For publication in Method section of AEM

Functional Screening of Metagenome and Genome Libraries for Detection of Novel Flavonoid-Modifying Enzymes

U. Rabausch¹, J. Juergensen¹, N. Ilmberger¹, S. Böhnke¹, S. Fischer², B. Schubach², M. Schulte² and W. R. Streit*¹

Abteilung für Mikrobiologie und Biotechnologie, Biozentrum Klein Flottbek, Universität Hamburg, Germany¹

Merck KGaA D 64293 Darmstadt, Germany²

Corresponding author:

Wolfgang R. Streit, Abteilung für Mikrobiologie und Biotechnologie,

Biozentrum Klein Flottbek, Universität Hamburg,

Ohnhorststraße 18, D-22609 Hamburg, Germany,

Telephone: (49) 428 16 463

Electronic mail: wolfgang.streit@uni-hamburg.de
The functional detection of novel enzymes other than hydrolases from metagenomes is limited since only very few reliable screening procedures are available that allow the rapid screening of large clone libraries. For the discovery of flavonoid-modifying enzymes in genome and metagenome clone libraries, we have developed a new screening system based on high performance thin layer chromatography (HPTLC). This metagenome extract thin layer chromatography analysis (META) allows the rapid detection of glycosyltransferase (GT) but also other flavonoid-modifying activities. The developed screening method is highly sensitive and the amount of 4 ng modified flavonoid molecules can be detected. This novel technology was validated against a control library of 1,920 fosmid clones generated from a single Bacillus cereus isolate, and then used to analyze more than 38,000 clones derived from two different metagenomic preparations. Thereby we identified two novel UDP glycosyltransferases (UGT) genes. The metagenome-derived gtfC gene encoded for a 52 kDa protein and the deduced amino acid sequence was weakly similar to putative UGTs from Fibrisoma and Dyadobacter. GtfC mediated the transfer of different hexose moieties and exhibited high activities on flavones, flavonols, flavanones and stilbenes but accepted also isoflavones and chalcones. From the control library we identified a novel macroseide glycosyltransferase (MGT) with a calculated MW of 46 kDa. The deduced amino acid sequence was highly similar to MGTs from Bacillus thuringiensis. Recombinant MgtB transferred sugar residue from UDP-glucose effectively to flavones, flavonols, isoflavones and flavanones. Moreover MgtB exhibited high activity on larger flavonoid molecules as tiliroside.
INTRODUCTION

For more than a decade, metagenome research has demonstrated that it is a powerful tool for the discovery of novel biocatalysts and other valuable biomolecules by using either function- or sequence-based screening technologies (1-3). Sequence-based approaches allow the identification of candidate genes. Especially the development of next generation sequencing (NGS) technology and improved bioinformatic tools have significantly advanced this methodology (4). However, a major drawback of sequence-based screening technologies is that they do not allow direct conclusions on the functionality and biochemical parameters of the encoded enzymes. Furthermore sequence-based searches are limited to the identification of homologs of already known motifs (5). Yet another problem associated with the sequence-based approach is that it often reveals only partial genes, which make a subsequent expression and a detailed biochemical analysis of the gene products difficult if not impossible. In contrast, the function-driven approach is usually much slower, more labor and cost intensive but results in the detection of complete and active enzyme clones. It is of course well known, that the function-driven metagenomics is hampered due to expression problems of the genes, and incorrect processing of enzymes (6-8). However, the function-based approach allows the identification of truly novel enzymes and it allows a first judgment on the actual enzyme activities and physicochemical parameters already during the screening process.

It is not surprising that the majority of metagenome-derived enzymes that have been characterized biochemically mainly originated from function-based screenings (9, 10). Interestingly, the majority of biocatalysts that have been identified through functional approaches were hydrolytic enzymes, mainly esterases and glycoside hydrolases (11, 12). This is perhaps linked to the simple plate-based screening procedures required for rapid detection of these hydrolytic enzymes (13). In this context it is noteworthy, that the function-based screening technology is limited by the availability of sensitive and reliable assays for enzymes other than hydrolases and which are of importance for biotechnology. Since the overall hit rates are usually low, metagenome screens often require high throughput
screening (HTS) technologies to be efficient and the screenings need to be done under production near conditions (6). Further, enzyme screens often require complex substrates and sophisticated chromogenic assays as well as HPLC or similar analytical methods. Clearly, the setup and development of novel function-driven metagenome screening assays is a very tedious and time-consuming process. This may be one reason why only rather few function-based metagenome screening techniques have been developed during the last decade focusing on enzymes other than hydrolases and with relevance to biotechnological processes (14-20). Thus, there is an urgent need to develop function-based screening methods for genes and enzymes belonging to enzyme classes other than hydrolases and that are of relevance to biotechnology.

Flavonoids as natural substances in fruits and vegetables are part of our daily nutrition. They are well known for their anti-oxidative and radical scavenging nature and even more having various beneficial effects on human health (21). Because of these broad effects there is an increasing demand for specific flavonoids in the cosmetic, the pharma- and nutraceutical industries (22-24). Meeting this demand a major problem arises from their limited availability. Flavonoids are exclusively produced in plants at low levels. The extraction is linked to the use of large quantities of solvents, and the chemical modification is not easily accomplished due to their rather complex structure (25).

The regio-specific modification of flavonoids remains difficult as the directed chemical modification mostly fails. Thus flavonoid-modifying enzymes have gained interest as they can mediate the regio- and stereochemical modification of flavonoids (26). Especially the specific glycosylation of flavonoids lies in the focus of research as a modification to influence water solubility and bioavailability of the polyphenolics (27, 28). Enzymes that catalyze this reaction are glycosyltransferases (GTs). Generally, GTs mediate the transfer of sugar residues from a donor substrate to acceptor molecules. Based on their sequence similarities GTs are currently classified into 94 families (29). GT family 1 (GT1) comprises of enzymes that catalyze the glycosylation of small lipophilic molecules (30). These enzymes (EC 2.4.1.x) that use a nucleotide-activated donor belong to the UDP-glycosyltransferase (UGT) superfamily.
and are also referred as Leloir enzymes (31, 32). Glycosyltransferases acting on flavonoids also belong to GT1 (33). Enzymes of GT1 possess a GT-B fold structure and present an inverting reaction mechanism concerning the linkage of the transferred sugar moiety (34). Up to now very few flavonoid-acting GT1s of prokaryotic origin have been identified and characterized in detail. The currently known flavonoid accepting UGTs derived from Gram-positive bacteria all belong to the macroside glycosyltransferase (MGT) subfamily and originate from *Bacilli* and *Streptomycetes* (35-37). Furthermore a single flavonoid acting UGT derived from the Gram-negative *Xanthomonas campestris* is known (38).

