Beneficial Microbes, 2023; 14 (3): 255-268

Wageningen Academic
Publishers

Assessment of short-read shotgun sequencing and microbiome analysis of faecal samples to discriminate between equol producers and non-producers

L. Vázquez[1,2], R. Cabrera-Rubio[3,4], J. Tamames[5], B. Mayo[1,2] and A.B. Flórez[1,2*]

[1]Departamento de Microbiología y Bioquímica, Instituto de Productos Lácteos de Asturias (IPLA), Consejo Superior de Investigaciones Científicas (CSIC), Paseo Río Linares s/n, 33300-Villaviciosa, Asturias, Spain; [2]Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Avenida de Roma s/n, 33011-Oviedo, Asturias, Spain; [3]Alimentary Pharmabiotic Centre (APC), Microbiome Institute, University College Cork, Cork, Ireland; [4]Moorepark Teagasc Food Research Centre, Fermoy, Ireland; [5]Departamento de Biología de Sistemas, Centro Nacional de Biotecnología (CNB), Consejo Superior de Investigaciones Científicas (CSIC), Cantoblanco, Madrid, Spain; abflorez@ipla.csic.es

Received: 11 March 2022 / Accepted: 17 January 2023

RESEARCH ARTICLE

Abstract

Among the isoflavones and isoflavone-derived metabolites, equol, which in the human gut is synthesised from daidzein by minority bacterial populations, shows the strongest estrogenic and antioxidant activity. The beneficial effects on human health of isoflavone consumption might be partially or indeed totally attributable to this equol. Although some of the bacterial strains involved in its formation have been identified, the interplay between the composition and functionality of the gut microbiota and equol producer phenotype has hardly been studied. In this study, after shotgun metagenomic sequencing, different pipelines for the taxonomic and functional annotation of sequencing data were used in the search for similarities and differences in the faecal metagenome of equol-producing (n=3) and non-producing (n=2) women, with special focus on equol-producing taxa and their equol-associated genes. The taxonomic profiles of the samples differed significantly depending on the analytical method followed, although the microbial diversity detected by each tool was very similar at the phylum, genus and species levels. Equol-producing taxa were detected in both equol producers and non-producers, but no correlation between the abundance of equol-producing taxa and the equol producing/non-producing phenotype was found. Indeed, functional metagenomic analysis was unable to identify the genes involved in equol production, even in samples from equol producers. By aligning equol operons with the collected metagenomics data, a small number of reads mapping to equol-associated sequences were recognised in samples from both equol producers and equol non-producers, but only two reads mapping onto equol reductase-encoding genes in a sample from an equol producer. In conclusion, the taxonomic analysis of metagenomic data might not be suitable for detecting and quantifying equol-producing microbes in human faeces. Functional analysis of the data might provide an alternative. However, to detect the genetic makeup of the minority gut populations, more extensive sequencing than that achieved in the present study might be required.

Keywords: equol, daidzein, microbiota, microbiome, shotgun metagenomics

1. Introduction

The consumption of isoflavone-containing foods (mostly soy and soy-derived products) has traditionally been linked with positive health effects in humans (for recent reviews see Kim, 2021 and Nakai *et al.*, 2020). Epidemiological studies have further reinforced the relationship between regular isoflavone intake and the relief of menopause symptoms in women, along with a reduction in the risk of cardiovascular and neurodegenerative diseases and certain types of cancer (Chen and Chen, 2021; Finkeldey *et al.*, 2021). Among all isoflavones and their metabolites, equol, which is synthesised from daidzein, has the strongest estrogenic and antioxidant activity (Kładna *et al.*, 2016; Setchell *et al.*, 2005) – activities by which isoflavones and equol are thought to confer their claimed health benefits.

The production of equol from daidzein has been confirmed in all animal species tested so far (Schwen *et al.*, 2012), but only 25-50% of the human population can produce equol, the determining factors including community and dietary habits (Franke *et al.*, 2014). It is now firmly established that the production of equol requires the presence of equol-producing bacteria in the gut microbiota (Bolca *et al.*, 2007; Hall *et al.*, 2007; Setchell and Cole, 2006; Yoshikata *et al.*, 2019), but the actual microbes involved in isoflavone metabolism and equol biosynthesis have not been fully documented (Goris *et al.*, 2021; Rafii *et al.*, 2015; Vázquez *et al.*, 2020). Currently, only a small number of strains belonging to a few gut species have been identified as capable of transforming daidzein into equol (Mayo *et al.*, 2019).

With rare exceptions, equol-producing bacteria from the gut belong to the family *Eggerthellaceae*, and usually to the genera *Adlercreutzia*, *Eggerthella*, *Paraeggerthella* or *Slackia* (Clavel *et al.*, 2014). Strains of *Adlercreutzia equolifaciens* subsp. *equolifaciens*, *A. equolifaciens* subsp. *celatus*, *Adlercreutzia mucosicola*, *Slackia equolifaciens*, *Slackia isoflavoniconvertens*, *Senegalimassilia faecalis* and *Hugonella massiliensis* have been reported to synthetise equol (Mayo *et al.*, 2019; Soukup *et al.*, 2021). Some studies involving 16S rRNA-targeted and shotgun metagenomic techniques have reported a greater relative abundance and prevalence of these species in faecal samples from equol producers than from non-producers (Iino *et al.*, 2019; Zheng *et al.*, 2019), but other studies have reported no such findings (Guadamuro *et al.*, 2019; Nakatsu *et al.*, 2014). Equol-producing taxa have in fact been reported members of the gut microbiota of both equol producers and non-producers (Guadamuro *et al.*, 2019; Iino *et al.*, 2019; Zheng *et al.*, 2019), suggesting our knowledge of the diversity and activity of equol-related taxa in the human gut to be far from complete.

Genomic and proteomic analyses have allowed the biochemical and genetic machinery involved in equol production in some species to be deciphered (Mayo *et al.*, 2019). The products of three genes (*dzr*, *ddr*, and *tdr*), which respectively code for daidzein reductase, dihydrodaidzein reductase and tetrahydrodaidzein reductase, have been identified as the minimum required for equol production (Schröder *et al.*, 2013; Shimada *et al.*, 2010). Additionally, a group of highly conserved genes in the equol cluster has been observed to be induced by daidzein, suggesting their contribution to the synthesis of equol (Flórez *et al.*, 2019; Dufault-Thompson, *et al.*, 2022). Significant among these is a gene encoding a racemase. This is required for maximising equol formation via the synthesis of the *S*-dihydrodaidzein enantiomer, the preferred substrate of the dihydrodaidzein reductase (Ruiz de la Bastida *et al.*, 2021; Shimada *et al.*, 2012). Just as seen for equol-producing taxa, genes involved in equol biosynthesis have also been found in the faeces of both equol producers and equol non-producers (Braune and Blaut, 2018; Vázquez *et al.*, 2017). Clearly, further studies are needed if we are to understand the apparent lack of correlation between the equol-producing phenotype and the presence of equol-related taxa and genes.

In the present work, metagenomic shotgun sequencing was performed to explore the feasibility of using this technique for studying the microbial characteristics of equol production in the human gut. After shotgun sequencing of total microbial DNA in faecal samples from five postmenopausal women, three of whom were equol producers and two equol non-producers, different pipelines and tools for taxonomic and functional annotation were used to analyse the faecal metagenome. Marked differences in microbial diversity and metabolic functions were seen among the faecal samples, depending on the analytical software tool used. However, no conclusive relationships were seen among the taxa, their metabolic activities, and equol-producing phenotype, even when the analyses focused on microorganisms and genes reported to be involved in equol synthesis.

2. Material and methods

Ethical approval, donors and faecal sample collection

This study was approved by the Bioethics Committee of the CSIC (Consejo Superior de Investigaciones Científicas, Spain) and the Regional Ethics Committee for Clinical Research (Servicio de Salud del Principado de Asturias, Spain) (Ref. 84/14). In compliance with the declaration of Helsinki, all participants provided written, informed consent to be included. From a group of women enrolled in a previous study, five were selected on the basis of their equol-producing phenotype and their harbouring equol-related genes in their gut microbiome (Vázquez *et al.*, 2017). Three of the women (W9, W14 and W26) produced equol and harboured equol-associated genes; another woman (W11) did not produce equol and harboured no equol genes; and the last (W13) was negative for equol production but harboured (at least some) equol-related genes. Stool samples were collected in sterile vessels containing Anaerocult A (VWR International, Radnor, PA, USA) by the participants themselves. Samples were transported to the laboratory in <2 h and stored at -80 °C until analysis.

