Correspondence Jin-Woo Bae baejw@khu.ac.kr

# Blautia faecis sp. nov., isolated from human faeces

Seong-Kyu Park, Min-Soo Kim and Jin-Woo Bae

Department of Life and Nanopharmaceutical Sciences and Department of Biology, Kyung Hee University, Seoul 130-701, Republic of Korea

A strictly anoxic, Gram-stain-positive, non-motile *Blautia*-like bacterium, designated strain M25<sup>T</sup>, was isolated from a human faecal sample. Strain M25<sup>T</sup> was negative for both catalase and oxidase activity, utilized carbohydrates as fermentable substrates, produced lactate and acetate as the major end products of glucose fermentation in PYG medium, and had a DNA G+C content of 41.6 mol%. Comparative 16S rRNA gene sequencing showed that strain M25<sup>T</sup> was closely related to *Ruminococcus obeum* ATCC 29174<sup>T</sup> (96.40% 16S rRNA gene sequence similarity) and *Blautia glucerasea* HFTH-1<sup>T</sup> (96.17%) within the family *Lachnospiraceae*. Straight-chain saturated and monounsaturated cellular fatty acids were also detected, the majority being C<sub>14:0</sub>, C<sub>16:0</sub> and C<sub>16:0</sub> dimethyl acetal acids. Based on the phenotypic, genotypic and phylogenetic characteristics presented in this study, strain M25<sup>T</sup> represents a novel species within the genus *Blautia* for which the name *Blautia faecis* sp. nov. is proposed. The type strain is M25<sup>T</sup> (=KCTC 5980<sup>T</sup>=JCM 17205<sup>T</sup>).

In taxonomic studies using polyphasic approaches, the genus Blautia was created to encompass a number of misclassified organisms including one species of the genus Clostridium, five species of the genus Ruminococcus, and a number of previously misclassified ruminococci within the Firmicutes (Liu et al., 2008). At the time of writing, the genus consisted of nine species with validly published names: Blautia coccoides, Blautia glucerasea, Blautia hansenii, Blautia hydrogenophotrophica, Blautia luti, Blautia producta, Blautia schinkii, Blautia stercoris and Blautia wexlerae (Furuya et al., 2010; Liu et al., 2008; Park et al., 2012). Members of the genus Blautia are Grampositive, non-motile, coccoid or oval-shaped obligate anaerobes that produce acetate, ethanol, hydrogen, lactate and succinate as the end products of glucose fermentation (Liu et al., 2008). The present study examined the microbial diversity of the human gastrointestinal tract and resulted in the isolation of a Blautia-like strain from human faeces. On the basis of the distinct morphological, biochemical, genomic and phylogenetic characteristics, a novel species belonging to the genus Blautia is described.

Strain M25<sup>T</sup> was isolated from faecal samples obtained from a healthy Korean male (28 years old). The samples were immediately placed in an anaerobic chamber (Bactron II; Shellab) containing 90 % nitrogen, 5 % hydrogen and 5 % carbon dioxide. After serial dilution in 0.1 M PBS (pH 7.0) and plating on Gifu anaerobic medium agar (GAM, pH 6.2; Nissui Pharmaceutical), the samples were incubated for 3 days at 37 °C. A pure culture of one isolate was obtained after several rounds of streaking onto fresh medium. The isolate was preserved in 10 % skimmed milk (BBL) and stored at -80 °C.

Morphological and physiological studies were performed using cells grown on GAM medium at 37 °C for 3 days. Cell morphology was observed under a light microscope (ECLIPSE 50*i*; Nikon). Gram staining was performed using a Gram-staining kit (bioMérieux) according to the manufacturer's instructions. Malachite green dye was used to determine the presence of spore formation, and motility was examined by stabbing cells into GAM medium containing 0.4 % agar. Oxidase activity was assessed with 1 % (v/v) *p*-tetramethyl phenylenediamine (bioMérieux), and catalase activity was determined by the observation of bubbles following addition of 3 % hydrogen peroxide.

Colony PCR using a PCR Pre-Mix (SolGent) and two bacterial universal primers (8F and 1492R) was performed to obtain the 16S rRNA gene sequence (Baker et al., 2003) of strain M25<sup>T</sup>. Amplified 16S rRNA gene sequences were sequenced using a BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) and analysed using a PRISM 3730XL DNA analyser (Applied Biosystems). After sequence assembly using the SeqMan software (DNASTAR), the 16S rRNA gene sequence of strain M25<sup>T</sup> was compared with other sequences on the EzTaxon-e server (http://eztaxon-e.ezbiocloud.net/; Kim et al., 2012). The results showed clear similarities with the 16S rRNA gene sequences of B. glucerasea HFTH-1<sup>T</sup> (96.17%), B. schinkii BT (95.57%), B. luti  $BInIX^T$  (95.54%), B. wexlerae WAL  $14507^{T}$  (95.15%), *B. producta* ATCC 27340<sup>T</sup> (94.77%), *B. coccoides* ATCC 29236<sup>T</sup> (94.70%), GAM6-1<sup>T</sup> (94.56%) and B. hansenii DSM 20583<sup>T</sup> (93.83%). A high level of 16 S

Abbreviations: DMA, dimethyl acetal; FAME, fatty acid methyl ester.