Within the current publication we report on a semi-automated TLC screening for clone pools of metagenome libraries. The novel method allows the rapid identification of flavonoid-modifying enzyme clones. Using this technology, we have screened more than 40,000 fosmid clones and thereby identified two positive clones that showed significant flavonoid GT activities. The two novel enzymes, designated MgtB and GtfC, belong to GT family 1 and are highly active on flavonoids and similar molecules. While MgtB is highly similar to a hypothetical *B. thuringiensis* MGT, GtfC is weakly similar to a hypothetical protein from *Fibrisoma limi*. 
Bacterial strains and plasmids used in the present work are listed in TABLE S1 and primers are listed in TABLE S2. If not otherwise stated Escherichia coli was grown at 37°C in LB medium (1% tryptone, 0.5% yeast extract, 0.5% NaCl) supplemented with appropriate antibiotics. Bacillus isolates were grown at 30°C in the same medium. All used chemical reagents were of analytical laboratorial grade. Polyphenolic substances were purchased from the following companies located in Germany: Merck KGaA, Darmstadt; Carl Roth GmbH, Karlsruhe; Sigma-Aldrich, Heidelberg and Applichem GmbH, Darmstadt. Additional flavonoids were ordered from Extrasynthese (Lyon, France). Stock solutions of the polyphenols were prepared in DMSO in concentrations of 100 mM.

Isolation of DNA and fosmid library construction

Strain Bacillus sp. HH1500 was originally isolated from a soil sample of the botanical garden of the University of Hamburg. DNA from Bacillus sp. HH1500 was isolated using the peqGOLD Bacterial DNA Kit (PEQLAB Biotechnologie GmbH, Erlangen, Germany) following the manufacturer protocol. The sample for the construction of the elephant feces library was derived from the Hagenbeck Zoo (Hamburg, Germany). Fresh feces of a healthy six year old female Asian elephant (Elephas maximus) named Kandy were taken and stored at -20°C in TE buffer (10 mM TRIS-HCl, 1 mM EDTA, pH 8) containing 30% (v/v) glycerol until DNA extraction. For DNA extraction the QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany) was used. The kit was applied according the manufacturer protocol. As recommended we increased the incubation temperature in ASL buffer to 95°C. Isolation of DNA from Elbe river sediment was performed with sediment samples from the tidal flat zone of the river Elbe nearby Glückstadt (Germany) at low tide (53°44’40” N, 009°, 26°14’ E). Environmental DNA was extracted using the SDS-based DNA extraction method published by Zhou and coworkers (39).
Construction of the genomic and metagenomic libraries in *E. coli* EPI300 cells harboring fosmid pCC1FOS was achieved with the CopyControl™ Fosmid Library Production Kit (Epicentre Biotechnologies, Madison, USA) according to the manufacturer protocol using minor modifications as previously published (40). Clones were transferred into 96 well microtitre plates containing 150 µL liquid LB medium with 12.5 µg/mL of chloramphenicol and allowed to grow overnight. Libraries were stored at -70°C after adding 100 µL of 86% glycerol to each microtitre well. The genomic fosmid library of *Bacillus* sp. HH1500 comprised 1,920 clones; a total of 35,000 clones were obtained for the river Elbe sediment library and the elephant feces library encompassed a total of 20,000 clones. All libraries contained fosmids with average insert sizes of 35 kb.

**Molecular cloning strategies**

Fragments of pCC1FOS fosmids were subcloned into pBluescript II SK+ vector using *Hind*III according to the restriction of the fosmid clones pFOS4B2 and pFOS144C11. The resulting plasmids were designated pSK4B2 and pSK144C11, respectively. Further subcloning of pSK144C11-derived fragments was achieved in pTZ19R-Cm with restriction enzymes *Eco*RI and *Pst*I. The obtained clones were designated as pTZ144E and pTZ144P, respectively. *E. coli* DH5α was transformed with the plasmids by heat shock and the plasmid carrying subclones were identified by blue white screening on LB agar plates containing 10 µM 5-bromo-4-chloro-indolyl-β-D-galactopyranoside (X-Gal) and 400 µM isopropyl-β-D-thiogalactopyranoside (IPTG) after overnight growth. Different clones were analyzed by plasmid purification, followed by enzymatic digestion and agarose gel electrophoresis and/or DNA sequencing.

PCR Amplification of open reading frames (ORFs) was performed with fosmid DNA as a template. The reactions were performed in 30 cycles. To amplify *mgtB* the primers mgt1-*Xho*I-for and mgt1-*Xho*I-rev were used, inserting an *Xho*I endonuclease restriction sites 5’ and 3’ of the ORF (see TABLE S2). For cloning of *gtfC* primer pair gtf-*Nde*-for and gtf-*Bam*-rev was used, inserting an *Nde*I site including the start codon 5’ and a *Bam*HI site 3’ of the ORF (TABLE S2). PCR fragments were ligated into pDrive using the QIAGEN PCR
Cloning Kit (Qiagen, Hilden, Germany) and cloned into *E. coli* DH5α. Resulting clones designated as pDmgtB and pDgtfC, respectively, were analyzed for activity in biotransformation and by DNA sequencing for the correct insert. Ligation of *mgtB* and *gtfC* into expression vector pET19b (Merk, KGaA, Darmstadt, Germany) was achieved using the inserted endonuclease restriction sites of each ORF. Plasmids containing the correct insert were designated pET19mgtB and pET19gtfC, respectively. *E. coli* DH5α clones harboring the desired plasmids were detected by direct colony PCR using T7 terminator primer and mgt1-XhoI-for to confirm *mgtB* and T7 terminator primer and gtf-Nde-for to verify *gtfC*, respectively. Additionally, the inserts of pET19mgtB and pET19gtfC were sequenced using T7 promoter and T7 terminator primers (TABLE S2) to verify the constructs.

**Overproduction and purification of enzymes**

For overproduction of *deca*-histidin (His10-) tagged proteins *E. coli* BL21 (DE3) was transformed with pET19b constructs. An overnight preculture was harvested by centrifugation and 1% was used to inoculate an expression culture. Cells carrying pET19mgtB were grown at 22°C until 0.7 OD600. The culture was transferred to 17°C and induced by 100 µM IPTG. After 16 h, the culture was harvested by centrifugation at 7,500 g at 4°C. Cells were resuspended in 50 mM phosphate buffer saline (PBS) with 0.3 M NaCl at pH 7.4 and disrupted by ultrasonication with a S2 sonotrode in a UP200S (Hielscher, Teltow, Germany) at a cycle of 0.5 and an amplitude of 75%.

The overproduction of *deca*-histidin-tagged GtfC was induced at 37°C at an OD600 of 0.6, with 100 µM IPTG. Cells were then incubated for four hours, harvested and lysed as stated above for MgtB.

Crude cell extracts were centrifuged at 15,000 g and 4°C to sediment the cell debris. The clarified extracts were loaded on 1 mL HisTrap FF Crude columns using the ÄKTAprime plus system (GE Healthcare). The enzymes were purified according to the manufacturer protocol for gradient elution of His-tagged proteins. Eluted protein solutions were dialyzed twice against 1,000 vol. 50 mM PBS pH 7.4 with 0.3 M NaCl at 4°C. The purification was...
analyzed on a 12% SDS-PAGE. The concentration of protein was determined by Bradford method using Roti-Quant (Carl Roth GmbH, Karlsruhe, Germany).

