Flavonoid-modifying capabilities of the gut microbiome – an in silico study

Tobias Goris (✉ tobias.goris@dife.de)
German Institute of Human Nutrition Potsdam-Rehbruecke: Deutsches Institut fur Ernahrungsforschung
Potsdam-Rehbrucke
https://orcid.org/0000-0002-9977-5994

Rafael Cuadrat
German Institute of Human Nutrition Potsdam-Rehbruecke: Deutsches Institut fur Ernahrungsforschung
Potsdam-Rehbrucke

Annett Braune
German Institute of Human Nutrition Potsdam-Rehbruecke: Deutsches Institut fur Ernahrungsforschung
Potsdam-Rehbrucke

Research Article

Keywords: flavonoids, metagenome-assembled genomes (MAGs), biotechnological applications

DOI: https://doi.org/10.21203/rs.3.rs-588616/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Flavonoids are a major group of dietary plant polyphenols and have a positive health impact, but their modification and degradation in the human gut is still widely unknown. Due to the rise of human gut metagenome data and the assembly of hundreds of thousands of bacterial metagenome-assembled genomes (MAGs), large-scale screening for potential flavonoid-modifying enzymes is now feasible. With sequences from characterized flavonoid-transforming enzymes as queries, the Unified Human Gastrointestinal Protein catalog was analyzed and quantification of putative flavonoid-modifying enzymes was carried out. The results revealed that flavonoid-modifying enzymes are often highly abundant in bacteria hitherto not considered as flavonoid-modifying gut bacteria. The enzymes for the physiologically important daidzein to equol conversion, well studied in *Slackia isoflavonicvertens*, were encoded only to a low extent in Slackia MAGs, but more abundant in *Adlercreutzia equolifaciens* and an uncharacterized *Eggerthellaceae* species. In addition, a high abundance of genes with a similarity of only about 35% in uncultivated Collinsella species suggest a hitherto uncharacterized Daidzein-to-equol potential in these bacteria. Of all potential flavonoid modification steps, *O*-deglycosylation (including derhamnosylation) was by far the most abundant in this analysis. In contrast, enzymes putatively involved in *C*-deglycosylation were detected less often in human gut bacteria and mainly found in *Agathobacter faecis* (formerly *Roseburia faecis*). Phloretin hydrolase, flavanonol/flavanone-cleaving reductase and flavone reductase (all three most abundant in *Flavonifractor plautii*) and *O*-demethylase (*Intestinibacter bartlettii*) homologs were of intermediate prevalence (several hundreds of MAGs). This first comprehensive insight into the black box of flavonoid modification in the human gut highlights many hitherto overlooked and uncultured bacterial genera and species as key organisms in flavonoid modification by the human gut microbiota. This could lead to a significant contribution to future biochemical-microbiological investigations on gut bacterial flavonoid transformation. In addition, our results are important for individual nutritional recommendations and for biotechnological applications which rely on novel enzymes catalyzing potentially useful flavonoid modification reactions.

Introduction

Flavonoids, which are plant secondary metabolites exclusively, are considered to bear many beneficial effects on health [1-5] and are assumed to contribute especially to a lower cardiovascular disease and cancer-related mortality [6, 7]. Most flavonoids are taken up as a part of human diet mainly as glycosides, with a few exceptions, such as flavan-3-ols in tea or cocoa. A portion of dietary flavonoids is modified in the small intestine, where e.g. *O*-deglycosylation by lactase-phlorizin hydrolase secreted by epithelial cells can occur. The produced aglycons may enter epithelial cells through passive diffusion. Absorbed flavonoids then undergo phase II transformation such as glucuronidation, methylation, and sulfonation [8-10] before entering the bloodstream. Another part of dietary flavonoids (or their phase I/II metabolites) reach the large intestine, where they are subject to transformation by gut bacteria [11-13]. Especially members of the genera *Bifidobacterium* and *Lactobacillus* were observed to *O*-deglycosylate, e.g. flavanone and isoflavone glycosides [11]. While *O*-deglycosylation was the most extensively described flavonoid deglycosylation in the literature with many glycosides from flavonols (e.g. rutin in several vegetables), flavanones (often found in citrus fruits) and isoflavones (present mainly in soy products) as substrates, several bacteria from the phyla *Coriobacteriaceae* and *Lachnospiraceae* families also mediate *C*-deglycosylation from compounds such as
isovitexin, e.g. from buckwheat. Additional flavonoid transformations involve demethylation, dehydroxylation, reduction, and C-ring cleavage [11].

Many described gut bacterial flavonoid-converting activities lack biochemical or genetic analyses and therefore sequences of the involved enzymes. Of the O-deglycosylating enzymes, six beta-galactosidases [14, 15] and a set of flavonoid-active rhamnosidases from *Bifidobacterium* spp. [16, 17], *Bacteroides thetaiotaomicron* [18], *Lactobacillus* spp. [19, 20] and taxonomically unresolved enzymes isolated from stool samples [21] have been characterized. Elucidation of the more complex C-deglycosylation pathway was tackled in *Eubacterium cellulosolvens*, *Catenibacillus scindens* and strain PUE (likely a bacterium of the *Dorea* genus), but not completely resolved [22-24]. Phloretin hydrolysis as carried out by phloretin hydrolase [25, 26] and the three-step conversion of daidzein to equol [27-29] are relatively well studied, while gut bacterial enzymes involved in dehydroxylation and O-demethylation of flavonoids were not yet described in the literature. However, an O-demethylase from *Eubacterium limosum* ZL-II demethylates secoisolariciresinol, a polyphenolic lignan, to 4,4′-dihydroxyenterodiol [30]. Like other characterized O-demethylases [31, 32], this enzyme system consisted of two methyltransferases (MT), a corroinoid protein (CP) and an activating enzyme (AE). Since the type strain of *E. limosum* and another species, *E. ramulus*, were observed to O-demethylate flavonoids [33, 34], the characterized O-demethylase might also carry out flavonoid O-demethylation.

The six flavonoid O-deglycosylating hydrolases characterized so far were found in *Bifidobacterium animalis*, *Bi. pseudocatenulatum* and *Bacteroides thetaiotaomicron*. [14, 15, 35]. Several flavonols and isoflavones (e.g. genistin as a major flavonoid of soy) were deglycosylated by these beta-glucosidases. Most of them show a high promiscuity, deglycosylating also several polysaccharides and several glycosylated phenolic compounds. As a subgroup of flavonoid O-deglycosylating enzymes, rhamnosidases cleave usually the terminal rhamnose subunit of flavonoid rhamnosides (e.g. rutin or naringin), resulting in the corresponding flavonoid glucosides [36]. In contrast to these relatively simple one-step hydrolysis reactions, C-deglycosylation is dependent on several enzymes for which two different operons were described. The C-deglycosylation system described in strain PUE is dependent on at least three enzymes, DgpABC, which are part of a large gene cluster including transporters and other accessory genes whose function in C-deglycosylation is unclear [23, 24]. The C-deglycosylating enzyme system of *Eubacterium cellulosolvens* includes the five enzymes DfgABCDE [22]. Both corresponding enzyme complexes include at least one glycosyltransferase and an oxidoreductase.

Daidzein-to-equol conversion in the human gut has been detected in many individuals (estimated as one half of the population) [37-39]. The involved enzymes and their encoding genes have been characterized in several different species such as *Slackia isoflavoniconvertens*, *Slackia* sp. NATTS, *Eggerthella* sp. YY7918 and *Lactococcus garviae* [27, 29, 40, 41]. Equol formation in the human gut is important, since this compound is more easily absorbed and has stronger estrogenic activities than daidzein [42]. At least three enzymes are involved in the complete conversion of daidzein to equol. All three were initially proposed to be reductases [27, 29], although one of them was suggested to be a dismutase recently [40]. In addition, a didhydrodaidzein (DHD) racemase is employed for stereochemical conversion of DHD in some bacteria [28]. Horizontal gene transfer seemed to have occurred in the case of the closely related enzymes (more than 98% amino acid sequence identity [PID]) from *Eggerthella* sp. YY7918 and *L. garviae* [40]. Besides the complete conversion
from daidzein to equol, a large number of individuals seem to carry out only the reduction of daidzein to O-desmethylangolensin [39, 43], which is based on bacteria encoding only daidzein reductase.

Phloretin hydrolase (Phy), a Friedel-Crafts hydrolase, hydrolytically cleaves the C-C bond adjacent to the aromatic A-ring of phloretin and thereby produces phloroglucinol and 3-(4-hydroxyphenyl)propionic acid. The first purified and characterized Phy was that of *Eubacterium ramulus* [26] in the apigenin and naringenin degradation pathway [44]. Only one other Phy sequence from the intestinal pathogenic *Mycobacterium abscessis* showing a 30% amino acid sequence identity to that of *E. ramulus* was published [25].

The NADH-dependent flavanone- and flavanol-cleaving reductase (Fcr), cleaving the heterocyclic C-ring of flavanones and flavanonols, was characterized in detail recently [45]. Suggested to be an enoate reductase earlier and acting in a concerted pathway together with a chalcone isomerase (CHI, [46]), the Fcr of *E. ramulus* was described to catalyze the conversion of e.g. naringenin to phloretin also without CHI. Still, CHIs [46-49] responsible for auronol production from flavanonols [47], could play a substantial role in flavonoid transformation in the human gut. In addition, the specific reaction(s) carried out by CHI might be of biotechnological interest in auronol production. Very recently, the characterization of an ene-reductase (Flr), catalyzing the reduction of the B ring double bond of flavones and flavonols, was published [50] and the sequences of the two characterized enzymes were included in this analysis.