Total DNA extraction

For total microbial DNA extraction, 0.2 g of faeces were suspended in 1.8 ml of phosphate-buffered saline (PBS; VWR International). To eliminate solid and insoluble wastes and intestinal cells, the homogenised suspensions were centrifuged at low speed (60×*g*) for 10 min at 4 °C. Supernatants were transferred to a new tube and centrifuged at 18,400×*g* for 5 min at 4 °C. Bacterial pellets were washed twice with PBS, and total DNA extracted and purified following the protocol described by Zoetendal *et al.* (2006) with some modifications as reported by Yu and Morrison (2004). Briefly, pellets were suspended in 200 µl of a solution containing 20 mmol/l Tris-HCl pH 8.0, 2 mmol/l EDTA, 1.20% Triton X-100, 20 mg/ml lysozyme (Merck, Darmstadt, Germany), 20 U mutanolysin (Sigma-Aldrich, St. Louis, MO, USA) and 10 mg/ml lysostaphin (Sigma-Aldrich), and incubated for 40 min at 37 °C. In addition to chemical lysis, cells were also subjected to treatment with zirconia-silica beads (0.1 and 0.5 mm) using a FastPrep FP120 Cell Disrupter (Qbiogene, Carlsbad, CA, USA) at 5.5 m/s for 30 s; this procedure was repeated three times with a cooling step on ice for 1 min between repetitions. The DNA of the lysate was precipitated overnight at -20 °C with 1/4 volume of 10 mol/l of ammonium acetate (Merck) and 1 volume of 2-propanol (Sigma-Aldrich). After centrifugation, the DNA was suspended in 100 µl of Tris-EDTA (TE) buffer, and incubated consecutively with 0.2 mg/ml RNase (15 min at 37 °C) and 3 mg/ml proteinase K (15 min at 70 °C) (both enzymes from Sigma-Aldrich). For DNA purification, the suspension was mixed with ethanol and transferred to a column of the QIAamp DNA Stool Minikit (Qiagen, Hilden, Germany). The DNA was eluted with 150 µl of sterile molecular grade water (Sigma-Aldrich) and its concentration and quality determined in an Epoch microvolume spectrophotometer (BioTek Instruments, Winooski, VT, USA) to comply with requirements for shotgun sequencing (DNA concentration >300 ng/µl, ratio of A260/280 >1.5 and A26/230 >1.5). Finally, all DNA samples were stored at -20 °C until required.

PCR amplification

The sequences of the *tdr* gene from equol-producing bacteria deposited in the NCBI database were aligned using Clustal Omega software (EMBL-EBI, Hinxton, UK). Degenerated oligonucleotide primers – tdr-F (5’GASAACTCMGCHGAKACCAT 3’) and tdr.qPCR-R (5’ GGMAYYTCCATGTTGTAGGA 3’) – were manually designed matching conserved regions of the gene. The PCR reaction mixture (50 µl) contained 2 µl of total microbial DNA (100 ng), 25 µl of 2× Taq Master Mix Red (Ampliqon, Odense, Denmark), 1.5 µl of each primer (10 µmol/l) and 20 µl of molecular-grade H_2O. Purified DNA of *Adlercreutzia equolifaciens* subsp. *equolifaciens* DSM 19450^T was used as a positive control. The standard PCR programme included one denaturing cycle at 94 °C for 5 min, 35 amplification cycles at 94 °C for 30 s, 50 °C for 1 min, 72 °C for 1 min, and a final extension step at 72 °C for 10 min. Amplicons were analysed on 1% agarose gels after fluorescent staining (GreenSafe Premium; NZYTech, Lisbon, Portugal). Finally, the desired DNA bands in the agarose gel were excised and purified using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, Amersham, UK), and sequenced by cycle extension in an ABI 373 DNA sequencer (Applied Biosystems; ThermoFisher Scientific, Waltham, MA, USA).

Library construction and shotgun sequencing

Library construction and sequencing were performed at Life Sequencing (Valencia, Spain). Briefly, 150 pg of total DNA from each faecal sample were employed to prepare a shotgun metagenomic sequencing library using the Nextera XT DNA Library Prep Kit (Illumina, San Diego, CA, USA). Fragmented DNA was purified using the AMpure XP System (SPRI beads) (Beckman Coulter, Brea, CA, USA), and validated and quantified using a Qubit spectrofluorometer (ThermoFisher Scientific) and an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). After passing quality checks, the libraries were pair-end sequenced using the Illumina HiSeq™ 500 platform (2×150-bp reads), following the manufacturer’s standardised protocols.

Sequences and data analysis

To eliminate adapters and reads with a Phred quality (Q) score of <20, sequencing data were processed using bcl2fastq 2.20 conversion software (https://support.illumina.com/downloads/bcl2fastq-conversion-software-html). The resulting fastq files were pre-processed by sequence quality (removal of all bases at the 3’ end with an average Phred score of <25 over a sliding window of 10 bp) and length (only reads with a minimum length of 120 bases were considered for analysis) using PRINSEQ-Lite v.0.20.4 software (Schmieder and Edwards, 2011). Using BMTagger v.3.101 software (Rotmistrovsky and Agarwala, 2011), clean sequences were screened against the human reference genome (GRCh37d5 Homo sapiens) (https://support.illumina.com/sequencing/sequencing_software/igenome.html, 2019) to remove host reads. Paired-end reads were joined and filtered for duplicates using Picard Tools 2.7.1 –specifically the FastqToSam and MarkDuplicates tools (‘Picard Toolkit’ 2019, Broad Institute, GitHub Repository, http://broadinstitute.github.io/picard/). Reads were then trimmed for bases with low quality scores (≤30) using a modified version of the script trimBWAstyle.pl tool that works directly with BAM files (TrimBWAstyle.usingBam.pl, 2010; https://github.com/genome/genome/blob/master/lib/perl/Genome/Site/TGI/Hmp/HmpSraProcess/trimBWAstyle.usingBam.pl). The ≤30 threshold was chosen to delete all bases with an uncertain quality as defined by Illumina’s EAMMS (End Anchored Max Scoring Segments) filter. The microbial composition within the clean reads was determined using MetaPhlAn2 Species Classifier software (Segata *et al.*, 2012) with default parameters for use with the mpa_v20_m200 marker database, and Kraken 2 software (Wood and Salzberg, 2014) using a k-mer matching algorithm and the standard Kraken 2 database containing all RefSeq for bacterial, archaeal, and viral domains. According to Lu *et al.* (2017), to re-estimate bacterial abundances, the Kraken 2 outputs were examined using the Bracken tool. For each classifier, taxa with a relative abundance of <0.1% were categorised as 'Other'. Subsequently, the function of the reads was assigned using the HUMAnN2 tool by searching the ChocoPhlan and UniRef databases. The HUMAnN2 gene abundance table was regrouped by a mapping of the gene ontology (GO) terms for all categories of bacterial metabolism and dividing the functional data into two files (one stratified and one non-stratified). The SUPER-FOCUS program (Silva *et al.*, 2016) was also used, with the help of the RAPSearch and DIAMOND programs, to assign functionality to the clean reads. Filtered reads were finally assembled using IDBA_UD software (kmers 20-120) (Peng *et al.*, 2012), and contigs with a length of >500 bp kept for analysis.

To analyse the genes involved in equol production and their taxonomic assignment, the fully automatic SqueezeMeta pipeline (Tamames and Puente-Sánchez, 2019) was employed. SqueezeMeta was run using the co-assembly option, pooling the reads from all samples before assembly with Megahit software (Li *et al.*, 2015). Binning and bin-checking procedures were employed to retrieve individual genomes before merging. Results were analysed using SQMtools (Puente-Sánchez *et al.*, 2020). The BLAST tool was also used to compare the obtained metagenomic data against the sequences for equol operons (*A. equolifaciens*, *S. isoflavoniconvertens*, and *S. equolifaciens*) and the sequences for all three reductase genes (*dzr*, *ddr* and *tdr*) deposited in the NCBI database, plus those obtained by PCR from the faecal microbial DNA provided by samples of the equol producer women of this study. In BLAST analyses, cut-offs of 90% identity, 80% minimum query coverage, and 0.0001 e-value were routinely used. When a more permissive filtering was applied, percentages of identity and query coverage were lowered to 70 and 50%, respectively.

Statistical analysis

Statistical analyses were performed in R v.6.0 (https://www.r-project.org/) using tools of Vegan package (Oksanen *et al.*, 2018). The normality of the data was evaluated using the Shapiro-Wilk test. The alpha diversity of faecal microbiotas was determined by the Shannon, Simpson and Observed Species indices. Significant differences in microbiota composition were assessed by permutational multivariate analysis of variance (PERMANOVA) using Vegan software (Adonis 2 function) (Oksanen *et al.*, 2018). Microbial differences between samples (beta diversity) were visualised using non-metric multidimensional scaling (NMDS) plots with the Bray-Curtis dissimilarity index. Continuous variables were described as mean ± standard deviation or median. Significance was set at $P<0.05$.

Metagenomics sequencing data

The faecal metagenomic sequences obtained in this study were deposited into the European Nucleotide Archive (ENA) under project ERP138460. The raw metagenomic data of W9, W11, W13, W14, and W26 samples are available under accession numbers ERR9867199, ERR9867200, ERR9867201, ERR9867202, and ERR9867203, respectively.

3. Results

Sequencing data

After the trimming of low-quality reads and the removal of human-derived contaminating sequences (which ranged from 0.16 to 1.15% per sample), a mean of 26,839,078±2,432,955 reads per faecal sample were obtained with a length varying from 146 to 150 bp, and with an average Phred quality score (Q) of 32 (Supplementary Table S1). After assembly, sequencing data yielded a mean 79,728±29,091 contigs per sample with a length greater than 500 bp.

Taxonomic composition of the microbiota

Although volunteer women constituted a single group, results were often analysed separately for equol producers and non-producers. Differences in community composition of the stools from equol-producing and equol non-producing women was assessed by calculating alpha diversity indices. No significant differences were observed in the overall microbial richness between the samples from women of either phenotype (P=0.4 for Shannon index and P=0.8 for Simpson index) (Supplementary Figure S1). To determine whether there were any differences between samples in terms of microbial relative abundance at the family, genus and species levels, a non-metric multidimensional scaling (NMDS) analysis was conducted (Supplementary Figure S2). Using MetaPhlAn2-derived data, the NMDS analysis revealed the bacterial community structure of the faeces from the equol producers and non-producers to segregate at all taxonomic levels (Supplementary Figure S2A1-A3). However, when using Bracken-derived data, the sample from the equol non-producer harbouring equol-related genes (W13) clustered within the samples from the equol producers (Supplementary Figure S2B1-B3).