**Biotransformations and Biocatalyses**

For the detection of flavonoid modifications in bacteria we used a biotransformation approach. Cultures were grown in LB medium with appropriate antibiotic overnight. Expression cultures were prepared as stated above for overproduction of enzymes. The cells were sedimented by centrifugation at 4,500 g and resuspended in 50 mM sodium phosphate buffer pH 7 supplemented with 1% (w/v) α-D-glucose. Biotransformations with a final concentration of 100 µM flavonoid inoculated from stock solutions of 100 mM in DMSO, i.e. 0.1%, were incubated in Erlenmeyer flasks at 30°C and 175 rpm up to 24 hours. Samples of 4 mL were withdrawn and acidified with 100 µL 1M H₃PO₄ for extraction in 2 mL ethyl acetate. They were shaken for 1 minute and phase separated by centrifugation at 2,000 g and 4°C. The supernatant was applied in TLC analysis. For quantification, samples of 100 µL were taken and dissolved 1/10 in ethyl acetate/acetic acid 3:1. These acidified ethyl acetate samples were centrifuged at 10,000 g. The supernatant was used for quantitative TLC analysis as stated below.

Fosmid clones were grown in 96 deep well plates overnight. Clones were joined in 96, 48, eight or six clones per pool. The pools were harvested by centrifugation at 4,500 g and resuspended in 50 mL LB medium containing 12.5 µg/mL chloramphenicol, CopyControl™ Autoinduction Solution (Epicentre, Madison, WI) (5 mM arabinose final concentration) and 100 µM of flavonoid for biotransformation. Alternatively to deep well plates, clones were precultured on agar plates. After overnight incubation the colonies where washed off with 50 mM sodium phosphate buffer pH 7, harvested by centrifugation and resuspended as outlined above. The biotransformations were incubated in 300 mL Erlenmeyer flasks at 30°C with shaking at 175 rpm. Single clones were tested analogously but precultured in 5 mL LB and resuspended in 20 mL biotransformation media in 100 mL flasks. Samples of 4 mL were taken from the reactions after 16, 24 and 48 hours acidified with 40 µL HCl and prepared for TLC analysis as stated above. Positive pools were verified in a second biotransformation
and then systematically downsized to detect the corresponding hit in a smaller pool until the responsible single clone was identified.

Biocatalytic reactions of 1 mL contained 5 µg of purified His-tagged enzyme and were performed in 50 mM sodium phosphate buffer pH 7 at 37°C. UDP-α-D-glucose or UDP-α-D-galactose was added to final concentrations of 500 µM as donor substrate from 50 mM stock solutions in 50 mM sodium phosphate buffer pH 7. Acceptor substrates were used in concentrations of 100 µM and were added from stock solutions of 100 mM in DMSO leading to a final content of 0.1% in the reaction mixture. The reaction was stopped dissolving 100 µL reaction mixture 1/10 in ethyl acetate/acetic acid 3:1. These samples were used directly for quantitative TLC analysis.

**TLC analyses**

The supernatant transferred into HPLC flat bottom vials was used for TLC analysis. Samples of 20 µL were applied on 20×10 cm² (HP)TLC silica 60 F₂₅₄ plates (Merck KGaA, Darmstadt, Germany) versus 200 pmol of reference flavonoids. To avoid carryover of substances, i.e. prevent false positives, samples were spotted with double syringe rinsing in between by the ATS 4. The sampled TLC plates were developed in ethyl acetate/acetic acid/formic acid/water 100:11:11:27 (‘Universal Pflanzenlaufmittel’) (41). After separation the TLC plates were dried in an oven at 80°C for five minutes. The absorbance of the separated bands was determined densitometrically depending on the absorbance maximum of the applied educt substances at 285 to 370 nm using the deuterium lamp in a TLC Scanner 3 (CAMAG, Muttenz, Switzerland). Subsequently, the substances on developed TLC plates were stained by either dipping or spraying the plates in a methanolic solution 1% (w/v) of ‘Naturstoffreagenz A’, containing diphenyl boric acid β-aminoethyl ester (42), available from Carl Roth GmbH, Karlsruhe, Germany. Following this step the TLC plates were dried and then dipped in or sprayed with a 5% (w/v) solution of polyethylene glycol 4000 in ethanol (70%, v/v). For dipping a chromatogram Immersion device (CAMAG, Muttenz, Switzerland) was used. After two minutes the bands were visualized at 365 nm with a UV hand lamp and photographed. Alternatively, fluorescence of the bands was determined densitometrically by...
the TLC Scanner 3 depending on the absorbance maximum of the applied substances at 320 to 370 nm.

**Quantification of flavonoids by TLC**

To quantify flavonoids in biotransformation and biocatalytic reactions, samples were diluted 1/10 in ethyl acetate/acetic acid 3:1 to stop the reaction. Samples of 20 µL were sprayed by an ATS 4 (CAMAG, Muttenz, Switzerland) on HPTLC silica 60 F254 plates (Merck KGaA, Darmstadt, Germany) versus different amounts of respective standard educt and product substances. TLC plates were developed, dried, derivatized and analyzed as stated above. Regression curves were calculated from the peak area of the applied reference substances to determine the amount of produced and residual flavonoids.

**HPLC-ESI-MS analysis**

HPLC was carried out on a Purospher Star RP-18e 125-4 column (Merck, Darmstadt, Germany), particle size of 3 μm, with a Rheos 2000 pump (Flux Instruments, Suisse) and set pressure limits of 0 bar minimum and a maximum of 400 bar. Injection volumes of 10 µL were separated with solvent A, water supplemented with 0.1% TFA; and solvent B, acetonitrile with 0.1% TFA in following gradient HPLC conditions: From 0 min, 0.6 mL/min 90% A, 10% B; from 14 min, 0.6 mL/min 75% A, 25% B; from 18 min, 0.6 mL/min 5% A, B=95%; from 22 min, 0.6 mL/min 5% A, 95% B; from 22.1 min, 0.6 mL/min, 90% A, 10% B; and from 28.1 min, 0.6 mL/min 90% A, 10% B. Elution was monitored with a Finnigan Surveyor PDA detector and fractions were collected by a HTC PAL autosampler (CTC Analytics). Mass spectrometry (MS) was performed on a Thermo LCQ Deca XP Plus with an ESI interface in positive ionization.

**Sequence analysis and Genbank entries**

Automated DNA sequencing of small insert plasmids was performed using ABI377 and dye terminator chemistry following the manufacturer’s instructions. Large fosmid sequences were established by 454 sequencing technology. The sequences were assembled by using Gap 4 software. ORF finding was performed with Clone manager 9.
Professional software. All sequences mentioned in this work were deposited at GenBank.
The DNA sequences of the *Bacillus* sp. HH1500 16S rRNA gene has the GenBank accession number KC145729. The fosmid derived genes from *B*. sp. HH1500 identified on subclone pSK4B2 are *bsp*A (JX157885), *mgt*B (JX157886) and *bsp*C (JX157887) and have the GenBank accession numbers AGH18135 - AGH18137, respectively. The Elbe sediment metagenome derived fosmid subclone pSK144C11 comprised genes *esmA* (JX157626), *gtf*C (JX157627), *esm*B (JX157628), and *esm*C (JX157629). The deduced proteins have the GenBank numbers AGH18138 to AGH181341.
RESULTS

Screening method: Setup of a TLC-based screening method for the detection of flavonoid-modifying enzyme clones.