*Table 1: Characterized flavonoid-modifying enzymes of human gut bacteria.*
| Accession no. | Enzyme          | Reaction                              | Source bacterium                                   | Reference     |
|--------------|-----------------|---------------------------------------|--------------------------------------------------|---------------|
| KEF29323.1  | Beta-glucosidase| 7-O-deglycosylation of isoflavones    | Bifidobacterium pseudocatenulatum IPLA36007       | [14]          |
| KEF27912.1  | Beta-glucosidase| 7-O-deglycosylation of isoflavones    | Bifidobacterium pseudocatenulatum IPLA36007       | [14]          |
| KEF28010.1  | Beta-glucosidase| 7-O-deglycosylation of isoflavones    | Bifidobacterium pseudocatenulatum IPLA36007       | [14]          |
| KEF28001.1  | Beta-glucosidase| 7-O-deglycosylation of isoflavones    | Bifidobacterium pseudocatenulatum IPLA36007       | [14]          |
| AFS33105.1  | Beta-glucosidase| 7-O-deglycosylation of isoflavones, 3-/4-/7-O-deglycosylation of queretin glycosides | Bifidobacterium animalis ssp. lactis SH5         | [15]          |
| AAO76887.1  | Beta-glucosidase| 7-O-deglycosylation of isoflavones    | Bacteroides thetaiotaomicron VPI-5482             | [35]          |
| AGS77942.1  | Alpha-L-Rhamnosidase | Derhamnosylation of 1->2 and 1->6 O-glycosidic bond of rutinosylated flavonoids | Bifidobacterium dentis ssp. lactis SH5         | [16]          |
| AHJ22585.1  | Alpha-L-Rhamnosidase | Derhamnosylation of 1->6 O-glycosidic bond of rutin | Bifidobacterium breve 689b                     | [17]          |
| WP_011107561| Alpha-L-Rhamnosidase | Derhamnosylation of 1->2 O-glycosidic bond of rutinosylated flavonoids | Bacteroides thetaiotaomicron VPI-5482             | [18]          |
| QBM20340.1  | Alpha-L-Rhamnosidase HFM-RhaA (MGR1) | Derhamnosylation of 1->2 and 1->6 O-glycosidic bond of rutinosylated flavonoids | Human gut metagenome                          | [21]          |
| QBM20341.1  | Alpha-L-Rhamnosidase HFM-RhaB (MGR2) | Derhamnosylation of 1->2 O-glycosidic bond of rutinosylated flavonoids | Human gut metagenome                          | [21]          |
| QBM20342.1  | Alpha-L-Rhamnosidase HFM-RhaC (MGR3) | Derhamnosylation of 1->2 and 1->6 O-glycosidic bond of rutinosylated flavonoids | Human gut metagenome                          | [21]          |
| CCC80440.1  | Alpha-L-Rhamnosidase LpR1 | Derhamnosylation of 1->6 O-glycosidic bond of rutinosylated flavonoids | Lactobacillus plantarum WCFS1*                  | [19]          |
| CCC80442.1  | Alpha-L-Rhamnosidase LpR2 | Derhamnosylation of 1->6 O-glycosidic bond of rutinosylated flavonoids | Lactobacillus plantarum WCFS1*                  | [19]          |
| AAV43293.1  | Alpha-L-Rhamnosidase LaR | Derhamnosylation of 1->2 and 1->6 O-glycosidic bond of rutinosylated flavonoids | Lactobacillus acidophilus NCFM                 | [19]          |
| EUA80835.1  | Phloretin hydrolase | Phloretin hydrolysis                 | Mycobacteroid abscessus ssp. bolletii 103         | [25]          |
| AAQ12341.1  | Phloretin hydrolase | Phloretin hydrolysis                 | Eubacterium ramulus DSM 16296                    | [26]          |
| AKC35075.1  | DfgCD           | 7-O-deglycosylation of isoflavones and flavones | Catenibacillus scindens DSM 106146               | [22]          |
| AKC35076.1  | DfgCD           | C-deglycosylation of flavones        | Eubacterium cellulosolvens ATCC 43171            | [22]          |
| EIM58373.1  | DfgABCDE        | C-deglycosylation of flavones        |                                  |               |
| EIM58372.1  | DfgABCDE        | C-deglycosylation of flavones        |                                  |               |
| EIM58371.1  | DfgABCDE        | C-deglycosylation of flavones        |                                  |               |
| EIM58370.1  | DfgABCDE        | C-deglycosylation of flavones        |                                  |               |
| EIM58369.1  | DfgABCDE        | C-deglycosylation of flavones        |                                  |               |
| BBG22493.1  | DgpABC          | C-deglycosylation of isoflavones     | Dorea sp. strain PUE                          | [23]          |
| BBG22494.1  | Flavone/flavonol reductase (Fr) | C-ring double bond reduction        | Flavonifractor plautii DSM 6740               | [50]          |
| BBG22495.1  | Flavone/flavonol | C-ring double bond                   | Clostridium ljungdahlil DSM 6740              | [50]          |
| ANU40626.1  | Flavone/flavonol | C-ring double bond                   |                                  |               |
| ADK16070.1  | Flavone/flavonol | C-ring double bond                   |                                  |               |
| Accession   | Description                                      | Function                                    | Organism                          | Reference(s) |
|------------|--------------------------------------------------|---------------------------------------------|-----------------------------------|--------------|
| AGS82961.1 | Flavonone/flavanonol-cleaving reductase (Flr)   | Reduction of flavanones and flavonols      | Eubacterium ramulus DSM [45]      | 16296        |
| WP_154024723.1 | Flavonone/flavanonol-cleaving reductase (Fcr) | Reduction of flavanones and flavonols      | Flavonifractor plautii DSM [45]   | 6740         |
| AIS36173.1 | Chalcone isomerase (CHI)                         | Isomerization of chalcones and flavanones  | Eubacterium ramulus DSM [46, 47, 49] | 16296        |
| EHM54434.1 | Chalcone isomerase (CHI)                         | Isomerization of chalcones                 | Flavonifractor plautii ATCC 29863 (formerly Clostridium orbiscindens) | 46, 47, 49   |
| AFV15450.1 | Tetrahydrodaidzein reductase (TDR)              | Reduction of tetrahydrodaidzein            | Slackia isoflavoniconvertens DSM 22006 | 27           |
| AFV15451.1 | Dihydrodaidzein reductase (DDR)                 | Reduction of dihydrodaidzein               | Slackia isoflavoniconvertens DSM 22006 | 27           |
| AFV15453.1 | Daidzein reductase (DZR)                         | Reduction of daidzein                      | Slackia isoflavoniconvertens DSM 22006 | 27           |
| BAL46928.1 | Tetrahydrodaidzein reductase (TDR)              | Reduction of tetrahydrodaidzein            | Slackia sp. NATTS                 | 29           |
| BAL46929.1 | Dihydrodaidzein reductase                       | Reduction of dihydrodaidzein               | Slackia sp. NATTS                 | 29           |
| BAL46930.1 | Daidzein reductase                              | Reduction of daidzein                      | Slackia sp. NATTS                 | 29           |
| BAM25050.1 | Dihydrodaidzein Racemase                       | Racemization of dihydrodaidzein            | Lactococcus garviae 20-92         | 28           |
| BAJ7244.1  | Tetrahydrodaidzein reductase                    | Reduction of tetrahydrodaidzein            | Lactococcus garviae 20-92         | 51           |
| BAJ7245.1  | Dihydrodaidzein reductase                       | Reduction of dihydrodaidzein               | Lactococcus garviae 20-92         | 51           |
| BAJ22678.1 | Daidzein reductase                              | Reduction of daidzein                      | Lactococcus garviae 20-92         | 41           |
| WP_013979960.1 | Tetrahydrodaidzein reductase          | Reduction of daidzein                      | Eggerthella sp. YY7918            | 40           |
| WP_013979959.1 | Dihydrodaidzein reductase                  | Reduction of dihydrodaidzein               | Eggerthella sp. YY7918            | 40           |
| WP_013979957.1 | Daidzein reductase O-demethylase (ODem)         | Activation of CP                           | Eggerthella sp. YY7918            | 40           |
| ANI69995.1 | O-demethylase (ODem) activating enzyme (AE)    | O-Demethylation of secoisolariciresinol    | Eubacterium limosum ZL-II         | 30           |
| ANI69960.1 | ODem Methyltransferase (MT) 1                  | Methyl transfer to CP                      | Eubacterium limosum ZL-II         | 30           |
| ANI69961.1 | ODem Corrinoid protein (CP)                    | Methyl transfer                            | Eubacterium limosum ZL-II         | 30           |
| ANI69962.1 | ODem MT2                                       | Methyl transfer to CP                      | Eubacterium limosum ZL-II         | 30           |

*An identical rhamnosidase was identified in Lactobacillus plantarum NCC245 (Avila et al., 2009).

All these studies were performed with few flavonoid-modifying bacteria and the corresponding enzymes originate from a limited number of bacterial species. Isolation of these strains from fecal samples could be biased, because of selective enrichment of culturable bacteria. Therefore, the research on gut-bacterial flavonoid modification published up to now does not display a complete picture of flavonoid transformation in the human gut. Phenotypic studies on human fecal samples rely mainly on (partial) 16S rRNA gene sequencing and thus do not resolve the species level and miss flavonoid-converting genera and quantitative studies of genes encoding flavonoid-converting enzymes are scarce. A recent study on the quantification of
daidzein-to-equol-converting bacteria relied on primers specific for the different dihydrodaidzein reductase and tetrahydrodaidzein reductase genes [52]. In this pilot study, only 17 individuals were tested, of which less than half showed daidzein conversion, four assigned to *Adlercreutzia equolifaciens* and three to *Slackia isoavoniconvertens* (or related species). With the advent of metagenomic studies, a closer look into the gene catalog of the human gut microbiome became feasible [53], but still, quantification and especially taxonomic affiliation of the genes of interest were difficult or impossible. Current bioinformatic methods offered the opportunity to assemble genomes from metagenomic data [54], which also was applied recently to human gut metagenome studies [55-57], opening up new ways to reveal flavonoid modification in the human gut.