Marked differences in taxonomic assignment and relative abundance at the phylum and genus levels were seen between the results obtained with the MetaPhlAn2 and Bracken analyses (Supplementary Figure S3). When combining the results obtained with both classifiers, the reads were affiliated to 31 phyla, 234 families, 619 genera, and 1,303 microbial species. Among them, *Agathobacter rectalis*, *Faecalibacterium prausnitzii*, *Roseburia hominis*, *Parabacteroides distasonis*, and *Bifidobacterium longum* were the five most abundant species (Figure 1). Heat map analysis using species with >0.1% relative abundance in samples showed no significant correlation between microbial populations and the equol-producing phenotype (Figure 2). However, after manual inspection of the heat map graph, the MetaPhlAn2 analysis identified the differential presence (only present, or present in greater abundance) of two species in samples from the women with an equol-producing phenotype – *Alistipes putredinis* and *Bilophila* spp. (Figure 2A). In contrast, closer inspection of the Bracken analysis detected no difference in the samples that could tentatively be associated with equol (Figure 2B).

Equol-producing taxa

Most equol-producing bacteria of intestinal origin belong to the phylum *Actinomycetota*, class *Coriobacteriia* and family *Eggerthellaceae*. In the present work, reads of the *Coriobacteriia* class were detected by MetaPhlAn2 and Bracken in samples from both equol-producing and equol non-producing women, their relative abundance ranging from 0.9 to 3.3% of the total (Figure 3A). The largest number of *Coriobacteriia* reads was detected in samples from the equol-producing W9 and W26. However, the smallest percentage (0.9%) was also detected in the faeces of an equol-producing woman (W14). In fact, reads for *Adlercreutzia equolifaciens* and *Eggerthella* spp. were

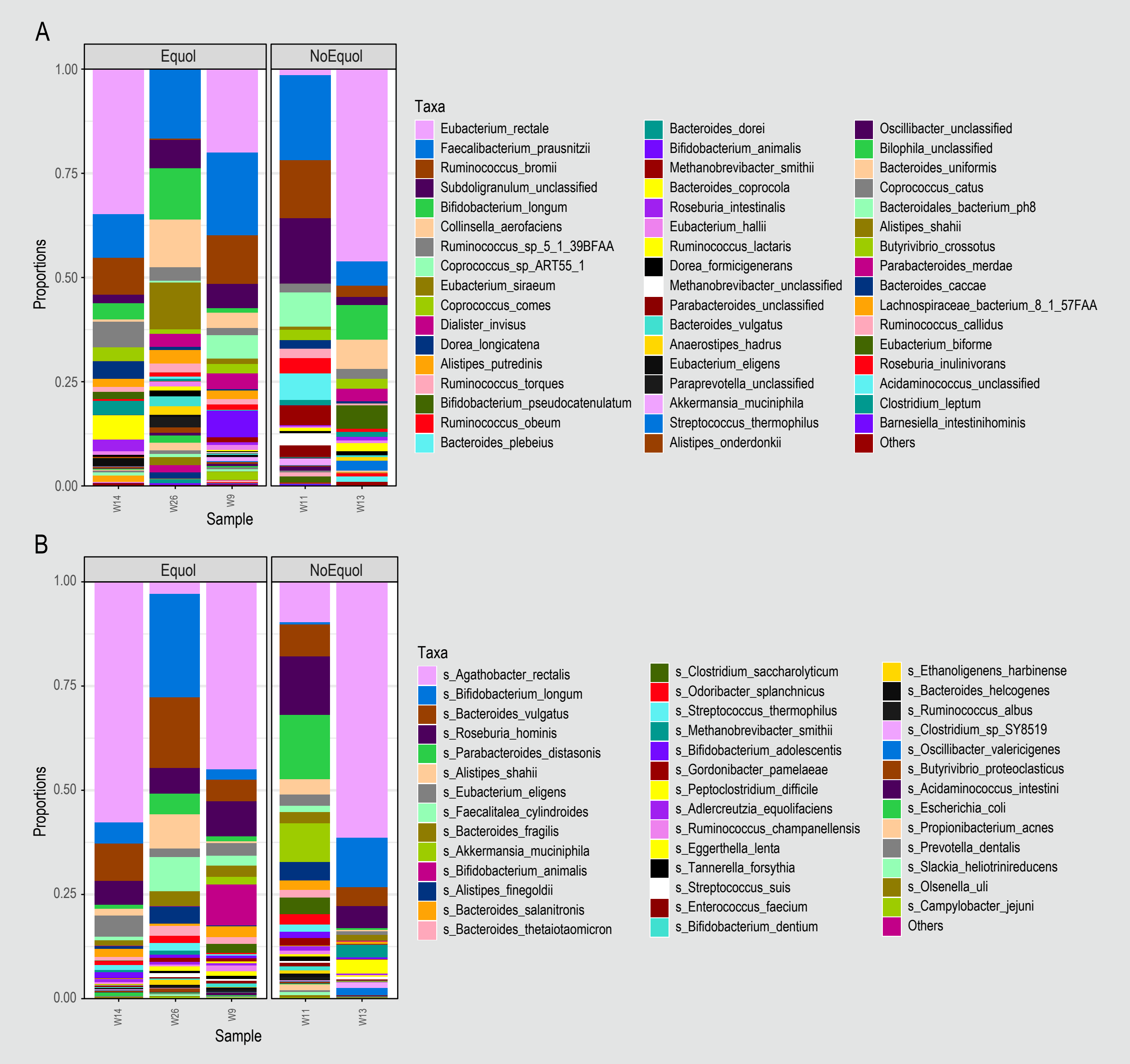

Figure 1. Microbial taxa composition of faecal samples from women of equol-producing (Equol) and equol non-producing (NoEquol) phenotype using two different classifiers: MetaPhlAn2 (A) and Bracken (B). Each stacked bar plot shows the relative abundance of the bacterial communities (>0.1%) at the species level.

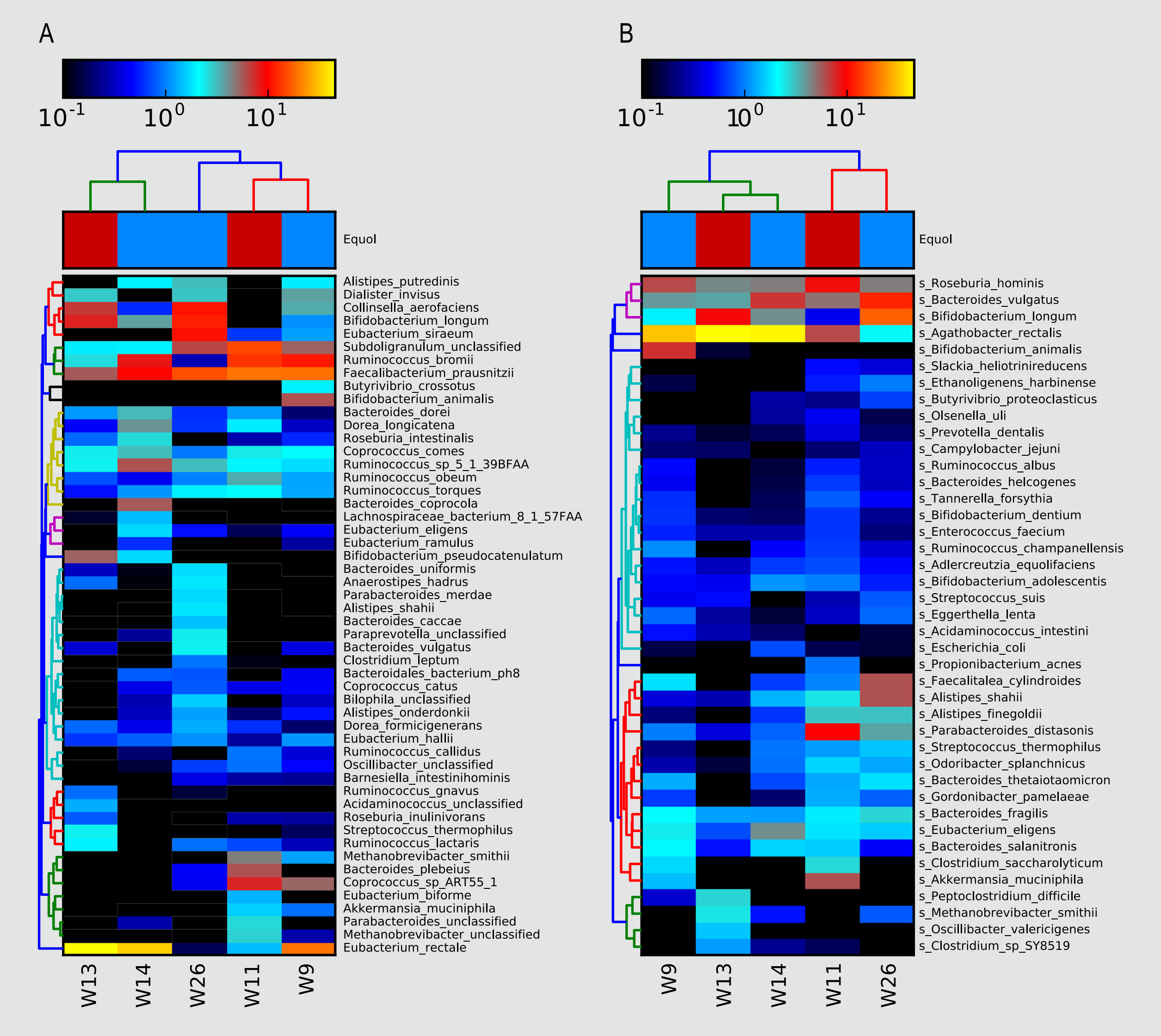

Figure 2. Heat map diagram correlating the relative abundance of bacterial species identified by the software programs MetaPhlAn2 (A) or Bracken (B) in the faecal microbiota of equol-producing (blue squares) and non-producing (red squares) women.