Since it is known that B. cereus and B. subtilis encode for glycosyltransferases mediating the glucosylation of flavonoids (36), we initially tested several single bacterial isolates from our strain collections with respect to their flavonoid modifying activities. Biotransformations using whole cells of wild type isolates confirmed the presence of flavonoid modifying enzymes in one of the strains. This strain was originally isolated from a soil sample of the botanical garden in Hamburg and was designated Bacillus sp. HH1500. Sequence analysis of a 16S rRNA gene (GenBank entry KC145729) showed a 100% identity to members of the B. cereus group (data not shown). In order to use this strain as a positive control, we constructed a fosmid library of its genomic DNA in pCC1FOS. The obtained library contained 1,920 clones with an average insert size of 35 kb. Thus, the library encompassed approximately 67 Mb of cloned gDNA hence covering the average size of a genome from B. cereus group members about ten times (43). Further, the sensitivity of the (HP)TLC-based assay was verified using a serial dilution of isoquercitrin, the 3-O-β-D-glucoside of quercetin, by spraying 10 µL of 0.78 µM up to 100 µM solutions of isoquercitrin on TLC plates and measuring the absorbance at 365 nm (TABLE S3). In addition, 10 µL of other glycosylated flavonoids were assayed at 10 µM concentrations and could be detected as clear peaks on the absorbance chromatograms (TABLE S3, and data not shown).

Based on the observed sensitivities, we designed a systematic screening scheme as outlined in FIGURE 1. Initially 96 fosmid clones were grown in deep well microtitre plates at 37°C overnight. Cultures were then pooled and following this step, the cells were sedimented by centrifugation and resuspended in fresh LB medium containing the appropriate antibiotics and 100 µM of quercetin as acceptor substrate. After incubation for 16, 24 and 48 hours at 30°C, 4 mL samples of the pooled cultures were withdrawn and extracted with half the volume of ethyl acetate. Of these extracts 20 µL were applied on TLC silica plates and
separated using ‘Universal Pflanzenlaufmittel’ as a solvent. The absorbance of the
developed sample lanes was determined densitometrically at 365 nm. Additionally, bands of
substrates and modified flavonoids were visualized by staining with ‘Naturstoffreagenz A’ as
outlined in the Material and Methods section. In our hands the sensitivity of the assay was
high enough to detect a single flavonoid modifying enzyme clone in a mixture of 96 clones
(FIGURE 1). After the detection of a positive signal, we divided the 96 fosmid clones into
pools of 48 to locate the same peak in one of the resulting two half microtitre plates.
Following this procedure, we divided the 48 clones to six times eight clones (FIGURE 1) and
finally analyzed the eight individual clones. This strategy was applied successfully to identify
six overlapping positive clones in the Bacillus sp. HH1500 fosmid library testing all 20
microtitre plates with 1,920 clones, totally.

Of these six fosmid clones, one clone pFOS4B2 of approximately 46 kb was
subcloned using the HindIII restriction site of pBluescript II SK+ vector. The obtained
subclones were analyzed using the above-mentioned TLC screening technology. Thereby, a
positive subclone designated pSK4B2 was identified and completely sequenced (GenBank
entry JX157885 - JX157887). Subclone pSK4B2 carried an insert of 3,225 bp (FIGURE S1A)
and encoded for a gene, designated mgtB, encoding for a 402 aa protein. The identified ORF
was subcloned creating plasmid pDMgtB and again assayed for activity. TLC analysis clearly
confirmed the glycosylation activity of the MgtB enzyme in this construct as well. The
deduced aa sequence of MgtB (GenBank # AGH18136) was highly similar to a predicted B.
thuringiensis macroside glycosyltransferase (TABLE 1). The mgtB-surrounding DNA
sequences in plasmid pSK4B2 represented two truncated genes that consistently were
almost identical to genes from B. thuringiensis (TABLE 1). This phylogenetic relation was in
accordance to the preliminary sequence analysis of the 16S rRNA gene of Bacillus sp.
HH1500 (see above).

These tests suggested that the screening procedure was suitable for the functional
screening of large insert metagenome libraries. For the function-based screening of
metagenomes we termed this methodology META: Metagenome Extract TLC Analysis.
Although it is not fully automated high-throughput screening (HTS) technology, META allows screening of about 1,200 clones per TLC plate within a time of 48 hours for preculture, biotransformation and analysis. This number of clones appeared to be feasible, if a single person did the screening. Generally, the sampling of about one TLC plate per hour by the ATS 4 is the time limiting step of the method. But this still allows the pooled screening of several plates a day and hence throughput of numerous thousand clones a day by META.

Identification of a novel glycosyltransferase from a metagenome library

To further apply the screening for enzyme discovery in metagenome libraries, we tested two fosmid libraries constructed in our laboratory. One library was constructed from DNA isolated from river Elbe sediment the other from isolated DNA out of fresh elephant feces. Altogether both libraries encompassed approximately 50,000 clones with an average insert size of 35 kb. Both libraries were screened using quercetin as a substrate. Using the described strategy we discovered one positive microtitre plate pool in the river Elbe-sediment-library. Further screening of this pool resulted in the identification of a single positive fosmid clone designated pFOS144C11 (FIGURE 2). Biotransformations of quercetin (Q) with 48 clone pools presented one product peak (P2) by TLC separation with an Rf value comparable to that of quercitrin, the quercetin-3-O-β-L-rhamnoside (FIGURE 2A). A second peak (P3) with a Rf value higher than the available reference quercetin glycones was observed in conversions with the six-clone-pool and the single fosmid clone, respectively (FIGURE 2B, C). Clone pFOS144C11 carried a fosmid of approximately 40 kb. Subsequent restriction fragment subcloning into pBluescript II SK+ with HindIII yielded in the identification of the positive *E. coli* DH5α subclone pSK144C11. However, biotransformations with pSK144C11 showed two product peaks, a major one (P2) with an Rf value comparable to that of quercitrin and a minor one (P1) similar to isoquercitrin (FIGURE 2D). The subclone pSK144C11 still had an insert of approximately 8.5 kb size. Further sequencing and subcloning of pSK144C11 finally identified the gene putatively responsible for the modifications which we designated *gtfC* (FIGURE S1B). The deduced 459 amino acid
sequence of the corresponding enzyme revealed motif similarities to UDP-glucuronosyl/UDP-glucosyltransferases. GtfC (GenBank # AGH18139) showed a similarity of 71% to the putative glycosyltransferase of the Gram-negative bacterium *Fibrisoma limi* covering 92% of the protein (TABLE 1). Further cloning of the *gtfC* ORF into pDrive vector and biotransformation with *E. coli* DH5α carrying the respective construct pDgtfC confirmed the flavonoid-modifying activity of GtfC (FIGURE 2E).

In summary, these results demonstrated that the developed screening procedure META is sufficiently sensitive to allow the identification of large insert clones from individual bacterial genomes (i.e. *Bacillus* sp. HH1500) and complex metagenome libraries (i.e. the river Elbe sediment library) showing flavonoid-modifying activities.