Here, we aim to identify and quantify flavonoid-modifying bacteria in the human gut microbiota. For this, we screened the most recent combined human gut MAGs collection, the Unified Human Gastrointestinal Protein or Genome (UHGP/UHGG) catalog of more than 280,000 assembled genomes with a BLAST search using protein sequences of characterized flavonoid-modifying enzymes (table 1) as queries. In addition, the presence, prevalence and abundance of bacterial species described to modify flavonoids was investigated. Based on the obtained/collected data, a possible scenario of flavonoid modification by the different identified bacteria in the human gut, based on the experimental evidence from known flavonoid-converting enzymes, is presented.

**Results And Discussion**

**Distribution of flavonoid-modifying bacterial species across human gut MAGs**

First, we screened the MAGs database accompanying metadata for the abundance of bacterial species described to transform flavonoids. For this, a literature-based overview of currently known flavonoid-converting bacterial strains present in the human gut was compiled, resulting in 45 distinct bacterial strains of 43 different species (Table 2). Of these, eleven were not taxonomically validly published, either due to lack of a 16S rRNA gene sequence or because the authors chose an informal taxonomic designation. If a 16S rRNA sequence was available, these unclassified isolates were provisionally assigned to the closest taxonomically described species using a 98% sequence identity cutoff to the respective strain. *O*-deglycosylation was the most abundant flavonoid modification performed by human gut bacteria (31 strains). *C*-deglycosylation, *C*-ring cleavage, reduction, dehydroxylation or *O*-demethylation of flavonoids were far less abundant (between three and seven strains each).

We then performed a text-based filtering and counting of species in the lineage column of the UHGG metadata with the aim to specify the MAG abundances of described flavonoid-transforming species as percentage of the overall number of MAGs in the human gut. Important to note here is that the taxonomic assignments of these metadata are based on the genome taxonomy database (GTDB, [https://gtdb.ecogenomic.org/](https://gtdb.ecogenomic.org/)). Thus, species without a sequenced genome (*Catenibacillus scindens*, *Lactobacillus leichmanii*, *Slackia* sp. NATTS) or species lacking taxonomic assignment in the GTDB to a given genome (*S. isoavoniconvertens*) could not be found. Besides the overall abundance of MAGs, we also calculated the prevalence of MAGs across individuals as percentage of the overall number of metagenome samples in the metadata.
The most prevalent flavonoid-modifying bacteria were *Bacteroides uniformis* (18.5% prevalence across all 21,508 individuals), *Parabacteroides distasonis* (14.5%), *Bifidobacterium infantis* (8.4%), and *Bi. adolescentis* (7.4%; Table 2). All of these species were described to be capable of flavonoid O-deglycosylation, mainly flavonols, isoflavones and flavanones[11]. The remaining species were detected in less than 5% of individuals (Table 2). The overall abundance across all MAGs were similarly distributed (Supplementary Table S1). The numbers for a single *E. coli* strain described to O-deglycosylate flavonoids [58] should be handled with care, since this activity was as yet not confirmed for other *E. coli* strains and the MAGs database does not allow for strain level assignment. The lack of *Bi. longum* in the MAGs metadata is remarkable, but can be explained by the fact that *Bi. longum* ssp. *infantis* [59] is classified as *Bi. infantis* in the GTDB.

In addition, potential flavonoid-modifying bacterial species from a recently published BRENDA enzyme reaction database search [60] were screened for their abundance in the MAGs metadata. However, these bacteria showed only very low abundance (Supplementary Table S2) and were not included in this study.

**Table 2:** Flavonoid-modifying gut bacterial species and their prevalence (percent of MAGs from all samples involved) as MAGs derived from the taxonomic classification in the metadata of the UHGG [61].
| Species | Flavonoid class | Prevalence |
|---------|----------------|------------|
| **O-Deglycosylation** | | |
| *Bacteroides ovatus* | Flavonols | 3.1 |
| *Bacteroides uniformis* | Flavonols, flavanones | 18.5 |
| *Bacteroides thetaiotaomicron*\(^1\) | Isoflavones | 1.9 |
| *Bifidobacterium adolescentis* | Flavanones, isoflavones | 7.4 |
| *Bifidobacterium angulatum* | Isoflavones | 0.3 |
| *Bifidobacterium animalis* | Anthocyanidins, isoflavones | 0.3 |
| *Bifidobacterium bifidum* | Flavanols, isoflavones | 3.3 |
| *Bifidobacterium breve* | Flavonols, isoflavones, flavanones | 1.5 |
| *Bifidobacterium catenulatum* | Flavonols, isoflavones, flavanones | 0.8 |
| *Bifidobacterium dentium* | Flavonols, flavanones | 0.3 |
| *Bifidobacterium infantis* | Flavonols, isoflavones, flavanones | 8.4 |
| *Bifidobacterium longum* | Isoflavones | 0 |
| *Bifidobacterium pseudocatenulatum* | Flavonols, flavanones | 3.5 |
| *Bifidobacterium pseudolongum* | Isoflavones | <0.1 |
| *Blautia producta* (MRG-PMF1, 99%) | Flavonols, flavones, flavanones, isoflavones | <0.1 |
| *Catenibacillus scindens* | Flavonols, isoflavones | n.d. (<0.1) |
| *Enterobacter cloacae* | Flavanones | 0.1 |
| *Enterococcus avium* | Flavonols | 0.1 |
| *Enterococcus casseliflavus* | Flavonols | <0.1 |
| *Enterococcus faecalis* | Flavanones | 2.9 |
| *Eubacterium cellulosolvens* | Flavonols, isoflavones | 0 |
| *Eubacterium ramulus* | Flavonols, flavones, dihydrochalcones, isoflavones | 0.5 |
| *Escherichia* sp. HGH21 (99% to *E. coli* (MIDI\(^*\)) | Isoflavones | (28) |
| *Escherichia* sp. 4 (*E. fergusonii*) | Flavones | <0.1 |
| *Parabacteroides distasonis* | Flavanols, flavanones | 14.5 |
| *Lactobacillus acidophilus* | Flavonones | <0.1 |
| *Lactobacillus buchneri* | Flavonones | <0.1 |
| *Lactobacillus casei* | Flavonones, anthocyanidins | 0 |
| *Lactobacillus leichmanii* | Flavanones | n.d. |
| *Lactobacillus plantarum* | Flavanones, anthocyanidins | 0.2 |
| *Lactococcus lactis* | Flavonols, isoflavones, flavanones | 0.2 |
| *Lactococcus paracasei* [62] | Isoflavones, flavonols | 0.2 |
| **C-Deglycosylation** | | |
| *Eubacterium cellulosolvens* | Flavonols, isoflavones | 0 |
| Strain PUE (1346 nt), *Dorea longicatena* (98%) | Isoflavones | n.d. (2.9) |
| *Catenibacillus scindens* | Flavonols, isoflavones | n.d. |
| *Enterococcus casseliflavus* (sp. 45, 99%) | Flavones | <0.1 |
| *Enterococcus faecium* (MRG-IFC-2, 99%) | Isoflavones | 1.8 |
| *Lactococcus lactis* (MRG-IFC-1, 99%) | Isoflavones | 0.2 |
| **C-Ring cleavage** | | |
| *Flavonifractor plautii* (formerly *C. orbiscindens*) | Flavonols/flavanonols, flavones/flavanones | 2.1 |
| *Catenibacillus scindens* | Flavones | n.d. |
| *Eubacterium ramulus* | Flavonols/flavanonols, flavones/flavanones, isoflavones | 0.5\(^1\) |
| *Clostridium butyricum* | Flavanones | 0.3 |
| *Lactobacillus plantarum* | Flavan-3-ols | 0.2 |
| Strain SY8519 (*Eubacterium* I sp000270305) | Isoflavones | n.d. |
| *Adlercreutzia equolifaciens* | Flavan-3-ols | 0.8 |
Sequence identities are taken from [11] or the reference given in the species column. When the bacterium was classified into the genus, but not into a species, we performed a 16S rRNA gene analysis and list the original species/strain designation and the 16S rRNA sequence identity to the identified strain in brackets behind the species name. *: as per MIDI technique (Hur Lay et al., 2000). ¹: This number is the combined value from two phylogenetic groups (possibly subspecies) given in the GTDB.

**Screening the MAGs amino acid sequence database for potential flavonoid-modifying enzymes**

In the following, the MAGs database of human gut bacteria v1.0 was screened with sequences from characterized flavonoid-modifying enzyme as queries. In total, 1,586,499 hits were found with an e-value of less than 1e-60, 1e-25 or 1e-20, filtered for a coverage of at least 75% and a PID of 30. In particular O-glycosidases and rhamnosidases gave a large number of hits, while O-demethylases and C-glycosidases resulted in rather low number of hits, because we filtered for the abundance of several genes from a larger gene cluster in the single genomes (Supplementary Excel File, pivot table).

