H.J., Davies, G.J., and Brumer, H. (2011) Structural and enzymatic characterization of a glycoside hydrolase family 31 α -xylosidase from Cellvibrio japonicus involved in xyloglucan saccharification. *Biochem J* 567–580.

Larsbrink, J., Rogers, T.E., Hemsworth, G.R., McKee, L.S., Tauzin, A.S., Spadiut, O., *et al*. (2014) A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**: 498–502.

Livak, K.J., and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods* **25**: 402–408.

Loaces, I., Amarelle, V., and Muñoz-Gutierrez, I. (2015) Improved ethanol production from biomass by a rumen metagenomic DNA fragment expressed in *Escherichia coli* MS04 during fermentation. *Appl Environ Microbiol* **99**: 9049–9060.

Markowitz, V.M., Chen, I.M.A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., *et al*. (2012) IMG: the integrated microbial genomes database and comparative analysis system. *Nucl Acids Res* **40**: D115–D122.

Martens, E.C., Chiang, H.C., and Gordon, J.I. (2008) Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**: 447–457.

Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N. A., Wu, M., McNulty, N.P., *et al*. (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* **9**: e1001221.

Mastropaolo, M.D., Thorson, M.L., and Stevens, A.M. (2009) Comparison of Bacteroides thetaiotaomicron and *Escherichia coli* 16S rRNA gene expression signals. *Microbiology* **155**: 2683–2693.

Nielsen, H.B., Almeida, M., Juncker, A.S., Rasmussen, S., Li, J., Sunagawa, S., *et al.* (2014) Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol* **32**: 822–828.

Noinaj, N., Guillier, M., Barnard, T.J., and Buchanan, S.K. (2010) TonB-dependent transporters: regulation, structure, and function. *Annu Rev Microbiol* **64**: 43–60.

Nouaille, S., Even, S., Charlier, C., Loir, Y.L., Cocaign-Bousquet, M., and Loubière, P. (2009) Transcriptomic response of *Lactococcus lactis* in mixed culture with *Staphylococcus aureus*. *Appl Environ Microbiol* **75**: 4473–4482.

Okuda, S., and Tokuda, H. (2011) Lipoprotein sorting in bacteria. *Annu Rev Microbiol* **65**: 239–259.

Petersen, T.N., Brunak, S., Heijne, G.V., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**: 785–786.

Phansopa, C., Roy, S., Rafferty, J.B., Douglas, C.W.I., Pandhal, J., Wright, P.C., *et al.* (2014) Structural and functional characterization of NanU, a novel high-affinity sialic acid-inducible binding protein of oral and gut-dwelling Bacteroidetes species. *Biochem J* **458**: 499–511.

Qian, Y., Yomano, L.P., Preston, J.F., Aldrich, H.C., and Ingram, L.O. (2003) Cloning, characterization, and functional expression of the *Klebsiella oxytoca* xylodextrin utilization operon (xynTB) in *Escherichia coli*. *Appl Environ Microbiol* **69**: 5957–5967.

Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., *et al.* (2015) Glycan complexity dictates microbial resource allocation in the large intestine. *Nat Commun* **6**: 7481.

Roy, S., Douglas, C.W.I., and Stafford, G.P. (2010) A novel sialic acid utilization and uptake system in the periodontal pathogen *Tannerella forsythia*. *J Bacteriol* **192**: 2285–2293.

Schauer, K., Rodionov, D.A., and Reuse, H. D. (2008) New substrates for TonB-dependent transport: do we only see the "tip of the iceberg?". *Trends Biochem Sci* **33**: 330–338.

Seydel, A., Gounon, P., and Pugsley, A.P. (1999) Testing the '+ 2 rule' for lipoprotein sorting in the *Escherichia coli* cell envelope with a new genetic selection. *Mol Microbiol* **34**: 810–821.

Shin, H., Mcclendon, S., Vo, T., and Chen, R.R. (2010) *Escherichia coli* binary culture engineered for direct fermentation of hemicellulose to a biofuel. *Appl Environ Microbiol* **76**: 8150–8159.

Shipman, J.A., Cho, K.H., Siegel, H.A., and Salyers, A.A. (1999) Physiological characterization of SusG, an outer membrane protein essential for starch utilization by *Bacteroides thetaiotaomicron*. *J Bacteriol* **181**: 7206–7211.

Shultzaberger, R.K., Chen, Z., Lewis, K.A., and Schneider, T.D. (2007) Anatomy of *Escherichia coli* s 70 promoters. *Nucleic Acids Res* **35**: 771–788.

Singh, S.S., Typas, A., Hengge, R., and Grainger, D.C. (2011) *Escherichia coli* p 70 senses sequence and conformation of the promoter spacer region. *Nucleic Acids Res* **39**: 5109–5118.

Solovyev, V., and Salamov, A. (2011) Automatic annotation of microbial genomes and metagenomic sequences. In *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*. Li, R.W. (ed.). New York: Nova Science Publishers.

Spurgeon, S.L., Jones, R.C., and Ramakrishnan, R. (2008) High throughput gene expression measurement with real time pcr in a microfluidic dynamic array. *PLoS One* **3**: e1662.

Stafford, G., Roy, S., Honma, K., and Sharma, A. (2012) Sialic acid, periodontal pathogens and *Tannerella forsythia*: stick around and enjoy the feast!. *Mol Oral Microbiol* **27**: 11–22.

Strachan, C.R., Singh, R., VanInsberghe, D., Ievdokymenko, K., Budwill, K., Mohn, W.W., *et al.* (2014) Metagenomic scaffolds enable combinatorial lignin transformation. *Proc Natl Acad Sci USA* **111**: 10143–10148.

Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., *et al.* (2010) Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res* **20**: 1605–1612.

Tauzin, A.S., Kwiatkowski, K.J., Orlovsky, N.I., Smith, C.J., Creagh, A.L., Haynes, C.A., *et al.* (2016) Molecular dissection of xyloglucan recognition in a prominent human gut symbiont. *MBio* **7**: e02134–e02115.

Terrapon, N., Lombard, V., Gilbert, H.J., and Henrissat, B. (2015) Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* **31**: 647–655.

Tokuda, H., and Matsuyama, S.I. (2004) Sorting of lipoproteins to the outer membrane in *E. coli*. *Biochim Biophys Acta* **1693**: 5–13.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., and Gordon, J.I. (2007) The human microbiome project. *Nature* **449**: 804–810.

Vimrt, E.R., and Troy, F.A. (1985) Identification of an inducible catabolic system for sialic acids (nan) in *Escherichia coli*. *J Bacteriol* **164**: 845–853.

Wang, Y., Chen, Y., Zhou, Q., Huang, S., Ning, K., Xu, J., *et al.* (2012) A culture-independent approach to unravel uncultured bacteria and functional genes in a complex microbial community. *PLoS One* **7**: e47530.

Weglenska, A., Jacob, B., and Sirko, A. (1996) Trancriptional pattern of *Escherichia coli* ihfB (himD) gene expression. *Gene* **181**: 85–88.

Yamaguchi, K., Yu, F., and Inouye, M. (1988) A single amino acid determinant of the membrane localization of lipoproteins in *E. coli*. *Cell* **53**: 423–432.

Yan, N. (2015) Structural biology of the major facilitator superfamily transporters. *Annu Rev Biophys* **44**: 257–283.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.