**Sequence based classification of MgtB and GtfC**

To analyze the affiliation of MgtB and GtfC, we calculated a phylogenetic tree using the MEGA version 5 software (44). The amino acid (aa) sequences of MgtB and GtfC, and their closest sequence-based relatives determined by pBlast were aligned by ClustalW. Additionally, the sequences of the actually published prokaryotic flavonoid active GTs were aligned and finally as an outer group two eukaryotic enzymes, the flavonoid glucosyltransferase UGT85H2 from *Medicago truncatula* and the flavonoid rhamnosyltransferase UGT78D1 from *Arabidopsis thaliana* (45, 46). Thereof a neighbor-joining tree with 100 bootstraps was computed (FIGURE 3). As expected, MgtB from *Bacillus* sp. HH1500 clustered with other MGTs from the *B. cereus* group. At time of writing, the MGT of *B. thuringiensis* IBL 200 and the MGT of *B. cereus* G9842 turned out to be the closest relatives with an aa identity to MgtB of 98 % each. Both MGTs were annotated as predicted enzymes and no substrate data were available. From the MGT cluster five other enzymes already were reported to mediate the glucosylation of flavonoids. Three of them BcGT-1 the nearest relative reported to be flavonoid active, BcGT-4, and BcGT-3 all originated from *B. cereus* ATCC10987 (47-49). Another flavonoid active MGT, designated BsGT-3, originates from *B. subtilis* strain 168 (36). The remaining flavonoid active MGT is the well-studied OleD from *Streptomyces antibioticus* (50, 51). GtfC was located in a distinct cluster of UGTs and...
appeared to be somewhat related to hypothetical enzymes from *Cytophagaceae* bacteria as *Dyadobacter fermentans* and *Fibrisoma limi* (FIGURE 3). Within this cluster only the UGT XcGT-2 is known to accept flavonoid substrates (38). Interestingly, rhamnosyltransferases like BSIG 4748 from *Bacteroides* sp. 116 and RtfA from *Mycobacterium avium* phylogenetically also show affiliation to this cluster but forming a separate branch (FIGURE 3).

To further characterize the identified metagenome-derived GTs, the aa residues of the C-terminal donor binding regions were compared to the motifs of the closest relatives and the known flavonoid active GTs (FIGURE 4). Here, the Rossmann fold α/β/α subdomain, the conserved donor-binding region of UGTs, is located (52). Plant UDP-glycosyltransferases like UGT85H2 and UGT78D1 exhibit a highly conserved motif in this region which is termed the (Plant Secondary Product Glycosyltransferase) PSPG motif (45, 46, 53, 54). By alignment we were able to identify key aa known to be of importance for NDP-sugar binding. While MgtB revealed a clear UDP-hexose binding motif consisting of highly conserved Gln289 and Glu310 residues for ribose binding and a conserved DQ, GtfC lacked this motif (45, 55, 56). Instead, GtfC presented typical residues Phe336 and Leu357 for deoxy-ribose nucleotide utilization (57). Moreover we were able to identify the pyrophosphate binding sites in the MgtB aa sequence (FIGURE 4). However, GtfC does not possess these conserved phosphate binding residues suggesting that GtfC and related enzymes have another donor binding mode. In this context GtfC seemed to belong to a novel enzyme class underlining the low level of sequence homology.

**Overexpression and glycosylation patterns of MgtB and GtfC**

To further characterize the novel enzymes and verify their functions, we overexpressed and purified MgtB and GtfC as His-tagged proteins in *E. coli* BL21 (DE3). Both genes *mgtB* and *gtfC* were ligated into the expression vector pET19b. The recombinant enzymes containing N-terminal His10-tags were purified by Ni-affinity chromatography in native conditions and gradient elution. MgtB could be purified with more than 5 mg/g cell pellet (wet weight). The maximum yield of GtfC was 3 mg/g of cell pellet. The molecular
weights of the proteins were verified by SDS-PAGE analysis in denaturing conditions according to Laemmli. After Coomassie-staining, His<sub>10</sub>-MgtB was visible as a single band with a MW of approximately 50 kDa on a 12% SDS-PAGE (FIGURE S2A). This was in accordance with the calculated molecular weight (MW) of 51.2 kDa including the N-terminal His-tag. His<sub>10</sub>-GtfC revealed a MW of about 55 kDa on a 12% SDS-PAGE which was in well accordance to the calculated MW of 54.7 kDa including the N-terminal His-tag. While virtually no additional bands were visible on SDS-PAGEs with purified recombinant MgtB protein, some minor contaminating bands were still visible on the SDS-PAGE loaded with purified GtfC (FIGURE S2B). In summary both proteins could be purified to allow further biochemical characterization.

The purified His<sub>10</sub>-MgtB protein was able to use UDP-α-D-glucose as a donor substrate. The recombinant enzyme catalyzed the transfer of α-D-glucose residues to various polyphenols. Biocatalytic reactions were performed with 500 µM UDP-α-D-glucose as donor and 100 µM of acceptor substrate. The following flavonoids served as acceptor substrates and were modified with high yields: Luteolin, quercetin, kaempferol, tiliroside, naringenin, genistein (TABLE 2). Thereby flavonols turned out to be the best acceptor molecules. Generally, the conversion during a two-hour assay ranged from 52% for naringenin and approximately 100% for quercetin and kaempferol. Interestingly, in the presence of quercetin and kaempferol no residual educts could be monitored by HPTLC analysis. The specific educts and their observed glycones of the biocatalytic reactions are summarized in TABLE 2 together with the respective Rf values. MgtB favored the glucosylation at the C3 hydroxy group if accessible like in the aglycone flavonols quercetin and kaempferol. Further, the C7-OH was attacked and glucosylated by the enzyme which could be shown for the flavone luteolin but also the flavanone naringenin and the isoflavone genistein (TABLE 2). MgtB glucosylated luteolin also at the C3' hydroxy group forming the 3',7-di-O-glucoside of luteolin if the C7-OH was glucosylated previously. MgtB also catalyzed the conversion of the kaempferol derivative tiliroside, the kaempferol-3-(O-6''trans-p-coumaroyl)-glucoside. One glucosylated product with an Rf values of 0.54 was detected.
Finally, the chalcone xanthohumol and the stilbene t-resveratrol were tested in biotransformation reactions with *E. coli* expressing *mgtB* but conversions were not quantified (data not shown). Xanthohumol yielded three detectable products whereas the biotransformation of t-resveratrol yielded one observed product by absorbance TLC analysis.

Tests with recombinant and purified GtfC using UDP-α-D-glucose and UDP-α-D-galactose and quercetin as acceptor molecule suggested that dTDP-activated sugar moieties were transferred by this enzyme. This finding was confirmed by HPLC-ESI-MS analyses of biotransformation assays (see following paragraph). Unfortunately, deoxy-ribose nucleotide activated hexoses e.g. dTDP-rhamnoside were commercially not available to further analyze the obtained reaction products in more detail (58).