**O-Deglycosylation of flavonoids**

The five bifidobacterial flavonoid O-deglycosylating enzyme sequences, which showed a maximum amino acid percent sequence identity (PID) of 35 to each other, resulted in a high number of hits in MAGs classified to six different *Bifidobacterium* sp. (Figure 1) when a PID threshold of 50 and an occurrence of at least 50 for each sequence hit was applied, with the highest numbers observed in MAGs classified as *Bi. adolescentis* (overall 6,633 with a PID > 40), *Bi. infantis* and *Bi. pseudocatelunatum* (overall about 3,000 each, see also Supplementary Excel File). The *Ba. thetaiotaomicron* O-glycosidase query resulted mainly in MAGs from six different *Bacteroides* species, with *Ba. ovatus* and *Ba. thetaiotaomicron* (both about 500 occurrences) as most numerous (Figure 1, Supplementary Excel File). *Bacteroides* members not already characterized as flavonoid O-deglycosylating species include *Ba. cellulosilyticus* and *Ba. eggerthii*. 
Other bacterial genera besides *Bacteroides* or *Bifidobacterium* were detected at a PID of 50 to 60 especially to the three glucosidase sequences of *Bi. pseudocatenulatum* (Figure 1). At PID values of 55 to 60 to BpGluE, MAGs from *Faecalicatena gnarus* and the unclassified *Gemmiger* sp003476825 were highly abundant, with 3,170 and 8,329 overall occurrences, respectively (Supplementary Excel File). No *Gemmiger* bacterium was associated with the deglycosylation of flavonoids up to now, but the high PID to BpGluE and the highest overall abundance of *O*-glycosidase hits make *Gemmiger* sp003476825 (one of the most prevalent bacteria in the human gut [61]), a potentially highly relevant species for flavonoid deglycosylation. Two other taxonomically not described species with a high number of *O*-glycosidase hits with a PID of more than 50 were CAG-180 sp000432435 (Clostridiales, family *Acutalibacteriaceae* about 2,000 occurrences) and CAG-217 sp000436335 of the same family with about 1,000 occurrences. All of these show a PID of about 50 to *Bi. pseudocatenulatum* GluD. The high abundance and similarity to BpGluD indicates a potentially prominent role of these *Acutalibacteriaceae* members in deglycosylation of flavonoids. Furthermore, *Blautia wexlerae* and two taxonomically not described *Eubacterium* spp. with a high PID to BpGluD, were abundant as well. More distantly related glucosidases (PID between 40 and 50 to the queries) include those of *Faecalicatena gnarus*, several *Clostridium* and *Ruminococcus* species and *Gemmiger* sequences more closely related to the *Bifidobacterium breve* enzyme and a second *Blautia* enzyme (Supplementary Figure 1). Since only isoavones were tested as substrates for the bifidobacterial enzymes, conclusions about the substrate specificity of the here detected homologs cannot be given. Usually also flavonones and flavonols are deglycosylated by *Bifidobacterium* spp. [11].

To unravel the phylogenetic relationships between the observed enzymes and to see whether the novel sequences can be classified into distinct phylogenetic classes with potential new catalytic properties, a phylogenetic tree was created. The largest part of the *Bl. wexlerae* sequences form a large distinct clade including sequences from *Faecalicatena gnarus* as well, while the *Gemmiger* glucosidases forms another clade, also distinct from all clades harboring characterized flavonoid *O*-glycosidases (Supplementary Figure 2). The latter forms one phylogenetic clade together with glucosidases of four *Acutalibacteriaceae* species. Only four of the sequenced enzymes (BpGluA, BpGluB, BpGluD) and the *Bacteroides* sequence form larger clades, with the *Bl. wexlerae*, the CAG-180 and the *F. prausnitzii* clades being much larger, suggesting that the *O*-deglycosylating potential of the human gut microbiota is largely uncharacterized and awaits biochemical investigation.

A second class of *O*-glycosidases was observed in *Catenibacillus scindens*, where a cluster of two genes, *dfgCD*, was characterized to encode an enzyme that 7-*O*-deglycosylated isoavones and flavones [22]. DfgCD homologs are found scarcely in human gut MAGs, mainly in *Hungatella hathewayi*, *Faecalicatena* species, and *Eisenbergiella* sp900066775, all with approximately 100 occurrences (Supplementary Figure 3).

**Derhamnosylation of flavonoids**

The BLAST search with queries from six rhamnosidases of *Lactobacillus*, *Bifidobacterium* and *Bacteroides* species resulted mainly in hits with MAGs classified as *Bifidobacterium* and *Bacteroides*. The low number of hits to *Lactobacillus* MAGs (only *L. plantarum* was detected with a PID > 65 and Freq > 20) was expected, given the low abundance of *Lactobacillus* species in the gut MAGs metadata (Table 2). One of the most abundant species was *Bacteroides dorei*, not yet described as rhamnoglycoside-hydrolyzing species (Figure
2), whereas the species described to de-rhamnosylate e.g. rutin, such as Bi. pseudocatenulatum, Bi. dentium, Bi. breve, and Ba. thetaiotaomicron were of comparably lower numbers. The highest number of MAGs in this analysis was observed when using the metagenome-derived rhamnosidase 3 sequence as query. These MAGs could be assigned to Gemmiger sp003476825, which also contains a putative flavonoid O-glucosidase (see above, no sequence similarities to the rhamnosidase described here). With a lower abundance (about 50 appearances), metagenome-derived rhamnosidase 2 orthologs were detected in MAGs of the genus CAG-170 (family Oscillospiraceae). Orthologs (PID of about 98) of the metagenome-derived rhamnosidase 1 were detected in only five MAGs from the family Monogloboaceae, which could not be assigned to a genus yet and differ from the only described member of this genus, the pectin-degrading intestinal Monoglobus pectiniylicus [65, 66], which does not encode the corresponding rhamnosidase. Below a 65 PID cutoff, mainly species from the genera Parabacteroides and Bacteroides and the species Alistipes onderdonkii (approximately 45 to 50 PID to MGR2) were found to be highly abundant (Supplementary Figure 4). Several distinct rhamnosidase families can be distinguished phylogenetically, of which the largest cluster does not include characterized ones and contains mainly sequence hits from Bacteroides and Parabacteroides genomes. Interestingly, most characterized rhamnosidases do not belong to phylogenetic nodes mainly representing the corresponding species: B. dentium is assigned to a cluster containing mostly B. pseudocatenulatum genomes, L. acidophilus belongs to a large node containing mostly Fusicatenibacter saccharivorans, the node including the characterized Bidobacterium breve rhamnosidase is dominated by Parabacteroides merdae, while the LpR2 produced mainly hits from Bacteroides dorei and Parabacteroides species (Supplementary Figure 4). The B. thetaiotaomicron rhamnosidase and LpR1 cluster together in a large node dominated by Bacteroides dorei. These results show that the characterized isolate species often do not represent the species mainly responsible for the de-rhamnosylating activity in the human gut.

C-Deglycosylation of flavonoids

The current gene clusters reported to play a role in C-deglycosylation are depicted in Figure 3. Of the Eubacterium cellulosolvens dfgABCDE gene cluster (Figure 3A, [22]), all sequences were considered as queries in this study, because it is not completely resolved which gene products are involved in catalysis. Of the dgp gene cluster of strain PUE (Figure 3 B), we used only the first three genes (dgpABC) as queries, since only these genes were shown to catalyze reactions in the C-deglycosylation pathway [23, 24].

To respond to the involvement of more than one enzyme in this reaction, we filtered for genomes encoding all three catalytic subunits dgpABC or at least three of dfgABCDE. The latter resulted in only a few similar hits with a PID higher than 40 (less than 100, Supplementary Figure 4), most of them in the Faecalicatena genus (Supplementary Figure 4). Since dfgD was not found in the majority of MAGs containing dfgABCE, it is likely that DfgD is not involved in C-deglycosylation. The effect of a lack of dfgD in the dfgABCDE gene cluster in C-deglycosylation was not specifically investigated [22]. A higher number of hits to DfgA (more than 300) with a PID lower than 40 was observed in Faecalicatena species, but these were single genes not part of a dfgABCDE cluster (Supplementary Figure 5). The dgpABC cluster was much more prominent in the human gut MAGs than the dfgABCE cluster, with most hits in Agathobacter faecis (formerly Roseburia faecis, PID of 50 to 85%) plus several dozens in Blautia, Dorea, Enterococcus and Faecalicatena MAGs (Figure 4). However, only a low
percentage of *A. faecis* MAGs carried the *dgpABC* cluster, i.e. 804 of the 2819 MAGs assigned to that species (~ 30%), so that the occurrence of *A. faecis* cannot be linked necessarily to the C-deglycosylation of flavonoids. Even less specific seems the occurrence of C-deglycosylating genes in *Faecallicatena gnavus*, with only 84 MAGs of 1197 encoding DgpABC (Supplementary Table 1).