identified in all samples. However, no correlation was found between their relative abundance and the equol-producing phenotype. In a further attempt to identify such correlations, the taxonomic assignment of the SqueezeMeta pipeline was used to identify and quantify equol-taxa associated reads. At the species level, reads of *S. isoflavoniconvertens*, *Eggerthella* spp., *Eggerthella lenta* and *A. equolifaciens* were the most abundant (Figure 3B), although none was exclusively associated to samples of equol producers. Surprisingly, more than 40% of the reads assigned by SqueezeMeta to *Eggerthellaceae* could not be confidently allocated further to any genus/species within this family.

Functionality of microbiota

Taking into account the annotations made by the HUMAnN2 and SUPER-FOCUS programs, the number of different metabolic pathways (Subsystem level 2) present with a relative abundance of >0.001% ranged from 139 to 153 across the samples. While the functional diversity was slightly higher in samples from equol-producers, no significant difference was seen with respect to non-producers (*P*=0.4 for the Shannon index and *P*=0.78 for the Simpson index; Supplementary Figure S4). The most abundant functional categories for the microbiota of both equol-producing and non-producing women were related to the metabolism of carbohydrates, protein and amino acids, nucleotides, cofactors and vitamins, followed by pathways involved in cell wall integrity, membrane transport, stress response and virulence. Over 62% of the

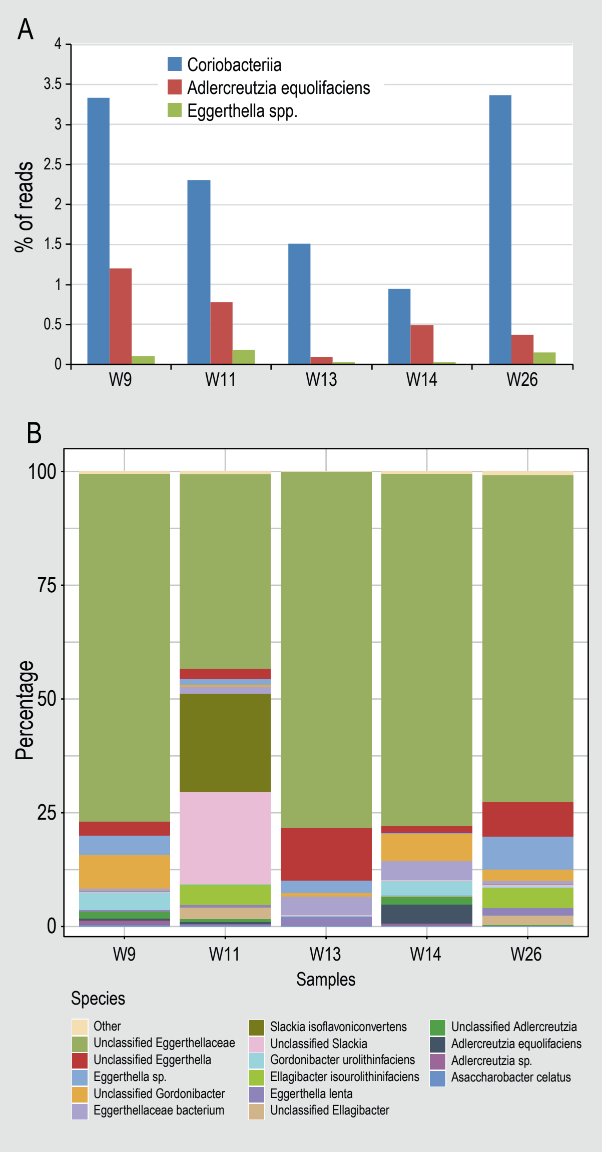

Figure 3. Relative abundance of *Coriobacteriia* and equol-producing bacterial genera and species in faecal samples from equol-producing (W9, W14, W26) and equol non-producing (W11, W13) women (A). Distribution of species belonging to the family *Eggerthellaceae* in the faecal microbiota of equol-producing and non-producing women (B).

total metagenomics reads belonged to genes involved in these metabolic modules. Of the 31 super-pathways (routes composed of at least one base pathway plus additional reactions) or pathways showing abundance differences between phenotypes, 18 were enriched in the faecal metagenome of equol producers, while 13 were enriched in the metagenome of equol non-producers (Figure 4). The largest increase in the metagenome of equol producers was for reads encoding enzymes of the tricarboxylic acid (TCA) cycle, the L-glutamate degradation pathway, and the glyoxylate cycle (in that order).

Equol-associated genes

Neither the HUMAnN2 nor SUPER-FOCUS tools, nor the SqueezeMeta software, identified sequences coding for any equol-related enzymes, nor any other element of the equol operon. Further attempts were therefore made to search for equol-associated reads using the BLAST program, taking as a reference sequences of reductase genes as well as equol operons from *A. equolifaciens*, *S. isoflavoniconvertens*, and *S. equolifaciens* deposited in the NCBI database (accessed in June 2022; Supplementary Table S2). Among the metagenomics data, 90 reads scattered among samples from equol producers and non-producers showed an identity score of up to 90% and a minimum coverage of 80% when compared against sequences from the equol operons utilised, while only a single read highly homologous (97% identity; 100% query coverage) to a segment of the *dzr* gene from *A. equolifaciens* was found in the sample from W14 (an equol producer). In addition, comparison of the metagenomic data to the sequences of the PCR amplicons for the *tdr* gene obtained from the faecal samples of the equol producers in this study, identified one read identical to *tdr* in W14.

Reads for FAD-dependent oxidoreductase- and glutamate synthase-encoding genes showing moderate homology to the reductases involved in equol biosynthesis were identified only when setting more permissive BLAST parameters (cut-offs of 70% identity and 50% minimum coverage). Once again, the number of sequences was the same in samples from the equol producers and non-producers. Finally, to verify that the equol phenotype of the women had not changed since the original study in which they were enrolled, qPCR was performed to newly detect and quantify the *tdr* genes involved in equol production. As in our previous study (Vázquez *et al.*, 2017), 10^3-10^4 copies of *tdr* genes per gram of fresh stools were recorded for the equol-producing women.

4. Discussion

Equol is produced from daidzein via the metabolic activities of certain bacteria in the gut (Mayo *et al.*, 2019). However, the relationships between the composition of the gut microbiota and the metabolic pathways involved in equol production are poorly understood. In this study, shotgun sequencing was used to explore the relationships between the compositional and functional diversity of the gut microbiota and the equol-producing phenotype. The work was performed under the premise that, if found to be useful in predicting the equol producing/non-producing phenotype, this approach would provide a less intrusive and more rapid way of stratifying the equol status of a population than traditional soy/isoflavone challenge tests (Kwak *et al.*, 2009). This would also allow large, publicly-available human gut metagenomic datasets to be mined in