Biotransformations with the *E. coli* strain expressing GtfC and using various polyphenols as substrates yielded in conversions ranging from 52% for xanthohumol up to almost 100% turnover for most flavonols tested (TABLE 3). Quercetin was transformed almost completely after four-hour biotransformations and yielded three detectable products (P1-P3). To further characterize these products UV absorbance spectra were recorded and compared to the reference glycones of quercetin isoquercitrin and quercitrin (59). P1 revealed an Rf value identical to the value of isoquercitrin. Further the UV absorbance spectrum of P1 matched the spectrum of isoquercitrin (FIGURE S3A). P2 revealed an Rf value identical to the one known for quercitrin. P2 also exhibited the same UV absorbance spectrum as quercitrin (FIGURE S3B). P3 revealed an Rf value of 0.82, which clearly differed from the RF values of known and available quercetin glycones. Compared to isoquercitrin, P3 showed a similar hypsochromic shift of band I to a $\lambda_{max}$ of 363 nm (FIGURE S3C); however it revealed a less hypsochromic shift in band II of only 5 nm to 272 nm with a shoulder at 280 nm. It is further notable that the HPLC-ESI-MS analysis of biotransformation products of quercetin consistently identified three distinct reaction products (FIGURE S4). P1 had a RT of 17.93 min in the HPLC analysis and revealed a molecular mass of 464 u, which is equivalent to isoquercitrin. P2 revealed a RT of 18.06 min and had a molecular mass of 448 u. This mass corresponds well with the molecular mass of quercitrin. Finally, P3 with a
RT of 18.31 min revealed a molecular mass of 446 u indicating the formation of a novel not further characterized quercetin glycoside.

Glycosylation patterns of GtfC on quercetin suggested a preference to act on the C3 hydroxy group mediating the transfer of different sugar residues. However, if a C3 OH-group was not available, GtfC efficiently catalyzed the glycosylation of other positions. Flavones lacking the hydroxy function at C3 were converted depending on the availability of other hydroxy groups. Pratol possessing only a single free C7-hydroxy group was converted weakly and resulted in a single detectable product. Further the biotransformation of 3',4'-dihydroxyflavone yielded three detectable glycones and 5-methoxy-eupatorin yielded two products (data not shown, TABLE S4); the biotransformation of the mono 4'-hydroxyflavanone yielded one glycosylated product and the glycosylation of naringenin yielded two products. The major biotransformation product of naringenin revealed the same Rf value and absorbance spectrum as prunin, the naringenin-7-O-glucoside (TABLE 3). The second naringenin glycone could not be further specified due to the lack of commercially available reference substances. Altogether these results suggested that GtfC acts on the C3, C3’, C4’ and C7 hydroxy groups of the flavonoid backbone.

In summary these data demonstrated that MgtB and GtfC possess interesting biocatalytic properties. While MgtB specifically mediated the transfer of glucose residues, GtfC transferred different hexose moieties. MgtB was capable to catalyze the glucosylation of already glycosylated flavonoids to form di-glycosides (e.g. formation of luteolin-3’,7-di-O-glucoside) and even tiliroside to generate novel glucosides not available from natural resources. In contrast, the glycosylation pattern of GtfC suggested the transfer of single sugar residues to only aglycone flavonoid forms. Interestingly, GtfC seemed to be very variable concerning its activity at various positions on the flavonoid backbone. This may lead to the formation of truly novel flavonoids naturally not available. Hence both enzymes might be helpful in the generation of new natural compounds.
Within this manuscript, we report on the development of a semi-automated TLC-based detection system for flavonoid-modifying enzyme clones. The screening assay was highly reproducible and highly sensitive. It allowed the detection of micromolar amounts of glycosylated flavonoids. Isoquercetin was detectable at 0.78 micromoles in the assay (TABLE S3). Using this assay we were able to systematically identify one positive clone out of pools of 96 metagenome clones (FIGURE 1 & 2). To our knowledge this is the first published TLC-based screening method for functional searches in metagenome libraries. We speculate that slight modifications of this screening system could easily allow the detection of other flavonoid-modifying enzyme clones such as acylation or methylation reactions.

Using this novel screening technology, we identified a macroside glycosyltransferase MgtB from a soil isolate (i.e. Bacillus sp. HH1500). A fosmid library established with DNA from this strain, which had been isolated from the local botanical garden, only recently, was initially used to develop and verify the outlined screening technology; and using the novel screening technology, MgtB was quickly identified from a pool of almost 2,000 clones. Isolation and purification of recombinant MgtB revealed a novel MGT. MgtB shared 89% aa identity with BcGT-1 from B. cereus ATCC 10987, the closest relative published to act on flavonoids. BcGT-1 was reported to catalyze the glucosylation of flavones, flavonols, flavanones and isoflavones (47). On flavonols BcGT-1 acted on C3-, C7- and C4’- hydroxy groups creating triglucosides of kaempferol (48). In contrast biocatalyses of kaempferol with MgtB yielded just two detectable glucosylated products. Instead reactions with quercetin resulted in three detectable glycones. These data suggested that MgtB acted at the C3’ OH-group. This hypothesis was also was supported by the observation that recombinant MgtB converted luteolin to luteolin-3’,7-di-O-glucoside as a byproduct. These results were in accordance with the glucosylation pattern of BcGT-3 yet another MGT from B. cereus ATCC10987 (49). Interestingly, BcGT-3 shares only 40% aa identity with MgtB but both enzymes act on the same flavonoids forming di-glucosides from flavones and flavonols at the same positions and only mono-glucosides from naringenin. The most spectacular conversion...
observed for MgtB was that of tiliroside. The product is likely to be the 7-O-glucoside taking
the glycosylation pattern of MgtB into account. Tiliroside glycosides yet were not reported in
scientific literature. This raises the possibility of the generation of new natural compounds.
The natural substrates of Bacillus MGTs still have not been reported. Other MGTs like OleD
usually detoxify macrolide antibiotics but often possess broad acceptor tolerance (35, 60).

The metagenome-derived GtfC turned out to be a completely novel enzyme. Only
seven flavonoid-active UGTs have been reported so far that originate from five different
prokaryotes (35, 36, 38, 47, 49). Without XcGT-2 from Gram-negative X. campestris ATCC
33913 all remaining are MGT enzymes from Gram-positive Bacilli and Streptomycetes.
MGTs play an important role in xenobiotic defense mechanisms of prokaryotes and thus
show broad acceptor specificities (55, 60). This also applies for eukaryotic UGTs pointing to
a biological principle of detoxification (61). To our knowledge GtfC is the first metagenome-
derived GT acting on flavonoids. Moreover, it is also the first bacterial enzyme reported to
transfer various dTDP activated hexose sugars to polyphenols (see below) in contrast to
usually stringent donor specificities like Gtfs (57). With respect to the notion that many NDP-
sugars in prokaryotes are dTDP and not UDP activated, GtfC might be a promising
biocatalyst in glycodiversification approaches (58, 62, 63). GtfC is similar to predicted GTs
from Cyphagaceae bacteria (64-66). These Gram-negative bacteria have large genomes
suggesting extensive secondary metabolic pathways and they are well known for the
presence of resistance mechanism to antibiotics as trimethoprim and vancomycin (67, 68).
As commonly known glycosylation of xenobiotics is a ubiquitous detoxification process in all
kingdoms of life. The phylogenetically divers members of Cytophagaceae have only recently
become an object of research and a concrete estimation about the phylogenetic wideness of
this family and exact taxonomic ranking still remain unclear (65, 69). Thus, the identification
of the metagenome-derived GtfC and its partial characterization suggest that this group of
microorganisms is perhaps highly promising resource for novel GTs and also other enzymes.
A ClustalW alignment of the donor-binding region of GtfC suggested the activated
donor substrates are of deoxy-thymidine nucleoside origin. GtfC possesses the typical aa
residues Phe336 for thymine base stacking and hydrophobic Leu357 for deoxy-ribose fitting (57). Concerning the donor binding of GTs GtfC appears to not exhibit the known aa residues for pyrophosphate binding (FIGURE 4). Instead of the conserved residue His/Arg in the up to date solved protein structures GtfC contains an Asn at the aa position 349 (52, 70). This applies also for the nearest GtfC relatives Dfer1940, UGT of F. limi BUZ 3 and Slin3970 as well as the NGTs RebG and BSIG4748. Further, GtfC does not show the conserved Ser/Thr residue responsible for α-phosphate binding. Instead the Gly354 appears to be of importance for the α-phosphate binding similar to the OleD transferase (55).