**Daidzein-to-equol conversion**

The three enzymes responsible for daidzein-to-equol conversion (DZR, DDR and TDR) plus the optional dihydrodaidzein racemase encoded by *Lactococcus garviae* [28] are encoded in a single cluster, which is very similar in *L. garviae*, *A. equolifaciens* and *S. isoavoniconvertens*. Due to the high sequence similarities of the enzymes from the three species, we used only the *S. isoavoniconvertens* sequences (plus the *L. garviae* racemase sequence) as queries. Up- and downstream genes were not considered in this study, since they were suggested to play accessory roles [27]. Hits in *Slackia*, *Lactococcus* or *Eggerthella* were scarce, which is contradicting the literature reports on a widely distributed daidzein-to-equol conversion by these species in the human gut. Four hits of *Slackia* MAGs containing the daidzein-to-equol gene cluster were assigned to the same species represented by GUT_GENOME145587. A comparison of this representative genome to the available *S. isoavoniconvertens* DSM22006 genome (GenBank assembly number GCA_003725955.1) using ANI calculator [67] resulted in an ANI value of 95%. Therefore, the classification of these MAGs to *S. isoavoniconvertens* is justified. Since the other 261 MAGs representing this species do not encode the daidzein-to-equol pathway enzymes, *S. isoavoniconvertens* appears not to be a typical equol producer in the human gut. Therefore, conclusions of equol production in an individual should not be based on 16S rRNA analyses, which obviously cannot resolve the daidzein-to-equol transforming potential of *Slackia* spp. Another bacterial species described to transform daidzein to equol, *Adlercreutzia equolifaciens*, was more abundant, with 33 occurrences of the *ddr* gene, but also here, the percentage of MAGs carrying the daidzein-to-equol gene cluster was relatively low with less than ten percent (a total of 271 *A. equolifaciens* MAGs). This observation is strengthened by a report of a human gut *A. equolifaciens* isolate, lacking the cluster and not able to transform daidzein to equol (Valezquez et al. 2020). Another species whose MAGs frequently carried daidzein-to-equol cluster genes was CAG-1427 sp000435745 (*Eggerthellaceae*, Figure 5), of which nearly 50% (a number of 35) of the MAGs (Supplementary Table 1) carry the daidzein-to-equol gene cluster. Therefore, the occurrence of these bacterial species cannot give a reliable conclusion on the main equol producers in the human gut. Specifically designed primers in a PCR-based method revealed that three of 17 arbitrarily selected individuals (of which 9 were “equol-producers”) carried equol-producing *Slackia*, another four *Adlercreutzia* [52], which together with the present study hints towards an at least equal or higher importance of *A. equolifaciens* compared to *S. equolifaciens* in equol production. However, the higher percentage of equol-producing individuals is in contrast to the low number of MAGs from potential equol-producing bacteria in the present study. This might be explained by a low abundance of equol-producers in most individuals, preventing an efficient assembly of MAGs from metagenomes, which of course is much less sensitive compared to sensitive HPLC-based equol detection or the highly specific PCR-based detection of the corresponding genes. Since two of the equol-producing individuals of the mentioned study did not result in a PCR product with the employed primer pairs, undiscovered equol-forming bacteria exist in the human gut. A high number of hits with a PID of 30 to 40 was identified in in *Agathobacter rectalis* (previously *Eubacterium rectale*) and several, mostly uncharacterized *Collinsella* species (Figure 5). The occurrence of this latter genus in human gut
microbiomes correlated with the ability to transform daidzein to equol in several studies [68-70]. However, since the PID to the characterized daidzein-transforming enzymes is rather low with only about 30 to 35 and the corresponding genes are not clustered, but spread out in the genome, this assumption warrants further research. A relatively high number of genomes showing only hits to the daidzein reductase sequence might be responsible for the large number of phenotypes transforming daidzein only to O-Desmethylangolensin. Especially bacteria assigned to Erysipelatoclostridium and Clostridium M genera carry putative dzr genes (Supplementary Figure 6, Supplementary Excel File).

**Flavone/flavonol reduction, flavanone/flavanonol ring cleavage, chalcone isomerization and phloretin hydrolysis**

Since these four reactions, the reduction of flavons and flavonols to flavanones and flavanonols, respectively (catalyzed by Flr), the C-ring cleavage of these products (via Fcr), isomerization of chalcones and auronols (via CHI), and the hydrolysis of the dihydrochalcone phloretin (via Phy) can be viewed as steps in the same pathway of flavone/flavonol degradation (e.g. of apigenin via naringenin, Figure 6), the results are compiled in one subchapter.

The very recently characterized Flrs from Flavonifractor plautii and Clostridium ljungdahlii [50] resulted in a large number of hits in *F. plautii* and *Clostridoides difficile* MAGs (the latter interestingly with a higher PID to the sequence of *F. plautii* than to the sequence of *C. ljungdahlii*), and, with a lower PID, in MAGs of *Olsenella, Eubacterium* and *Fusobacterium* species (Figure 7). Most of the *F. plautii* MAGs (also the type strain DSM 6740) harbor two copies of the *flr* gene, with one of them closer related (PID of ~ 60) to that of *Clostridium ljungdahlii* than to the “main” Flr of *F. plautii*. This second putative Flr of *F. plautii* didn’t show any apigenin-reducing activity [50], but might reduce other flavonoid classes, which were not tested in the corresponding study.

The BLAST analysis with the characterized flavanone- and flavanonol-cleaving reductase (Fcr) from *Eubacterium ramulus* resulted in hits identified in *F. plautii* MAGs (523 overall occurrences, Supplementary Excel File), while *E. ramulus* MAGs revealed a lower number of 188 occurrences. Similar to the situation with Flr, several *E. ramulus* genomes contained two *fcr* genes, one of which with a higher PID to the *F. plautii* gene, which might reflect a different substrate spectrum of the encoded Fcrs. Besides *F. plautii* and *E. ramulus*, MAGs assigned to Anaerostipes hadrus frequently (138 MAGs) encoded an Fcr-like enzyme with approximately 85% identity to Fcr from *E. ramulus*. Besides these three main species, a number of MAGs from *Clostridium boltae* encoded a putative Fcr with a PID of approximately 42 to the Fcr from *F. plautii*. Most hits to Fcr (several hundred up to nearly 1000) showed around 35 PID (Figure 8) and frequently occurred in particular in Faecalicatena spp. Whether these putative enzymes have a similar function and whether *Faecalicatena* and other species contribute to flavonoid degradation warrants further investigation.

Genes similar to that encoding CHI (Supplementary Figure 8) were highly abundant in MAGS assigned to Flavonifractor plautii (593 overall occurrences) and *Clostridoides difficile* (375) and, to a lower extent, present in *Eubacterium ramulus* (137), in which it was initially described. Below a PID of 40, only MAGs from *Fusobacterium mortiferum* were detected in a higher number (ca. 100).
The BLAST search using the phloretin hydrolase (Phy) sequence of *E. ramulus* resulted in several hits with a PID of more than 90, assigned to 159 *E. ramulus* MAGs (Figure 9). The most abundant Phy homologs with a PID of 46 (579 occurrences, Supplementary Excel file) were found in MAGs assigned to *Flavonifractor plautii*, which was described to hydrolyze phloretin [71]. Genes of two species never reported to transform flavonoids, *Dialister succinatophilus* (322 occurrences overall) and *Anaerostipes hadrus* (119 occurrences) were abundant with a PID of approximately 60 and 75 to the Phy query. Opposed to the Phy of *E. ramulus*, the BLAST search with the Phy of *M. abscessis* resulted in only few hits below 40 PID in a *Dakarella* species.

The co-occurrence of hits to all enzyme sequences (Flr, Fcr, CHI and Phy) was observed in most of the *F. plautii* and *E. ramulus* MAGs, emphasizing the co-functionality of these enzymes in a single pathway of flavone degradation. *F. plautii* was by far the most abundant species in this analysis, pointing toward its major role in flavone degradation in the human gut. Of the 610 MAGs assigned to *F. plautii*, most (509 to 610) contained at least one of the genes. In contrast, *A. hadrus*, containing only Fcr and Phy, showed only a minor amount of MAGs carrying avonoid-degrading enzymes (less than 5 % of the overall MAGs).

### O-Demethylation of flavonoids

Similar to C-deglycosylation, O-demethylation requires several proteins (MT1 and MT2, CP and AE) acting together in a complex reaction pathway. Since no specific flavonoid O-demethylation system has been characterized, the four enzyme sequences of the O-demethylase system from *Eubacterium limosum*, which acts on the polyphenolic lignan secoisolariciresinol [30], were used as queries. Above 70 PID, only a few dozen hits from the *Eubacterium* genus, notably from the species *E. callenderi* and to a lesser extent *E. limosum* (characterized for flavonoid O-demethylation), were observed (Supplementary Figure 9). The highest number of similar O-demethylase sequences was detected in MAGs of *Intestinibacter bartletti* (PIDs of 60 for MT1, ca. 40 to 50 for MT2 and CP; ca. 33 for AE). The high PID of MT1, which defines the substrate specificity of O-demethylation [32], suggests that *I. bartletti* (never tested for O-demethylation of flavonoids) has a similar substrate spectrum for O-demethylation as *E. limosum*. The MT1 gene is part of a complete O-demethylase gene cluster in *I. bartletti* (Supplementary Figure 10) including two CP genes and *cobW*, encoding a cobalt-binding protein. The high number of MAGs with hits to the CP of the O-demethylase operon sequences (Supplementary Excel File) can be explained by the high abundance of proteins homologous to CP due to the common B$_{12}$-binding and Fe-S cluster domains. In fact, only a low number of MAGs were found to contain the complete operon with a PID of above 40, with *I. bartletti* MAGs the most numerous one (69 occurrences, Supplementary Excel File).