The GenBank accession number for the 16S rRNA gene sequence of strain  $M25^{T}$  is HM626178.

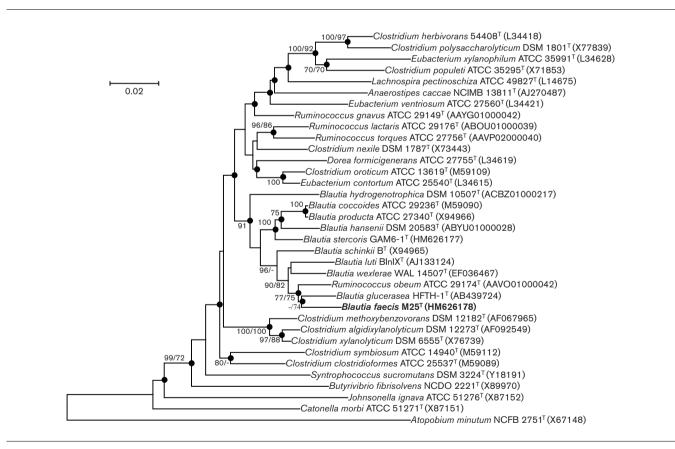
rRNA gene sequence similarity was also observed with Ruminococcus obeum ATCC 29174<sup>T</sup> (96.40%). R. obeum ATCC 29174<sup>T</sup> is considered by Liu *et al.* (2008) to be a member of the genus Blautia, however as it has not been possible to deposit this species in a second culture collection in a second country, it cannot be transferred to the genus Blautia because the new combination would not be considered to be validly published according to the rules of the Bacteriological Code. The multiple sequence alignment program, CLUSTAL X (Thompson et al., 1997), was used to align the 16S rRNA gene sequences from strain M25<sup>T</sup> and its close relatives within the family Lachnospiraceae. The neighbour-joining (Saitou & Nei, 1987) and maximum-likelihood methods (Cavalli-Sforza & Edwards, 1967; Felsenstein, 1981) were also used, based upon a consensus phylogenetic tree constructed using MEGA software version 5 (Tamura et al., 2011). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 bootstrap replicates for the neighbour-joining method and 500 replicates for the maximum-likelihood method) is shown at the nodes. The consensus phylogenetic tree (Fig. 1) showed that strain

M25<sup>T</sup> clearly belongs to the genus *Blautia* within the family *Lachnospiraceae*.

The closely related reference strains *B. stercoris* GAM6-1<sup>T</sup>, *B. coccoides* DSM 935<sup>T</sup>, *B. producta* DSM 2950<sup>T</sup>, *B. hansenii* DSM 20586<sup>T</sup>, *R. obeum* ATCC 29174<sup>T</sup> (ATCC) and *B. schinkii* DSM 10518<sup>T</sup> (DSMZ) were used to obtain comparative data. All bacterial strains, including strain M25<sup>T</sup>, were cultivated and maintained in GAM medium (pH 6.2) for at least 2 days at 37 °C unless stated otherwise.

API ZYM and Rapid ID 32A strips (bioMérieux) were used to determine enzyme activities and biochemical characteristics, and an API 50CHB/E strip (bioMérieux) was used to identify acid production from various carbohydrates. Table 1 shows characteristics of strain M25<sup>T</sup> in comparison with those of the other reference strains.

The cellular fatty acid composition of strain  $M25^{T}$  was compared with that of *B. stercoris* GAM6-1<sup>T</sup>, *B. coccoides* DSM 935<sup>T</sup>, *B. hansenii* DSM 20583<sup>T</sup>, *R. obeum* ATCC 29174<sup>T</sup>, *B. hansenii* DSM 20586<sup>T</sup> and *B. glucerasea* DSM 22028<sup>T</sup>. Briefly, cells were grown for 3 days on peptone-yeast-extract



**Fig. 1.** A consensus phylogenetic tree constructed using 16S rRNA gene sequences from strain M25<sup>T</sup> and close relatives within the family *Lachnospiraceae*. Filled circles indicate generic branches that were present in both phylogenetic trees generated by the neighbour-joining and maximum-parsimony methods. Numbers at nodes indicate the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 bootstrap replicates for the neighbour-joining method and 500 replicates for the maximum-likelihood method); only values >70 % are shown. *Atopobium minutum* NCFB  $2751^{T}$  (GenBank accession no. X67148) was included as an outgroup. Bar, 2 substitutions per 100 nt positions.