The assumption of dTDP activated co-substrates used by GtfC was supported by the observation that glucose, rhamnose and a third sugar residue with molecular weight of 446 were transferred by GtfC in biotransformations using intact E. coli cells. Besides, biocatalytic approaches with purified GtfC and either UDP-α-D-glucose or -galactose as donor substrates failed. In bacteria, the activated sugars, dTDP-α-D-glucose, -4-keto-6-deoxy-α-D-glucose or -4-keto-β-L-rhamnose, and -β-L-rhamnose are part of the dTDP-sugar biosynthesis pathway and are present in E. coli (71). Moreover, levels of dTDP-sugars are allosterically regulated by dTDP-rhamnose levels through activity of RmA (72).

In summary, the screening protocol developed within this manuscript is a very helpful tool for the identification of truly novel enzymes for the modifications of flavonoids and related substrates. Within the manuscript we have used the technology to identify two novel flavonoid-modifying enzymes. Both these enzymes would perhaps not have been detected without the above-developed screening technology. The partial biochemical characterizations either using biocatalysis or biotransformation suggest that MgtB and GtfC are both very interesting enzymes with a high potential for biotechnological applications with respect to flavonoid modifications at industrial scale. Thus future work will now refine this technology to also identify other enzymes linked to flavonoid modifications.
ACKNOWLEDGEMENTS

This work was funded by the ministry of education and research „Bundesministerium für Bildung und Forschung“ (BMBF) within the network program ‘Biokatalyse2021’ as project P18 (FKZ: 0315166C). The Bacillus Genetic Stock Center (BGSC, Columbus, OH, USA) kindly provided *Bacillus cereus* ATCC 10987 used as positive control.

REFERENCES


FIGURE LEGENDS

FIGURE 1. Outline of the metagenome screening for flavonoid-modifying clones. Schematic workflow of the function-based screening procedure for the systematic identification of flavonoid-modifying clones. Initially pools of 96 clones preincubated in deep-well plates were tested in biotransformation reactions with quercetin followed by culture extract TLC analysis as described in MATERIAL & METHODS. Once a positive pool forming a product peak was observed, the respective plate was divided in two half-plate pools of 48 clones each, which were tested similarly. The positive fraction was divided into six times eight clones per column (or vice versa 8 x 6 per row) that were tested for activity, until the candidate single clones were tested and the positive was identified. Putative positive clones were then reconfirmed and subcloned for further analysis. TLC analysis UV-chromatograms are displayed in relative absorbance units (AU) over Rf value measured at 365 nm in a densitometrical TLC Scanner 3 (CAMAG, Muttenz, Switzerland). Light grey peaks indicate remaining substrate quercetin (Q) absorbance, dark grey peaks (P1) the formed product isoquercitrin. Depicted chromatograms show UV-chromatograms of culture extracts in EtOAc after 24 hours biotransformations of pools with 96 (top), 48 (2nd row), and eight clones (3rd row) including fosmid clone pFOS19G2 and of single clone pFOS19G2 (last).

FIGURE 2. Iterative TLC analyses of culture extracts for the identification of a flavonoid modifying enzyme. Cells were grown in LB media with appropriate antibiotics. Biotransformation was performed according to MATERIAL & METHODS. Culture extracts in EtOAc from 24 hours biotransformation reactions with 100 µM quercetin as a substrate were applied on Merck Silica 60 F 254 TLC plates. UV-chromatograms are displayed in relative absorbance units (AU) to Rf value measured at 365 nm on a densitometrical TLC Scanner 3 (CAMAG, Muttenz, Switzerland) for activity determination. Remaining substrate quercetin peaks are depicted in light grey near the solvent front (Q); product peaks are colored in dark grey (P1, P2 and P3). TLC analyses of culture extracts from following biotransformations led to the final isolation of the GT encoding ORF gtfC: (A) Pool MT144R of 48 fosmid clones.
within the positive clone pFOS144C11, (B) Pool MT144C of six fosmid clones within the positive clone, (C) Positive single fosmid clone pFOS144C11, (D) Positive subclone pSK144C11 derived from pFOS144C11, (E) Active ORF gtfC derived from pFOS144C11 in clone pDgtfC.

FIGURE 3. Phylogenetic dendrogram of glycosyltransferases (GTs) related to the two GTs (MgtB and GtfC in black boxes) identified in this study. GTs known to act on flavonoids as acceptors are highlighted in grey. Phylogenetic analysis was conducted using MEGA version 5 (44) with ClustalW sequence alignment in a BLOSUM protein weight matrix. The neighbor-joining tree was calculated using bootstrap method with the Poisson model, bootstrap values higher 75 are indicated next to the branches. The scale represents amino acid changes per residue. All shown GTs belong to the family of Gtf_like GT1 and are further subclassified into MGTs, macroside glycosyltransferases (TIGR01426), NGTs, N-glycosyltransferases and UGT, a UDP-glucuronosyl/-glucosyltransferases (PF00201) as indicated on the right. GenBank numbers were retrieved from NCBI as following: YP002445489, MGT from B. cereus G9842; ZP04071678, MGT from B. thuringiensis IBL 200; AGH18136, MgtB from Bacillus sp. HH1500; AAP25969 B. anthracis strain Ames; AYB43166, MGT from B. weihenstephanensis KBAB4; NP978481, BcGT-1 from B. cereus ATCC 10987; NP979441, BcGT-4 from B. cereus ATCC 10987; AAS41737, BcGT-3 from B. cereus ATCC 10987; NP389104, BsGT-3 from B. subtilis strain 168; ABA42119, OleD from Str. antibioticus; AAM41712, XcGT-2 from X. campestris ATCC 33913; YP003086330, Dfer 1940 from Dyadobacter fermentans DSM 18053; AGH18139, GtfC from Elbe river sediment metagenome; CCH52088, UGT from Fibrisoma limi BUZ 3; YP003388759, Slin 3970 from Spirosona lingua DSM 74, ACV78946, Namu2594 from Nakamurella multipartita DSM 44233, YP712191, FRAAL1959 from Frankia alni ACN14a; AAN01207, RebG from Lechevalieria aerocolonigenes; ZP09941874, BSIG 4748 from Bacteroides sp. 116; AAC71702, RtfA from Mycobacterium avium; AFK05536, Emtol 0266 from Emticicia...
oligotrophica DSM 17448; 2PQ6A, UGT85H2 from Medicago truncatula, NP564357, UGT78D1 from Arabidopsis thaliana.