### Conclusion

The here presented *in silico* analysis of potential flavonoid-modifying enzymes of human gut bacteria is a first step to uncover the complete picture of flavonoid transformation in the human intestine. The findings constitute a basis for targeted testing of the detected potential novel enzymes and their bacterial sources. Figure 10 summarizes some representative flavonoid-transformations, the prevalence of MAGs encoding the corresponding flavonoid-transforming enzymes and the main bacterial species involved in these reactions. In addition to the already characterized gut bacteria involved in flavonoid transformation, potential species and genera so far not associated with the modification of these polyphenols in the human gut were identified. Of
these novel flavonoid-modifying candidates, several species are as yet not isolated. These include for example *Gemmiger* sp003476825 potentially involved in *O*-deglycosylation and derhamnosylation or the CAG-1427 genus in *O*-deglycosylation. Heterologous expression of the potentially involved enzymes could reveal their flavonoid deglycosylation potential. Some other bacteria were already known to play a crucial metabolic role in the human gut (e.g. production of short-chain fatty acids or polysaccharide degradation), but their ability to convert flavonoids may have been overlooked. An example is *Anaerostipes hadrus*, so far mainly known for butyrate production [72]. An (albeit minor) portion of MAGs of this highly abundant human gut bacterium was observed to encode several putative flavonoid-transforming enzymes including a Phy and an Fcr. A main flavonoid-degrading bacterium seems to be *Flavonifractor plautii*: Most of its MAGs contain Flr, Fcr, Phy and CHI genes, and it is relatively abundant. Thus, *F. plautii* might serve as a good indicator of the flavonoid degradation potential in 16S rRNA gene-based gut microbiota analyses.

On the other hand, our quantitative results on flavonoid-modifying bacteria implies that several key flavonoid modification processes in the human gut might have been so far assigned to bacteria, which are neither representative nor the main flavonoid-transforming bacteria in this ecosystem. For example, equol formation from daidzein appears to be catalyzed in the human intestine mainly by *Eggerthella* and maybe by *Collinsella* species probably more often than by the thoroughly studied *Slackia isoflavoniconvertens* and *Adlercreutzia equolificiens*.

The observation of the wide array of uncharacterized, putative flavonoid-transforming enzymes in human gut bacteria presented here might be a valuable source for identification of novel enzymes of biotechnological interest, especially for the production of unusual or difficult-to-produce flavonoids. Several bacterial species seem to be specialized in metabolizing flavonoids and possibly other polyphenols as well. For example, nearly all *Flavonifractor plautii* MAGs harbor the three enzymes Phy, Fcr and CHI, and therefore seem to be the main flavanone and flavanonol-degrading bacteria in the human gut. While further research on the effect of this degradation on human health is necessary, it might be one of the reasons, why the presence of *F. plautii* was associated with the development of colon cancer in an Indian population [73]. Finally, the outcome of the here presented *in silico* analysis could contribute to clarify the impact of gut bacteria on flavonoid-mediated health effects in humans.

**Methods**

**Retrieval of query sequences, MAGs data and set-up of UHGP BLAST database**

Besides extracting the most recent review on flavonoid modification [11], a literature search on the databases from Web of Science and Pubmed and in the Google scholar search engine was performed. The employed keywords were “flavonoid(s)” AND “modification” OR “deglycosylation” OR “demethylation” OR “cleavage”. Furthermore, screening in the citation network of the retrieved publications was performed and manually reviewed before Genbank accession numbers and the corresponding amino acid sequences were retrieved from the NCBI Genbank database (Table 1).

A BLAST-searchable database including all non-redundant amino acid sequences from the UHGP catalog was set up as follows: The UHGP-100 version 1.0 non-redundant protein fasta file (uhgp-100.faa) from 286,997
MAGs assembled from gut metagenomes [61] was downloaded as the packed uhgp-100.tar.gz file from the MGnify ftp server ((http://ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify_genomes/human-gut/v1.0/) on July 1st, 2020 and subsequently extracted. With BLAST+ (stand-alone, version 2.9.0 [74]) installed on Linux Ubuntu 18.04 and the makeblastdb command, a BLAST database was created from the uhgp-100 file. For quantification purposes, the uhgp-100.tsv table including all of the redundant amino acid sequence IDs (all assigned to a non-redundant “representative” sequence ID corresponding to the identical amino acid sequence in other MAGs) was extracted from the same packed file (uhgp-100.tar.gz). The UHGP metadata which included MAGs taxonomy assigned via the genome taxonomy database (GTDB, release 89), genome quality values, study and sample information and others were downloaded from the same mgnify ftp-address (genomes-all_metadata.tsv file). Please note that there might be differences in the taxonomy, since the current release of the GTDB is no. 202 and the metadata file was not updated.

BLAST searches, data mangling and quantification

The queries were searched against the UHGP-100 blast database using the blastp search function and an e-value cutoff of 1e-25 for pathways including short sequences (150 to 350 amino acids; daidzein-to-equol, CHI, O-demethylation, Phy, Flr, DfgCD) or 1e -60 for longer sequences (>350 amino acids; Fcr, O-glycosidases, rhamnosidases) and 1e-20 for the C-deglycosylation enzymes, which contained two sequences below 150 amino acids. The maximal number of target hits was set to 100,000 via the max_target_seqs argument. Blast table output format 6 was chosen with the specified columns (the Linux BASH commands are given in supplementary txt-file). The output files were concatenated and filtered for a PID of at least 30 and a query coverage value of at least 75 %. To avoid redundant hits to more than one query, only the hits with the best bitscore were kept. To retrieve the redundant amino acid sequence IDs for quantification of the hits, grep with a list of the non-redundant sequence IDs as input was used on the uhgp-100.tsv file. The table was merged with the metadata and subsequently filtered for MAGs only (isolate genomes also included in the uhgp catalog were discarded) and the occurrences of the redundant amino acid sequences were counted, these values accounting for the frequency in a given non-redundant sequence group were placed into a new column (Freq). For further filtering and data visualization, R (v 3.6.3) was used in RStudio (v 1.3) with ggplot2.

MAGs gene cluster analysis

To compare and analyze gene clusters among MAGs or isolate genomes, the corresponding genomes were downloaded from the NCBI database or the UHGP MAGs folder and uploaded to the rapid annotation subsystem technology (RAST) server (if possible with their original annotation preserved). The sequence comparison tool was used to compare up to ten genomes.

Construction of phylogenetic trees

Phylogenetic trees of the rhamnosidase and O-glycosidase sequences were constructed after clustering of the sequences with CD-hit (v 4.8.1 [75]) using standard settings (90% PID clustering threshold) and alignment with ClustalOmega (v 1.2.4, [76]) installed on Linux Ubuntu. The alignment was refined with Clipkit (v.1.1.2, [77]) before the tree was generated with Fasttree (v 2.1.10, standard settings [78]). Tree visualization was performed via iTol (v 6, [79]) and manually refined in Inkscape for Windows (v1.01). The size of the nodes
triangles was drawn to scale using the number of CD-hit clustered sequences as well as the redundant sequence IDs of the uhgp-100.tsv file.

**Declarations**

**Acknowledgement**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 814650 (SynBio4Flav).

**Conflicting interests**

The authors have no conflicting interests to declare.

**References**

1. Vargas F, Romecin P, Garcia-Guillen AI, Wangesteen R, Vargas-Tendero P, Paredes MD, Atucha NM, Garcia-Estan J: Flavonoids in Kidney Health and Disease. Front Physiol 2018, 9:394.

2. Panche AN, Diwan AD, Chandra SR: Flavonoids: an overview. J Nutr Sci 2016, 5:e47.

3. Pei R, Liu X, Bolling B: Flavonoids and gut health. Curr Opin Biotechnol 2020, 61:153-159.

4. Farzaei MH, Singh AK, Kumar R, Crole CR, Pandey AK, Coy-Barrera E, Kumar Patra J, Das G, Kerry RG, Annunziata G, et al: Targeting Inflammation by Flavonoids: Novel Therapeutic Strategy for Metabolic Disorders. Int J Mol Sci 2019, 20:4957.

5. Caro-Ordieres T, Marin-Royo G, Opazo-Rios L, Jimenez-Castilla L, Moreno JA, Gomez-Guerrero C, Egido J: The Coming Age of Flavonoids in the Treatment of Diabetic Complications. J Clin Med 2020, 9:346.

6. Bondonno NP, Dalgaard F, Kyro C, Murray K, Bondonno CP, Lewis JR, Croft KD, Gislason G, Scalbert A, Cassidy A, et al: Flavonoid intake is associated with lower mortality in the Danish Diet Cancer and Health Cohort. Nat Commun 2019, 10:3651.

7. Bondonno NP, Lewis JR, Blekkenhorst LC, Bondonno CP, Shin JH, Croft KD, Woodman RJ, Wong G, Lim WH, Gopinath B, et al: Association of flavonoids and flavonoid-rich foods with all-cause mortality: The Blue Mountains Eye Study. Clin Nutr 2019.

8. Cassidy A, Minihane AM: The role of metabolism (and the microbiome) in defining the clinical efficacy of dietary flavonoids. Am J Clin Nutr 2017, 105:10-22.

9. Chen Z, Zheng S, Li L, Jiang H: Metabolism of flavonoids in human: a comprehensive review. Curr Drug Metab 2014, 15:48-61.

10. Rodriguez-Mateos A, Vauzour D, Krueger CG, Shanmuganayagam D, Reed J, Calani L, Mena P, Del Rio D, Crozier A: Bioavailability, bioactivity and impact on health of dietary flavonoids and related compounds: an
11. Braune A, Blaut M: Bacterial species involved in the conversion of dietary flavonoids in the human gut. 
   *Gut Microbes* 2016, 7:216-234.

12. Duda-Chodak A, Tarko T, Satora P, Sroka P: Interaction of dietary compounds, especially polyphenols, 
   with the intestinal microbiota: a review. *Eur J Nutr* 2015, 54:325-341.

13. Rafii F: The role of colonic bacteria in the metabolism of the natural isoflavone daidzin to equol. 
   *Metabolites* 2015, 5:56-73.