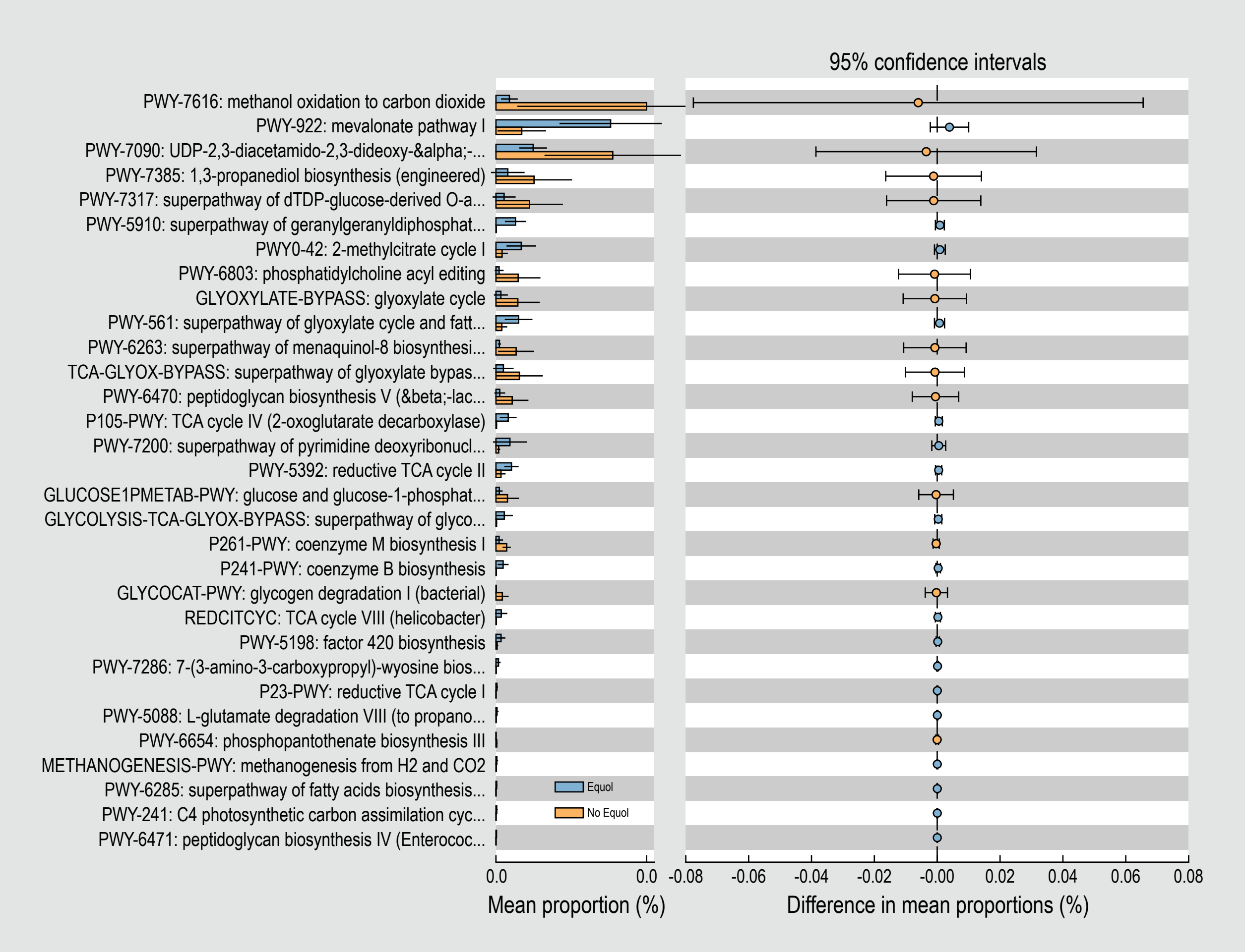

Figure 4. Difference in relative abundance of metabolic routes and pathways in the gut microbiota of equol-producing (Equol) and equol-non-producing (NoEquol) women. Bar plots show the average proportion of sequences, which were predicted to be associated with each pathway. Circles indicates the difference in mean proportions at a 95% confidence interval. Colour code: in blue, pathways enriched in equol-producing women; in orange, pathways enriched in equol-non-producing women.

the search for equol-producing taxa and equol biosynthesis genes and pathways (Goris *et al.*, 2021; Jia *et al.*, 2021). The analysis of microbiomes from healthy controls and from subjects with microbiota-driven diseases (Armour *et al.*, 2019; Mou *et al.*, 2021) would throw further light on the purported beneficial health properties of equol.

After shotgun sequencing the total microbial DNA from the present faecal samples, the community structure of the gut microbiota from the equol producers and equol non-producers was investigated using the standard phylogenetic classification software tools MetaPhlAn2 and Bracken. This taxonomic analysis estimated there to be (in total) over 1000 bacterial species making up the faecal microbiota of the examined women. Similar species diversity has been reported by other authors who performed metataxonomic and metagenomic analyses (Almeida *et al.*, 2019; Qin *et al.*, 2021; Yang *et al.*, 2020). However, large differences were seen in the taxonomic identification results depending on the software program selected, particularly at the genus and species levels. This is likely explained by the different taxonomy assignment strategies followed by MetaPhlAn2 and Bracken. MetaPhlAn2 only profiles taxa included in the clade-specific marker gene database (Lugli *et al.*, 2019a), which tends to underestimate certain microbial populations. In contrast, Bracken uses a k-mer-based methodology that seems to overestimate some other populations (Rajan *et al.*, 2019). As suggested by Lugli *et al.* (2019a), combining these methods might, therefore, provide more accurate identifications.

A correlation has been reported between daidzein intake by equol producers and the number of *Eubacterium*, *Faecalibacterium*, *Clostridium*, and *Collinsella* sequences in their faecal samples (Nakatsu *et al.*, 2014; Guadamuro et al, 2015). However, many sequences of these genera have also been detected in samples from equol non-producers. In the present work, significant differences in terms of species assignment between samples from equol producers and non-producers were found only for some subdominant bacterial populations, and no clear-cut core of bacterial species associated with equol production was identified by any of the software tools used. Reads for *A. putredinis*, *A. communis*, and *Bilophila* spp. – species that have been correlated with a healthy gut microbiome (Averina *et al.*, 2020; Kuhn *et al.*, 2018; Le Chatelier *et al.*, 2013; Wu *et al.*, 2020) – were found exclusively, or in greater abundance, in the faeces of equol producers. However, nothing is yet known on the role of these species in isoflavone metabolism or equol production (Goris *et al.*, 2021; Vázquez *et al.*, 2020). *A. rectalis*, which has previously been associated with isoflavone consumption and equol production (Clavel *et al.*, 2005), was detected as the majority population in samples from two of the equol producers (W9 and W14), but it was also found dominant in that provided by a non-producer (W13). Although interesting, the microbial trends observed between faecal samples of equol-producers and equol non-producers can result from the small size of the faecal samples analysed in this study and the large inter-individual diversity of the human gut microbiota (Qin *et al.*, 2010; Yang *et al.*, 2020).

In agreement with previous reports (Guadamuro *et al.*, 2015; Iino *et al.*, 2019; Nakatsu *et al.*, 2014; Sugiyama *et al.*, 2014; Zheng *et al.*, 2019), all three taxonomic tools used in the present study revealed the presence of sequences of equol-producing species, such as *A. equolifaciens* and *Eggerthella* spp. (Maruo *et al.*, 2008; Yokoyama and Suzuki, 2008). The relative abundance of equol-related bacteria, however, differed widely across the present samples, and did not correlate with the equol phenotype or the amount of equol produced. These data suggest that, as previously suggested (Goris *et al.*, 2021; Vázquez *et al.*, 2020), predicting the equol-related phenotype based on taxonomic analyses is unreliable. This idea is reinforced by a prior finding of an *A. equolifaciens* strain showing, as compared to the equol-producing strain *A. equolifaciens* DSM19450^{T}, a large deletion within the equol operon (Vázquez *et al.*, 2020). Indeed, these authors further reported that the deletions in the equol region were common in *A. equolifaciens* strains. Losses of equol genes in about 40% of the operons among bacteria from the *Eggerthellaceae* family has been recently confirmed (Dufault-Thompson, *et al.*, 2022). All these data together suggest that equol production seems to be a strain-specific character rather than a genus- or species-specific trait.

In the present work, the presence of large numbers of reads belonging to unclassified *Eggerthellaceae* species in the faeces of the equol producers suggests that as-yet unidentified bacteria involved in equol production might be present in these samples. In agreement with this, an *in silico* analysis of the metagenomes in the Human Reference Gut Microbiome (HRGM) database (https://www.mbiomenet.org/HRGM/) recently detected some metagenome-assembled genomes of unidentified members of *Eggerthellaceae* carrying equol-associated genes (Dufault-Thompson *et al.*, 2022; Goris *et al.*, 2021).

The metabolic pathways detected by the HUMAnN2 and SUPER-FOCUS tools in samples from equol producers and non-producers showed only slight differences, matching the scant variation observed for the taxonomic outputs. However, the number of genes associated with the TCA cycle was moderately increased in the samples from the equol producers. The TCA cycle plays a central role in the catabolism of carbohydrates, fats and proteins, synthesis certain amino acids, and provides the reducing power (NADH) required by multiple cellular reactions. Increases in the number of bacterial genes associated with it have also been reported after the intake of green tea polyphenols (Zhou *et al.*, 2020). The TCA, L-glutamate and glyoxylate pathways may contribute to butyrate and propionate formation from carbohydrates in the lumen of the colon, thus contributing to the maintenance of a healthy gut environment (Ahn *et al.*, 2016; Louis and Flint, 2016). Indeed, increased butyric and propionic acid production has been reported in faecal cultures from equol producers (Guadamuro *et al.*, 2017). Small increases in the number of sequences with homology to genes coding for enzymes involved in the synthesis of isopentenyl diphosphate and geranylgeranyl diphosphate (precursors of vitamin A; Srinivasan and Buys, 2019), as well as lipopolysaccharides and enzymes that take part in the 2-methylcitrate cycle and methanol oxidation reactions (Upton and McKinney, 2007), were also detected in the metagenome of the equol producer women. However, nothing is currently known about the connections between the abundance of genes associated with these pathways and equol production.