**FIGURE 4.** ClustalW alignment of the MgtB and GtfC (black boxes) aa sequences, their nearest sequence based relatives and other flavonoid active GTs (grey boxes). The window shows the C-terminal region of the Rossmann fold α/β/α subdomain, the conserved donor-binding region of UGTs. Plant UDP-glycosyltransferases like UGT85H2 (2PQ6A) from Medicago truncatula and UGT78D1 (NP564357) from Arabidopsis thaliana exhibit the highly conserved the (Plant Secondary Product Glycosyltransferase) PSPG motif in this region. Marked aa boxes indicate the residues’ role in donor nucleoside (blue), phosphate group (green) and hexose (red) binding according to given references (45, 52, 55-57, 73, 74).

GenBank numbers were retrieved from NCBI as following: AGH18136, MgtB from Bacillus sp. HH1500; ZP04071678, MGT from B. thuringiensis IBL 200; YP002445489, MGT from B. cereus G9842; ABY43166, MGT from B. weihenstephanensis KBAB4; AAP25969, MGT from B. anthracis strain Ames; NP978481, BcGT-1 from B. cereus ATCC 10987; NP979441, BcGT-4 from B. cereus ATCC 10987; AAS41737, BcGT-3 from B. cereus ATCC 10987; NP389104, BsGT-3 from B. subtilis strain 168; ABA42119, OleD from Streptomyces antibioticus; AAM41712, XcGT-2 from X. campestris ATCC 33913; AAK31352, GtfD from Amycolatopsis orientalis; AGH18139, GtfC from Elbe river sediment metagenome; YP003086330, Dfer 1940 from Dyadobacter fermentans DSM 18053; CCH52088, UGT from Fibrisoma limi BUZ 3; YP003388759, Slin 3970 from Spirosoma linguale DSM 74; ACV78946, Namu2594 from Nakamurella multipartita DSM 44233; YP712191, FRAAL1959 from Frankia alni ACN14a; AAN01207, RebG from Lechevalieria aerocolonigenes; ZP09941874, BSIG 4748 from Bacteroides sp. 116; AAC71702, RtfA from Mycobacterium avium; AFK05536; Emtol 0266 from Entmicicia oligotrophica DSM 17448; NP564357.
**CONTENT of TABLES**

**TABLE 1:** Open reading frames (ORF) identified on subclones pSK4B2 derived from the active *Bacillus* sp. HH1500 fosmid clone and pSK144C11 derived from the river Elbe sediment active fosmid clone.

<table>
<thead>
<tr>
<th>Subclone</th>
<th>ORF</th>
<th>AA</th>
<th>Homolog</th>
<th>Coverage (%)</th>
<th>% Identity/Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSK4B2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bspA</td>
<td>221</td>
<td>putative protein kinase</td>
<td>B. thuringiensis (ZP04101830)</td>
<td>100</td>
<td>99 / 99</td>
</tr>
<tr>
<td>mgtB</td>
<td>402</td>
<td>macrolide glycosyltransferase</td>
<td>B. thuringiensis (ZP04071678)</td>
<td>100</td>
<td>98 / 99</td>
</tr>
<tr>
<td>mgtC</td>
<td>261</td>
<td>hypothetical membrane protein</td>
<td>B. thuringiensis (ZP00741215)</td>
<td>100</td>
<td>99 / 100</td>
</tr>
<tr>
<td>pSK144C11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>esmA</td>
<td>80</td>
<td>putative UDP-NAc-muramate-L-alanin-ligase</td>
<td>Niabella soli (ZP09632598)</td>
<td>99</td>
<td>69 / 80</td>
</tr>
<tr>
<td>gtfC</td>
<td>459</td>
<td>putative UDP-glucosyltransferase</td>
<td>Fibrisoma limi (CCH52088)</td>
<td>92</td>
<td>51 / 71</td>
</tr>
<tr>
<td>esmB</td>
<td>170</td>
<td>hypothetical protein</td>
<td>Niastella koreensis (YP005009630)</td>
<td>95</td>
<td>63 / 77</td>
</tr>
<tr>
<td>esmC</td>
<td>150</td>
<td>putative membrane protein</td>
<td>Solitalea canadensis (YP006258217)</td>
<td>98</td>
<td>68 / 81</td>
</tr>
</tbody>
</table>
TABLE 2: Flavonoid substrates converted by recombinant MgtB in bioassays. Reactions of 1 mL were carried out at 37 °C for 2 hours in triplicate with 500 µM UDP-glucose, 100 µM of the respective flavonoid and 5 µg/mL of purified and recombinant MgtB.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Conversion (%)</th>
<th>Rf value a</th>
<th>Product(s) b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quercetin</td>
<td>~100%</td>
<td>0.79</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.64</td>
<td>Isoquercitrin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.27</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.25</td>
<td>-</td>
</tr>
<tr>
<td>Kaempferol</td>
<td>~100%</td>
<td>0.74</td>
<td>Astragalin</td>
</tr>
<tr>
<td>Luteolin</td>
<td>82%</td>
<td>0.65</td>
<td>Cynaroside</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.32</td>
<td>-3',7-di-O-Glc</td>
</tr>
<tr>
<td>Naringenin</td>
<td>52%</td>
<td>0.76</td>
<td>Prunin</td>
</tr>
<tr>
<td>Genistein</td>
<td>72%</td>
<td>0.69</td>
<td>Genistin</td>
</tr>
<tr>
<td>Tiliroside</td>
<td>83%</td>
<td>0.54</td>
<td>-</td>
</tr>
</tbody>
</table>

a Rf values and products in **bold** indicate the main product of the biocatalytic reactions.

b Products symbolized by “—” were not specified due to unavailable reference substances.
TABLE 3: Flavonoid substrates and products of biotransformation assays with recombinant GtfC. Quantification of the reaction was performed as stated in MATERIALS & METHODS. Triplicate reactions of 50 mL were performed in 50 mM sodium phosphate buffer (PB) pH 7.0 containing 1% (w/v) glucose and 200 µM of flavonoid at 30 °C.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Conversion (%)</th>
<th>Rf value&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Product(s)&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luteolin</td>
<td>86</td>
<td>0.81</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.73</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.68</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.58</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.58</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.51</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.46</td>
<td>-</td>
</tr>
<tr>
<td>Quercetin</td>
<td>~100</td>
<td>0.75 Quercitrin</td>
<td>Isoquercitrin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>Kaempferol</td>
<td>~100</td>
<td>0.80 Astragalin-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.85</td>
<td>-</td>
</tr>
<tr>
<td>Naringenin</td>
<td>76</td>
<td>0.84 Prunin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Genistein</td>
<td>68</td>
<td>0.76 Genistin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.83</td>
<td>-</td>
</tr>
<tr>
<td>t-Resveratrol</td>
<td>96</td>
<td>0.77 Genistin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.64</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.58</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.51</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.46</td>
<td>-</td>
</tr>
<tr>
<td>Xanthohumol</td>
<td>52</td>
<td>0.85</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.48</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>a</sup> Rf values and products in bold indicate the main product of the biotransformation reactions.

<sup>b</sup> Products symbolized by “–” were not specified due to unavailable reference substances.