14. Guadamuro L, Florez AB, Alegria A, Vazquez L, Mayo B: Characterization of four beta-glucosidases 
   acting on isoflavone-glycosides from *Bifidobacterium pseudocatenulatum IPLA 36007*. *Food Res Int* 2017, 
   100:522-528.

15. Youn SY, Park MS, Ji GE: Identification of the beta-glucosidase gene from *Bifidobacterium animalis* 
   subsp. lactis and its expression in *B. bifidum* BGN4. *J Microbiol Biotechnol* 2012, 22:1714-1723.

16. Bang SH, Hyun YJ, Shim J, Hong SW, Kim DH: Metabolism of rutin and poncirin by human intestinal 
   microbiota and cloning of their metabolizing alpha-L-rhamnosidase from *Bifidobacterium dentium*. 
   *J Microbiol Biotechnol* 2015, 25:18-25.

17. Zhang R, Zhang BL, Xie T, Li GC, Tuo Y, Xiang YT: Biotransformation of rutin to isoquercitrin using 
   recombinant alpha-L-rhamnosidase from *Bifidobacterium breve*. *Biotechnol Lett* 2015, 37:1257-1264.

18. Wu T, Pei J, Ge L, Wang Z, Ding G, Xiao W, Zhao L: Characterization of a alpha-L-rhamnosidase from 
   *Bacteroides thetaiotaomicron* with high catalytic efficiency of epimedin C. *Bioorg Chem* 2018, 81:461-467.

19. Beekwilder J, Marcozzi D, Vecchi S, de Vos R, Janssen P, Francke C, van Hylckama Vlieg J, Hall RD: 
   Characterization of Rhamnosidases from Lactobacillus plantarum and Lactobacillus acidophilus. 
   *Appl Environ Microbiol* 2009, 75:3447-3454.

20. Avila M, Jaquet M, Moine D, Requena T, Pelaez C, Arigoni F, Jankovic I: Physiological and biochemical 
   characterization of the two alpha-L-rhamnosidases of Lactobacillus plantarum NCC245. *Microbiology* 2009, 
   155:2739-2749.

21. Li BC, Zhang T, Li YQ, Ding GB: Target Discovery of Novel alpha-L-Rhamnosidases from Human Fecal 
   Metagenome and Application for Biotransformation of Natural Flavonoid Glycosides. *Appl Biochem 
   Biotechnol* 2019, 189:1245-1261.

22. Braune A, Engst W, Blaut M: Identification and functional expression of genes encoding flavonoid O- 
   and C-glycosidases in intestinal bacteria. *Environ Microbiol* 2016, 18:2117-2129.

23. Nakamura K, Zhu S, Komatsu K, Hattori M, Iwashima M: Expression and Characterization of the 
   Human Intestinal Bacterial Enzyme Which Cleaves the C-Glycosidic Bond in 3"-Oxo-puerarin. *Biol Pharm 
   Bull* 2019, 42:417-423.
24. Nakamura K, Zhu S, Komatsu K, Hattori M, Iwashima M: **Deglycosylation of the Isoflavone C-Glucoside Puerarin by a Combination of Two Recombinant Bacterial Enzymes and 3-Oxo-Glucose.** *Appl Environ Microbiol* 2020, **86**:e00607-00620.

25. Han JT, Zhang SP, Jia WJ, Zhang Z, Wang Y, He YX: **Discovery and structural analysis of a phloretin hydrolase from the opportunistic human pathogen Mycobacterium abscessus.** *FEBS J* 2019, **286**:1959-1971.

26. Schoefer L, Braune A, Blaut M: **Cloning and expression of a phloretin hydrolase gene from Eubacterium ramulus and characterization of the recombinant enzyme.** *Appl Environ Microbiol* 2004, **70**:6131-6137.

27. Schröder C, Matthies A, Engst W, Blaut M, Braune A: **Identification and expression of genes involved in the conversion of daidzein and genistein by the equol-forming bacterium Slackia isoflavonicconvertens.** *Appl Environ Microbiol* 2013, **79**:3494-3502.

28. Shimada Y, Takahashi M, Miyazawa N, Abiru Y, Uchiyama S, Hishigaki H: **Identification of a novel dihydrodaidzein racemase essential for biosynthesis of equol from daidzein in Lactococcus sp. strain 20-92.** *Appl Environ Microbiol* 2012, **78**:4902-4907.

29. Tsuji H, Moriyma K, Nomoto K, Akaza H: **Identification of an enzyme system for daidzein-to-equol conversion in Slackia sp. strain NATTS.** *Appl Environ Microbiol* 2012, **78**:1228-1236.

30. Chiu HH, Hsieh YC, Chen YH, Wang HY, Lu CY, Chen CJ, Li YK: **Three important amino acids control the regioselectivity of flavonoid glucosidation in glycosyltransferase-1 from Bacillus cereus.** *Appl Microbiol Biotechnol* 2016, **100**:8411-8424.

31. Engelmann T, Kaufmann F, Diekert G: **Isolation and characterization of a veratrol:corrinoid protein methyl transferase from Acetobacterium dehalogenans.** *Arch Microbiol* 2001, **175**:376-383.

32. Studenik S, Vogel M, Diekert G: **Characterization of an O-demethylase of Desulfitobacterium hafniense DCB-2.** *J Bacteriol* 2012, **194**:3317-3326.

33. Paraiso IL, Plagmann LS, Yang L, Zielke R, Gombart AF, Maier CS, Sikora AE, Blakemore PR, Stevens JF: **Reductive Metabolism of Xanthohumol and 8-Prenylnaringenin by the Intestinal Bacterium Eubacterium ramulus.** *Mol Nutr Food Res* 2019, **63**:e1800923.

34. Possemiers S, Rabot S, Espin JC, Bruneau A, Philippe C, Gonzalez-Sarrias A, Heyerick A, Tomas-Barberan FA, De Keukeleire D, Verstraete W: **Eubacterium limosum activates isoxanthohumol from hops (Humulus lupulus L.) into the potent phytoestrogen 8-prenylnaringenin in vitro and in rat intestine.** *J Nutr* 2008, **138**:1310-1316.

35. Byun DH, Choi HJ, Lee HW, Jeon HY, Choung WJ, Shim JH: **Properties and applications of -glycosidase from Bacteroides thetaiotaomicron that specifically hydrolyses isoflavone glycosides.** *International Journal of Food Science and Technology* 2015, **50**:1405-1412.
36. Yadav V, Yadav PK, Yadav S, Yadav KDS: alpha-L-Rhamnosidase: A review. Process Biochemistry 2010, 45:1226-1235.

37. Iino C, Shimoyama T, Iino K, Yokoyama Y, Chinda D, Sakuraba H, Fukuda S, Nakaji S: Daidzein Intake Is Associated with Equol Producing Status through an Increase in the Intestinal Bacteria Responsible for Equol Production. Nutrients 2019, 11:433.

38. Song KB, Atkinson C, Frankenfeld CL, Jokela T, Wahala K, Thomas WK, Lampe JW: Prevalence of daidzein-metabolizing phenotypes differs between Caucasian and Korean American women and girls. J Nutr 2006, 136:1347-1351.

39. Frankenfeld CL, Atkinson C, Thomas WK, Gonzalez A, Jokela T, Wahala K, Schwartz SM, Li SS, Lampe JW: High concordance of daidzein-metabolizing phenotypes in individuals measured 1 to 3 years apart. Br J Nutr 2005, 94:873-876.

40. Kawada Y, Yokoyama S, Yanase E, Niwa T, Suzuki T: The production of S-equol from daidzein is associated with a cluster of three genes in Eggerthella sp. YY7918. Biosci Microbiota Food Health 2016, 35:113-121.

41. Shimada Y, Yasuda S, Takahashi M, Hayashi T, Miyazawa N, Sato I, Abiru Y, Uchiyama S, Hishigaki H: Cloning and expression of a novel NADP(H)-dependent daidzein reductase, an enzyme involved in the metabolism of daidzein, from equol-producing Lactococcus strain 20-92. Appl Environ Microbiol 2010, 76:5892-5901.

42. Mayo B, Vazquez L, Florez AB: Equol: A Bacterial Metabolite from The Daidzein Isoflavone and Its Presumed Beneficial Health Effects. Nutrients 2019, 11:2231.

43. Frankenfeld CL: O-desmethylangolensin: the importance of equol's lesser known cousin to human health. Adv Nutr 2011, 2:317-324.

44. Schneider H, Blaut M: Anaerobic degradation of flavonoids by Eubacterium ramulus. Archives of Microbiology 2000, 173:71-75.

45. Braune A, Gutschow M, Blaut M: An NADH-Dependent Reductase from Eubacterium ramulus Catalyzes the Stereospecific Heteroring Cleavage of Flavanones and Flavanonols. Appl Environ Microbiol 2019, 85:e01233-01219.

46. Gall M, Thomsen M, Peters C, Pavlidis IV, Jonczyk P, Grunert PP, Beutel S, Scheper T, Gross E, Backes M, et al: Enzymatic conversion of flavonoids using bacterial chalcone isomerase and enoate reductase. Angew Chem Int Ed Engl 2014, 53:1439-1442.

47. Braune A, Engst W, Elsinghorst PW, Furtmann N, Bajorath J, Gutschow M, Blaut M: Chalcone Isomerase from Eubacterium ramulus Catalyzes the Ring Contraction of Flavanonols. J Bacteriol 2016, 198:2965-2974.
48. Herles C, Braune A, Blaut M: First bacterial chalcone isomerase isolated from Eubacterium ramulus. *Arch Microbiol* 2004, **181**:428-434.