Surprisingly, the functional metabolic analyses failed to detect sequences related to equol production beyond a single read each for *dzr* and *tdr* in a sample from an equol producing woman, even though the metagenome dataset of the samples in this work (>20 Mb each; Supplementary Table S1) is larger than those reported in other gut metagenomic studies (Lloyd-Price *et al.*, 2019; Oliver *et al.*, 2021). Extending the search to all genes of the equol operons (13-20 ORFs) (Dufault-Thompson *et al.*, 2022; Flórez *et al.*, 2019) an increased number of reads were identified. However, these were not enough to assemble into contigs or pathways. Furthermore, the reads were found in similar numbers in samples from equol-producing and non-producing women. In agreement with these results, Zheng *et al.* (2019) detected equol genes equally in the microbiomes of faeces from equol producers and non-producers, and detected the region encompassing the three reductase genes in only three out of 59 equol producers. Although of high prevalence, deletions within the equol operon do not always affect the same genes, as recently reported (Dufault-Thompson, *et al.*, 2022), which suggests that defining the equol producing phenotype based on the detection of equol operon genes different from those encoding the reductases might increase the probability of miscategorisation. Additionally, the difficulty in detecting such sequences might be explained by the low number of equol genes within the entire faecal metagenome, which have been shown to range (and thus the number of equol-producing microbes) between 10^3 to 10^5 copies/g faeces (Vázquez *et al.*, 2017). The results of this study are supported by recently published analysis of large metagenomics datasets (Dufault-Thompson *et al.*, 2022; Goris *et al.*, 2021). From out of 32,259 metagenome samples from 114 projects, the latter authors identified only 60 contigs containing putative equol-production gene clusters from 18 of the projects, while the equol-producing phenotype in the population (and thus the faecal samples containing such gene clusters) should range in between 25 to 50% (Bolca *et al.*, 2007; Hall *et al.*, 2007; Setchell *et al.*, 2006).

The presence in faecal samples of large amounts of DNA from the majority microbial populations may reduce the capacity to detect sequences low or very low in abundance (Pereira-Marques *et al.*, 2019). To overcome this restriction, recent studies have implemented different target enrichment sequencing approaches to allow detection of specific pathogen, viral or bacterial genomes present in low relative abundance in complex samples (Deng *et al.*, 2020; Kechin *et al.*, 2020; Lugli *et al.*, 2019b). Minority populations acting on dietary polyphenols (isoflavones and others) may have a pivotal influence on gut physiology and human health.

5. Conclusions

In summary, the gut bacterial communities of the equol producers and equol non-producers tested in this study were practically identical, even in terms of currently known equol-related taxa. Members of the genera *Alistipes* and *Bilophila* were, however, detected exclusively or in greater abundance in the faeces of the equol producers. The relationship between these species and the equol-producing phenotype has yet to be clarified. The functional analysis of the metagenomic data failed to detect any pathway or complete gene involved in equol synthesis. To defeat the challenges affecting the detection of minority taxa and genes within the gut microbiota and its associated microbiome, more extensive sequencing, target enrichment, and/or more powerful computational resources may be required.

Supplementary material

Supplementary material can be found online at https://doi.org/10.3920/BM2022.0027.

Figure S1. Box plots representing alpha Shannon Simpson, and Richenest (Rspecies specnumber) diversity indices of identified species in the faecal microbiota of women with an equol producing and equol non-producing phenotype using MethaPhlAn 2 and Bracken.

Figure S2. Non-metric Multidimensional Scaling (NMDS) analysis between the equol producers and non-producers, using taxonomic data from MethaPhlAn2 and Bracken at family, genus and species levels.

Supplementary

Figure S3. Microbial composition of faecal samples from women with equol producing and equol non-producing phenotypes using as a classifier MetaPhlAn 2 and Bracken.

Figure S4. Box plots depicting alpha diversity indices (Shannon, Simpson and Richenest) in the metagenome of faecal samples from women of equol producing and equol non-producing phenotypes.

Table S1. Summary of metagenomics data obtained after shotgun sequencing, removal of human-derived contaminant sequences, and assembly of reads from total microbial DNA of faecal samples from equol-producing and equol-non-producing women.

Table S2. Sequences used to search for equol-associated genes in the metagenomic data of this study.

Acknowledgements

This research was supported by projects from the Spanish Ministry of Science and Innovation (PID2019-110549RB-I00/AEI/10.13039/501100011033) and Asturias Principality (AYUD/2021/50916). L.V. was supported by a research contract of the FPI Program from MINECO (BES-2015-072285).

Conflict of interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

Ahn, S., Jung, J., Jang, I.A., Madsen, E.L. and Park, W., 2016. Role of glyoxylate shunt in oxidative stress response. Journal of Biological Chemistry 291: 11928-11938. https://doi.org/10.1074/jbc.M115.708149

Almeida, A., Mitchell, A.L., Boland, M., Forster, S.C., Gloor, G.B., Tarkowska, A., Lawley, T.D. and Finn, R.D., 2019. A new genomic blueprint of the human gut microbiota. Nature 568: 499-504. https://doi.org/10.1038/s41586-019-0965-1

Armour, C.R., Nayfach, S., Pollard, K.S. and Sharpton, T.J., 2019. A Metagenomic meta-analysis reveals functional signatures of health and disease in the human gut microbiome. mSystems 4: e00332-18. https://doi.org/10.1128/mSystems.00332-18

Averina, O.V., Kovtun, A.S., Polyakova, S.I., Savilova, A.M., Rebrikov, D.V. and Danilenko, V.N., 2020. The bacterial neurometabolic signature of the gut microbiota of young children with autism spectrum disorders. Journal of Medical Microbiology: 69: 558-571. https://doi.org/10.1099/jmm.0.001178

Bolca, S., Possemiers, S., Herregat, A., Huybrechts, I., Heyerick, A., De Vriese, S., Verbruggen, M., Depypere, H., De Keukeleire, D., Bracke, M., De Henauw, S., Verstraete, W. and Van de Wiele, T., 2007. Microbial and dietary factors are associated with the equol producer phenotype in healthy postmenopausal women. Journal of Nutrition 137: 2242-2246. https://doi.org/10.1093/jn/137.10.2242

Braune, A. and Blaut, M., 2018. Evaluation of inter-individual differences in gut bacterial isoflavone bioactivation in humans by PCR-based targeting of genes involved in equol formation. Journal of Applied Microbiology 124: 220-231. https://doi.org/10.1111/jam.13616

Chen, L.R. and Chen, K.H., 2021. Utilization of isoflavones in soybeans for women with menopausal syndrome: an overview. International Journal of Molecular Sciences 22: 3212. https://doi.org/10.3390/ijms22063212

Clavel, T., Fallani, M., Lepage, P., Levenez, F., Mathey, J., Rochet, V., Sérézat, M., Sutren, M., Henderson, G., Bennetau-Pelissero, C., Tondu, F., Blaut, M., Doré, J. and Coxam, V., 2005. Isoflavones and functional foods alter the dominant intestinal microbiota in postmenopausal women. Journal of Nutrition 135: 2786-2792. https://doi.org/10.1093/jn/135.12.2786

Clavel, T., Lepage, P. and Charrier, C., 2014. The family *Coriobacteriaceae*. In: Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E. and Thompson, F. (eds.) The Prokaryotes – *Actinobacteria*. Springer, Berlin, Germany, pp. 201-238.

Deng, X., Achari, A., Federman, S., Yu, G., Somasekar, S., Bártolo, I., Yagi, S., Mbala-Kingebeni, P., Kapetshi, J., Ahuka-Mundeke, S., Muyembe-Tamfum, J.-J., Ahmed, A.A., Ganesh, V., Tamhankar, M., Patterson, J.L., Ndembi, N., Mbanya, D., Kaptue, L., McArthur, C., Muñoz-Medina, J.E., Gonzalez-Bonilla, C.R., López, S., Arias, C.F., Arevalo, S., Miller, S., Stone, M., Busch, M., Hsieh, K., Messenger, S., Wadford, D.A., Rodgers, M., Cloherty, G., Faria, N.R., Thézé, J., Pybus, O.G., Neto, Z., Morais, J., Taveira, N., R. Hackett, J. and Chiu, C.Y., 2020. Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. Nature Microbiology 5: 443-454. https://doi.org/10.1038/s41564-019-0637-9

Dufault-Thompson, K., Hall, B. and Jiang, X., 2022. Taxonomic distribution and evolutionary analysis of the equol biosynthesis gene cluster. BMC Genomics 23: 182. https://doi.org/10.1186/s12864-022-08426-7

Finkeldey, L., Schmitz, E. and Ellinger, S., 2021. Effect of the intake of isoflavones on risk factors of breast cancer – A systematic review of randomized controlled intervention studies. Nutrients 13: 2309. https://doi.org/10.3390/nu13072309

Flórez, A.B., Vázquez, L., Rodríguez, J., Redruello, B. and Mayo, B., 2019. Transcriptional regulation of the equol biosynthesis gene cluster in *Adlercreutzia equolifaciens* DSM19450^{T}. Nutrients 11: 993. https://doi.org/10.3390/nu11050993

Franke, A.A., Lai, J.F. and Halm, B.M., 2014. Absorption, distribution, metabolism, and excretion of isoflavonoids after soy intake. Archives of Biochemistry and Biophysics 59: 24-28. https://doi.org/10.1016/j.abb.2014.06.007

Goris, T., Cuadrat, R.R.C. and Braune, A., 2021. Flavonoid-modifying capabilities of the human gut microbiome – An *in silico* study. Nutrients 13: 2688. https://doi.org/10.3390/nu13082688

Guadamuro, L., Azcárate-Peril, M.A., Tojo, R., Mayo, B. and Delgado, S., 2019. 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 Microbiology 5: 102-116. https://doi.org/10.3934/microbiol.2019.1.102

Guadamuro, L., Delgado, S., Redruello, B., Flórez, A.B., Suárez, A., Martínez-Camblor, P. and Mayo, B., 2015. Equol status and changes in fecal microbiota in menopausal women receiving long-term treatment for menopause symptoms with a soy-isoflavone concentrate. Frontiers in Microbiology 6: 777. https://doi.org/10.3389/fmicb.2015.00777