49. Thomsen M, Tuukkanen A, Dickerhoff J, Palm GJ, Kratzat H, Svergun DI, Weisz K, Bornscheuer UT, Hinrichs W: Structure and catalytic mechanism of the evolutionarily unique bacterial chalcone isomerase. *Acta Crystallogr D Biol Crystallogr* 2015, **71**:907-917.

50. Yang G, Hong S, Yang P, Sun Y, Wang Y, Zhang P, Jiang W, Gu Y: Discovery of an ene-reductase for initiating flavone and flavonol catabolism in gut bacteria. *Nat Commun* 2021, **12**:790.

51. Shimada Y, Takahashi M, Miyazawa N, Ohtani T, Abiru Y, Uchiyama S, Hishigaki H: Identification of two novel reductases involved in equol biosynthesis in *Lactococcus* strain 20-92. *J Mol Microbiol Biotechnol* 2011, **21**:160-172.

52. Braune A, Blaut M: Evaluation of inter-individual differences in gut bacterial isoflavone bioactivation in humans by PCR-based targeting of genes involved in equol formation. *J Appl Microbiol* 2018, **124**:220-231.

53. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al: A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010, **464**:59-65.

54. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li JH, Sunagawa S, Plichta DR, Gautier L, Pedersen AG, Le Chatelier E, et al: Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nature Biotechnology* 2014, **32**:822-828.

55. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, Lawley TD, Finn RD: A new genomic blueprint of the human gut microbiota. *Nature* 2019, **568**:499-504.

56. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC: New insights from uncultivated genomes of the global human gut microbiome. *Nature* 2019, **568**:505-510.

57. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P, et al: Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* 2019, **176**:649-662 e620.

58. Hur HG, Lay JO, Jr., Beger RD, Freeman JP, Rafii F: Isolation of human intestinal bacteria metabolizing the natural isoflavone glycosides daidzin and genistin. *Arch Microbiol* 2000, **174**:422-428.

59. Mattarelli P, Bonaparte C, Pot B, Biavati B: Proposal to reclassify the three biotypes of *Bifidobacterium longum* as three subspecies: *Bifidobacterium longum* subsp. *longum* subsp. nov., *Bifidobacterium longum* subsp. *infantis* comb. nov. and *Bifidobacterium longum* subsp. *suis* comb. nov. *Int J Syst Evol Microbiol* 2008, **58**:767-772.

60. Ivey KL, Chan AT, Izard J, Cassidy A, Rogers GB, Rimm EB: Role of Dietary Flavonoid Compounds in Driving Patterns of Microbial Community Assembly. *mBio* 2019, **10**:e01205-01219.
61. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, Pollard KS, Sakharova E, Parks DH, Hugenholtz P, et al: A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol* 2021, 39:105-114.

62. Shimojo Y, Ozawa Y, Toda T, Igami K, Shimizu T: Probiotic Lactobacillus paracasei A221 improves the functionality and bioavailability of kaempferol-glucoside in kale by its glucosidase activity. *Sci Rep* 2018, 8:9239.

63. Gaya P, Peiroten A, Alvarez I, Medina M, Landete JM: Production of the bioactive isoflavone O-desmethylangolensin by Enterococcus faecium INIA P553 with high efficiency. *Journal of Functional Foods* 2018, 40:180-186.

64. Gaya P, Peiroten A, Medina M, Landete JM: Isoflavone metabolism by a collection of lactic acid bacteria and bifidobacteria with biotechnological interest. *Int J Food Sci Nutr* 2016, 67:117-124.

65. Kim CC, Healey GR, Kelly WJ, Patchett ML, Jordens Z, Tannock GW, Sims IM, Bell TJ, Hedderley D, Henrissat B, Rosendale DI: Genomic insights from Monoglobus pectinilyticus: a pectin-degrading specialist bacterium in the human colon. *ISME J* 2019, 13:1437-1456.

66. Kim CC, Kelly WJ, Patchett ML, Tannock GW, Jordens Z, Stoklosinski HM, Taylor JW, Sims IM, Bell TJ, Rosendale DI: Monoglobus pectinilyticus gen. nov., sp. nov., a pectinolytic bacterium isolated from human faeces. *Int J Syst Evol Microbiol* 2017, 67:4992-4998.

67. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM: DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007, 57:81-91.

68. Nakatsu CH, Armstrong A, Clavijo AP, Martin BR, Barnes S, Weaver CM: Fecal bacterial community changes associated with isoflavone metabolites in postmenopausal women after soy bar consumption. *PLoS One* 2014, 9:e108924.

69. Guadamuro L, Dohrmann AB, Tebbe CC, Mayo B, Delgado S: Bacterial communities and metabolic activity of faecal cultures from equol producer and non-producer menopausal women under treatment with soy isoflavones. *BMC Microbiol* 2017, 17:93.

70. Guadamuro L, Azcarate-Peril MA, Tojo R, Mayo B, Delgado S: Use of high throughput amplicon sequencing and ethidium monoazide dye to track microbiota changes in an equol-producing menopausal woman receiving a long-term isoflavones treatment. *AIMS Microbiol* 2019, 5:102-116.

71. Schoefer L, Mohan R, Schwiertz A, Braune A, Blaut M: Anaerobic degradation of flavonoids by Clostridium orbiscindens. *Appl Environ Microbiol* 2003, 69:5849-5854.

72. Allen-Vercoe E, Daigneault M, White A, Panaccione R, Duncan SH, Flint HJ, O’Neal L, Lawson PA: Anaerostipes hadrus comb. nov., a dominant species within the human colonic microbiota; reclassification of Eubacterium hadrum Moore et al. 1976. *Anaerobe* 2012, 18:523-529.
73. Gupta A, Dhakan DB, Maji A, Saxena R, P KV, Mahajan S, Pulikkan J, Kurian J, Gomez AM, Scaria J, et al: Association of Flavonifractor plautii, a Flavonoid-Degrading Bacterium, with the Gut Microbiome of Colorectal Cancer Patients in India. mSystems 2019, 4:e00438-00419.

74. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: BLAST+: architecture and applications. BMC Bioinformatics 2009, 10:421.

75. Fu L, Niu B, Zhu Z, Wu S, Li W: CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 2012, 28:3150-3152.

76. Sievers F, Higgins DG: Clustal Omega for making accurate alignments of many protein sequences. Protein Sci 2018, 27:135-145.

77. Steenwyk JL, Buida TJ, 3rd, Li Y, Shen XX, Rokas A: ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol 2020, 18:e3001007.

78. Price MN, Dehal PS, Arkin AP: FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 2010, 5:e9490.

79. Letunic I, Bork P: Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 2021:gkab301.

Figures

Figure 1

Flavonoid O-glycosidase homologs in human gut MAGs. The PID (percent amino acid sequence identity) threshold to the queries (color code) was set to 50 (for abbreviations and details see Table 1). Hits were
filtered for at least 50 occurrences, so that each bubble represents a number of redundant sequences ranging from 50 to 772.

**Figure 2**

Rhamnosidase homologs in human gut MAGs. A PID threshold of 65 was chosen to the queries shown in the color code (for abbreviations and details see Table 1), those with a lower PID are shown in Supplementary Figure 4). Hits were filtered for at least 20 occurrences, so that each bubble represents a number of redundant sequences ranging from 20 to 2176.

**Figure 3**

Two characterized C-deglycosylation gene clusters. (A) Eubacterium cellulosolvens with five genes involved in C-deglycosylation [22] and (B) strain PUE [23, 24], with dgpA encoding an oxidoreductase and dgpBC as oxypuerarin-deglycosylating enzymes, of which all three are involved in C-deglycosylation. Genes encoding deglycosylating enzymes are shown in turquoise, accessory genes in violet. Full GenBank accession number is given for the first gene and only variable digits for the downstream genes.
Figure 4

Homologs to enzymes involved in flavonoid C-deglycosylation using sequences of strain PUE. Hits were filtered for the co-occurrence of all three genes required for C-deglycosylation (dgpABC) in the same MAG. Hits were filtered for at least 2 occurrences, so that each bubble represents a number of redundant sequences ranging from 2 to 820.

Figure 5

Homologs to enzymes involved in daidzein-to-equol transformation in human gut MAGs. Hits were filtered for the co-occurrence of all three genes required for daidzein-to-equol conversion (dzr, ddr, tdr) in a single MAG. For a plot showing the hits to individually occurring genes (mainly dzr), see Supplementary Figure 6.
Figure 6

Reductive degradation of flavonoids depicted with apigenin as flavone example.

Figure 7

Flr sequence hits of human gut MAGs. A PID threshold of 40 was chosen to the queries shown in the color code (for abbreviations and details see Table 1). Hits were filtered for at least 10 occurrences, so that each bubble represents a number of redundant sequences ranging from 10 to 245.

Figure 8
Fcr-like enzymes in human gut MAGs. Hits were filtered for at least 10 occurrences, so that each bubble represents a number of redundant sequences ranging from 10 to 804.

Figure 9

Distribution of Phy homologs in human gut MAGs. Depicted are MAGs with at least five identical amino acid sequences. Hits were filtered for at least ten occurrences, so that each bubble represents a number of redundant sequences ranging from 10 to 578.
Figure 10

Overview of the flavonoid conversion pathways in the human gut based on the MAGs study. Thickness of arrows reflects the abundance of observed hits to the sequences of the characterized flavonoid-transforming enzymes. For abbreviations, see text.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementBASHCommands.txt
- SupplementaryMaterialAB.docx
- SupplementaryTable1.xlsx