Guadamuro, L., Dohrmann, A.B., Tebbe, C.C., Mayo, B. and Delgado, S., 2017. Bacterial communities and metabolic activity of faecal cultures from equol producer and non-producer menopausal women under treatment with soy isoflavones. BMC Microbiology 17: 93. https://doi.org/10.1186/s12866-017-1001-y

Hall, M.C., O'Brien, B. and McCormack, T., 2007. Equol producer status, salivary estradiol profile and urinary excretion of isoflavones in Irish Caucasian women, following ingestion of soymilk. Steroids 72: 64-70. https://doi.org/10.1016/j.steroids.2006.10.010

Iino, C., Shimoyama, T., Iino, K., Yokoyama, Y., Chinda, D., Sakuraba, H., Fukuda, S. and Nakaji, S., 2019. Daidzein intake is associated with equol producing status through an increase in the intestinal bacteria responsible for equol production. Nutrients 11: 433. https://doi.org/10.3390/nu11020433

Jia, B., Han, X., Kim, K.H. and Jeon, C.O., 2021. Discovery and mining of enzymes from the human gut microbiome. Trends in Biotechnology 40: 240-254. https://doi.org/10.1016/j.tibtech.2021.06.008

Kechin, A., Borobova, V., Boyarskikh, U., Khrapov, E., Subbotin, S. and Filipenko, M., 2020. NGS-PrimerPlex: High-throughput primer design for multiplex polymerase chain reactions. PLoS Computational Biology 16: e1008468. https://doi.org/10.1371/journal.pcbi.1008468

Kim, I.S., 2021. Current perspectives on the beneficial effects of soybean isoflavones and their metabolites for humans. Antioxidants 10: 1064. https://doi.org/10.3390/antiox10071064

Kładna, A., Berczyński, P., Kruk, I., Piechowska, T. and Aboul-Enein, H.Y., 2016. Studies on the antioxidant properties of some phytoestrogens. Luminescence 31: 1201-1206. https://doi.org/10.1002/bio.3091

Kuhn, K.A., Schulz, H.M., Regner, E.H., Severs, E.L., Hendrickson, J.D., Mehta, G., Whitney, A.K., Ir, D., Ohri, N., Robertson, C.E., Frank, D.N., Campbell, E.L. and Colgan, S.P., 2018. Bacteroidales recruit IL-6-producing intraepithelial lymphocytes in the colon to promote barrier integrity. Mucosal Immunology 11: 357-368. https://doi.org/10.1038/mi.2017.55

Kwak, H.S., Park, S.Y., Kim, M.G., Yim, C.H., Yoon, H.K. and Han, K.O., 2009. Marked individual variation in isoflavone metabolism after a soy challenge can modulate the skeletal effect of isoflavones in premenopausal women. Journal of Korean Medical Science 24: 867-873. https://doi.org/10.3346/jkms.2009.24.5.867

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.-M., Kennedy, S., Leonard, P., Li, J., Burgdorf, K., Grarup, N., Jørgensen, T., Brandslund, I., Nielsen, H.B., Juncker, A.S., Bertalan, M., Levenez, F., Pons, N., Rasmussen, S., Sunagawa, S., Tap, J., Tims, S., Zoetendal, E.G., Brunak, S., Clément, K., Doré, J., Kleerebezem, M., Kristiansen, K., Renault, P., Sicheritz-Ponten, T., de Vos, W.M., Zucker, J.-D., Raes, J., Hansen, T., Guedon, E., Delorme, C., Layec, S., Khaci, G., van de Guchte, M., Vandemeulebrouck, G., Jamet, A., Dervyn, R., Sanchez, N., Maguin, E., Haimet, F., Winogradski, Y., Cultrone, A., Leclerc, M., Juste, C., Blottière, H., Pelletier, E., LePaslier, D., Artiguenave, F., Bruls, T., Weissenbach, J., Turner, K., Parkhill, J., Antolin, M., Manichanh, C., Casellas, F., Boruel, N., Varela, E., Torrejon, A., Guarner, F., Denariaz, G., Derrien, M., van Hylckama Vlieg, J.E.T., Veiga, P., Oozeer, R., Knol, J., Rescigno, M., Brechot, C., M'Rini, C., Mérieux, A., Yamada, T., Bork, P., Wang, J., Ehrlich, S.D., Pedersen, O. and Meta, H.I.T.c., 2013. Richness of human gut microbiome correlates with metabolic markers. Nature 500: 541-546. https://doi.org/10.1038/nature12506

Li, D., Liu, C. M., Luo, R., Sadakane, K. and Lam, T.W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31: 1674-1676. https://doi.org/10.1093/bioinformatics/btv033

Lloyd-Price, J., Arze, C., Ananthakrishnan, A.N., Schirmer, M., Avila-Pacheco, J., Poon, T.W., Andrews, E., Ajami, N.J., Bonham, K.S., Brislawn, C.J., Casero, D., Courtney, H., Gonzalez, A., Graeber, T.G., Hall, A.B., Lake, K., Landers, C.J., Mallick, H., Plichta, D.R., Prasad, M., Rahnavard, G., Sauk, J., Shungin, D., Vázquez-Baeza, Y., White, R.A., Bishai, J., Bullock, K., Deik, A., Dennis, C., Kaplan, J.L., Khalili, H., McIver, L.J., Moran, C.J., Nguyen, L., Pierce, K.A., Schwager, R., Sirota-Madi, A., Stevens, B.W., Tan, W., ten Hoeve, J.J., Weingart, G., Wilson, R.G., Yajnik, V., Braun, J., Denson, L.A., Jansson, J.K., Knight, R., Kugathasan, S., McGovern, D.P.B., Petrosino, J.F., Stappenbeck, T.S., Winter, H.S., Clish, C.B., Franzosa, E.A., Vlamakis, H., Xavier, R.J., Huttenhower, C. and the IBDMDB investigators, 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569: 655-662. https://doi.org/10.1038/s41586-019-1237-9

Louis, P. and Flint, H. J., 2017. Formation of propionate and butyrate by the human colonic microbiota. Environmental Microbiology 19: 29-41. https://doi.org/10.1111/1462-2920.13589

Lu, J., Breitwieser, F.P., Thielen, P. and Salzberg, S.L., 2017. Bracken: estimating species abundance in metagenomics data. PeerJ Computer Science 3: e104. https://doi.org/10.7717/peerj-cs.104

Lugli, G. A., Duranti, S., Milani, C., Mancabelli, L., Turroni, F., Sinderen, D. V., and Ventura, M. 2019b. Uncovering bifidobacteria via targeted sequencing of the mammalian gut microbiota. Microorganisms 7: 535. https://doi.org/10.3390/microorganisms7110535

Lugli, G.A., Milani, C., Mancabelli, L., Turroni, F., van Sinderen, D. and Ventura, M., 2019a. A microbiome reality check: limitations of *in silico*-based metagenomic approaches to study complex bacterial communities. Environmental Microbiology Reports 11: 840-847. https://doi.org/10.1111/1758-2229.12805

Maruo, T., Sakamoto, M., Ito, C., Toda, T. and Benno, Y., 2008. *Adlercreutzia equolifaciens* gen. nov., sp. nov., an equol-producing bacterium isolated from human faeces, and emended description of the genus *Eggerthella*. International Journal of Systematic and Evolutionary Microbiology 58: 1221-1227. https://doi.org/10.1099/ijs.0.65404-0

Mayo, B., Vázquez, L. and Flórez, A.B., 2019. Equol: A bacterial metabolite from the daidzein isoflavone and its presumed beneficial health effects. Nutrients 11: 2231. https://doi.org/10.3390/nu11092231

Mou, Z., Yang, Y., Hall, A.B. and Jiang, X., 2021. The taxonomic distribution of histamine-secreting bacteria in the human gut microbiome. BMC Genomics 22: 695. https://doi.org/10.1186/s12864-021-08004-3

Nakai, S., Fujita, M. and Kamei, Y., 2020. Health promotion effects of soy isoflavones. Journal of Nutritional Science and Vitaminology 66: 502-507. https://doi.org/10.3177/jnsv.66.502

Nakatsu, C.H., Arsmstrong, A., Cavijo, AP., Martin, B.R., Barnes, S. and Weaver, C.M., 2014. Fecal bacterial community changes associated with isoflavone metabolites in postmenopausal women after soy bar consumption. PLoS ONE 9: e108924. https://doi.org/10.1371/journal.pone.0108924

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlinn, D., Ouellette, M.H., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C.J.F. and Weedon, J., 2018. Ordination methods, diversity analysis and other functions for community and vegetation ecologists. Vegan: Community Ecology Package. Version 2.5-1. Available at: https://CRAN.R-project.org/package=vegan.

Oliver, A., Chase Alexander, B., Weihe, C., Orchanian Stephanie, B., Riedel Stefan, F., Hendrickson Clark, L., Lay, M., Sewall Julia, M., Martiny Jennifer, B.H. and Whiteson, K., 2021. High-fiber, whole-food dietary intervention alters the human gut microbiome but not fecal short-chain fatty acids. mSystems 6: e00115-00121. https://doi.org/10.1128/mSystems.00115-21

Peng, Y., Leung, H.C., Yiu, S.M. and Chin, F.Y., 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28: 1420-1428. https://doi.org/10.1093/bioinformatics/bts174

Pereira-Marques, J., Hout, A., Ferreira, R.M., Weber, M., Pinto-Ribeiro, I., van Doorn, L-J., Knetsch, C.W. and Figueiredo, C., 2019. Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis. Frontiers in Microbiology 10: 1277. https://doi.org/10.3389/fmicb.2019.01277

Puente-Sánchez, F., García-García, N. and Tamames, J., 2020. SQMtools: automated processing and visual analysis of 'omics data with R and anvi'o. BMC Bioinformatics 21: 358. https://doi.org/10.1186/s12859-020-03703-2

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Antolin, M., Artiguenave, F., Blottiere, H., Borruel, N., Bruls, T., Casellas, F., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Forte, M., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Jamet, A., Juste, C., Kaci, G., Kleerebezem, M., Knol, J., Kristensen, M., Layec, S., Le Roux, K., Leclerc, M., Maguin, E., Melo Minardi, R., Oozeer, R., Rescigno, M., Sanchez, N., Tims, S., Torrejon, T., Varela, E., de Vos, W., Winogradsky, Y., Zoetendal, E., Bork, P., Ehrlich, S.D., Wang, J. and Meta, H.I.T.C., 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464: 59-65. https://doi.org/10.1038/nature08821

Rafii, F. 2015. The role of colonic bacteria in the metabolism of the natural isoflavone daidzin to equol. Metabolites 5: 56-73. https://doi.org/10.3390/metabo5010056

Rajan, S.K., Lindqvist, M., Brummer, R.J., Schoultz, I. and Repsilber, D., 2019. Phylogenetic microbiota profiling in fecal samples depends on combination of sequencing depth and choice of NGS analysis method. PLoS ONE 14: e0222171. https://doi.org/10.1371/journal.pone.0222171

Rotmistrovsky, K. and Agarwala, R., 2011. BMTagger: best match tagger for removing human reads from metagenomics datasets. Available at: http://www.mmnt.net/db/0/0/ftp.ncbi.nlm.nih.gov/pub/agarwala/bmtagger.

Ruiz de la Bastida, A., Peirotén, Á., Langa, S., Arqués, J.L. and Landete, J.M., 2021. Heterologous production of equol by lactic acid bacteria strains in culture medium and food. International Journal of Food Microbiology 360: 109328. https://doi.org/10.1016/j.ijfoodmicro.2021.109328

Schmieder, R. and Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27: 863-864. https://doi.org/10.1093/bioinformatics/btr026

Schröder, C., Matthies, A., Engst, W., Blaut, M. and Braune, A., 2013. Identification and expression of genes involved in the conversion of daidzein and genistein by the equol-forming bacterium *Slackia isoflavoniconvertens*. Applied and Environmental Microbiology 79: 3494-3502. https://doi.org/10.1128/AEM.03693-12

Schwen, R.J., Nguyen, L. and Charles, R.L., 2012. Elucidation of the metabolic pathway of S-equol in rat, monkey and man. Food and Chemical Toxicology 50: 2074-2083. https://doi.org/10.1016/j.fct.2012.03.048

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O. and Huttenhower, C., 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nature Methods 8: 811-814. https://doi.org/10.1016/j.fct.2012.03.048

Setchell, K.D. and Cole, S.J. 2006. Method of defining equol-producer status and its frequency among vegetarians. Journal of Nutrition 136. 2188-2193. https://doi.org/10.1093/JN/136.8.2188

Setchell, K.D.R., Clerici, C., Lephart, E.D., Cole, S.J., Heenan, C., Castellani, D., Wolfe, B.E., Nechemias-Zimmer, L., Brown, N.M., Lund, T.D., Handa, R.J. and Heubi, J.E., 2005. S-Equol, a potent ligand for estrogen receptor β, is the exclusive enantiomeric form of the soy isoflavone metabolite produced by human intestinal bacterial flora1-42,3. American Journal of Clinical Nutrition 81: 1072-1079. /https://doi.org/10.1093/ajcn/81.5.1072

Shimada, Y., Takahashi, M., Miyazawa, N., Abiru, Y., Uchiyama, S., Hishigakia, H., 2012. Identification of a novel dihydrodaidzein racemase essential for biosynthesis of equol from daidzein in *Lactococcus* sp. strain 20-92. Applied and Environmental Microbiology 78: 4902-4907. https://doi.org/10.1128/AEM.00410-12

Shimada, Y., Yasuda, S., Takahashi, M., Hayashi, T., Miyazawa, N., Sato, I., Abiru, Y., Uchiyama, S. and Hishigaki, H., 2010. 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. Applied and Environmental Microbiology 76: 5892-5901. https://doi.org/10.1128/AEM.01101-10

Silva, G.G.Z., Green, K.T., Dutilh, B.E. and Edwards, R.A., 2016. SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data. Bioinformatics 32: 354-361. https://doi.org/10.1093/bioinformatics/btv584

Soukup, S.T., Stoll, D.A., Danylec, N., Schoepf, A., Kulling, S.E. and Huch, M. 2021. Metabolism of daidzein and genistein by gut bacteria of the Class *Coriobacteriia*. Foods 10: 2741. https://doi.org/10.3390/foods10112741

Srinivasan, K. and Buys, E.M., 2019. Insights into the role of bacteria in vitamin A biosynthesis: Future research opportunities. Critical Reviews in Food Science and Nutrition 59: 3211-3226. https://doi.org/10.1080/10408398.2018.1546670

Sugiyama, Y., Nagata, Y., Fukuta, F., Takayanagi, A., Masumori, N., Tsukamoto, T., Akasaka, H., Ohnishi, H., Saito, S., Miura, T., Moriyama, K., Tsuji, H., Akaza, H. and Mori, M., 2014. Counts of *Slackia* sp. strain NATTS in intestinal flora are correlated to serum concentrations of equol both in prostate cancer cases and controls in Japanese men. Asian Pacific Journal of Cancer Prevention 15: 2693-2697. https://doi.org/10.7314/apjcp.2014.15.6.2693

Tamames, J. and Puente-Sánchez, F., 2019. SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. Frontiers in Microbiology 9: 3349. https://doi.org/10.3389/fmicb.2018.03349

Upton, A.M. and McKinney, J.D., 2007. Role of the methylcitrate cycle in propionate metabolism and detoxification in *Mycobacterium smegmatis*. Microbiology 153: 3973-3982. https://doi.org/10.1099/mic.0.2007/011726-0

Vázquez, L., Flórez, A.B., Redruello, B. and Mayo, B., 2020. Metabolism of soy isoflavones by intestinal bacteria: genome analysis of an *Adlercreutzia equolifaciens* strain that does not produce equol. Biomolecules 10: 950. https://doi.org/10.3390/biom10060950

Vázquez, L., Guadamuro, L., Giganto, F., Mayo, B. and Flórez, A.B., 2017. Development and use of a real-time quantitative PCR method for detecting and quantifying equol-producing bacteria in human faecal samples and slurry cultures. Frontiers in Microbiology 8: 1155. https://doi.org/10.3389/fmicb.2017.01155

Wood, D.E. and Salzberg, S.L., 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biology 15: R46. https://doi.org/10.1186/gb-2014-15-3-r46.

Wu, Y., Bible, P.W., Long, S., Ming, W.-k., Ding, W., Long, Y., Wen, X., Li, X., Deng, X., Deng, Y., Guo, S., Doçi, C.L., Wei, L., Chen, H. and Wang, Z., 2020. Metagenomic analysis reveals gestational diabetes mellitus-related microbial regulators of glucose tolerance. Acta Diabetologica 57: 569-581. https://doi.org/10.1007/s00592-019-01434-2

Yang, J., Pu, J., Lu, S., Bai, X., Wu, Y., Jin, D., Cheng, Y., Zhang, G., Zhu, W., Luo, X., Rosselló-Móra, R. and Xu, J., 2020. Species-level analysis of human gut microbiota with metataxonomics. Frontiers in Microbiology 11: 2029. https://doi.org/10.3389/fmicb.2020.02029

Yokoyama, S. and Suzuki, T., 2008. Isolation and characterization of a novel equol-producing bacterium from human feces. Bioscience, Biotechnology and Biochemestry 72: 2660-2666. https://doi.org/10.1271/bbb.80329

Yu, Z. and Morrison, M., 2004. Improved extraction of PCR-quality community DNA from digesta and fecal samples. BioTechniques 36: 808-812. https://doi.org/10.2144/04365ST04

Zheng, W., Ma, Y., Zhao, A., He, T., Lyu, N., Pan, Z., Mao, G., Liu, Y., Li, J., Wang, P., Wang, J., Zhu, B. and Zhang, Y., 2019. Compositional and functional differences in human gut microbiome with respect to equol production and its association with blood lipid level: a cross-sectional study. Gut Pathogens 11: 20. https://doi.org/10.1186/s13099-019-0297-6

Zhou, J., Tang, L., Shen, C.L. and Wang, J.S., 2020. Green tea polyphenols boost gut-microbiota-dependent mitochondrial TCA and urea cycles in Sprague-Dawley rats. Journal of Nutritional Biochemistry 81: 108395. https://doi.org/10.1016/j.jnutbio.2020.108395

Zoetendal, E.G. Heilig, H.G.H.J., Klaassens, E.S., Booijink, C.C.G.M., Kleerebezem, M., Smidt, H. and de Vos, W.M., 2006. Isolation of DNA from bacterial samples of the human gastrointestinal tract. Nature Protocols 1: 870-873. https://doi.org/10.1038/nprot.2006.142