### **Table 1.** Characteristics that differentiate strain M25<sup>T</sup> from type strains of closely related species

Strain: 1,  $M25^{T}$ ; 2, *B. stercoris* GAM6-1<sup>T</sup>; 3, *B. coccoides* DSM 935<sup>T</sup>; 4, *B. producta* DSM 2950<sup>T</sup>; 5, *B. hansenii* DSM 20583<sup>T</sup>; 6, *R. obeum* ATCC 29174<sup>T</sup>; and 7, *B. schinkii* DSM 10518<sup>T</sup>. All data were obtained in this study. +, Positive; -, negative; (+), weakly positive; A, acetic acid; L, lactic acid; S, succinic acid.

Characteristic	1	2	3	4	5	6	7
End products of glucose fermentation	L, A	А	A, S	L, A	L, A	А	А
Enzyme activity (API ZYM)							
Alkaline phosphatase	_	_	—	-	_	+	-
Esterase (C4)	_	_	+	-	_	_	-
Leucine arylamidase	+	-	_	—	_	-	-
Acid phosphatase	+	-	_	+	_	+	+
Naphthol-AS-BI-phosphohydrolase	_	-	+	+	_	+	+
α-Galactosidase	+	+	+	+	-	+	+
$\beta$ -Galactosidase	+	+	+	+	+	-	-
α-Glucosidase	+	+	+	+	-	+	+
β-Glucosidase	+	+	+	+	-	-	+
N-Acetyl-β-glucosaminidase	_	-	_	—	+	-	-
α-Fucosidase	+	_	—	-	_	_	-
Enzyme activity (Rapid ID 32A)							
Arginine dihydrolase	+	+	—	—	—	_	+
α-Galactosidase	+	+	+	+	+	+	(+)
$\beta$ -Galactosidase	+	+	+	+	+	+	(+)
$\beta$ -Galactosidase 6-phosphate	(+)	_	_	_	_	_	_
β-Glucosidase	+	+	(+)	+	_	(+)	+
α-Arabinosidase	+	+	+	+	_	_	+
N-Acetyl-β-glucosaminidase	_	_	_	_	+	_	+
Mannose fermentation	_	_	+	+	—	_	_
Raffinose fermentation	_	_	+	+	—	_	_
Glutamic acid decarboxylase	_	_	_	+	+	_	_
α-Fucosidase	+	_	_	-	_	_	(+)
Alkaline phosphatase	(+)	_	_	_	_	_	_
Leucine arylamidase	+	_	—	—	—	_	_
Glycine arylamidase	+	_	—	—	—	_	_
Serine arylamidase	+	_	_	_	_	_	_
Fermentation of (API 50CHB/E):							
Glycerol	_	+	_	-	_	_	_
Erythritol	_	+	—	—	—	_	_
D-Arabinose	+	+	_	_	_	+	+
L-Arabinose	_	+	—	—	—	_	_
Methyl $\beta$ -D-xyloside	_	_	_	_	_	+	_
Methyl α-D-glucoside	_	_	_	-	—	+	-
N-Acetylglucosamine	_	_	_	-	_	+	_
Arbutin	_	_	_	_	_	+	_
Salicin	_	_	_	_	_	+	_
Glycogen	_	_	_	+	_	_	_
Turanose	_	+	_	+	+	+	_
l-Fucose	+	+	_	-	_	+	+
L-Arabitol	_	_	_	+	_	_	_

glucose (PYG) agar (Liu *et al.*, 2008) and the cellular fatty acids were extracted as described by the Sherlock Microbial Identification System (MIDI, 1999). The extracts were then analysed using GC (6890; Hewlett Packard) and the Microbial Identification software package, Sherlock Version 4.0 with the BHIBLA 3.80 library. The results showed that  $C_{14:0}$  (20.00%),  $C_{16:0}$  (22.60%) and  $C_{16:0}$  dimethyl acetal (DMA) acids (9.18%) were the predominant components of

straight-chain saturated and monounsaturated cellular fatty acids in strain  $M25^{T}$ , which was a similar profile to previously described members of the the genus *Blautia* (Liu *et al.*, 2008). The detailed fatty acid profiles of strain  $M25^{T}$  and the other reference species is shown in Table 2.

To analyse the end products of glucose fermentation, cells were cultivated at 37 °C for 5 days in PYG (Liu *et al.*,

**Table 2.** Fatty acid profiles of strain M25<sup>T</sup> compared with type strains of closely related species

Strain: 1, M25<sup>T</sup>; 2, *B. stercoris* GAM6-1<sup>T</sup>; 3, *B. coccoides* DSM 935<sup>T</sup>; 4, *B. hansenii* DSM 20583<sup>T</sup>; 5, *B. schinkii* DSM 10518<sup>T</sup>; 6, *R. obeum* ATCC 29174<sup>T</sup>; 7, *B. glucerasea* DSM 22028<sup>T</sup>. All data are from this study. Values represent the percentage of total fatty acids. ND, Not detected.

Fatty acid	1	2	3	4	5	6	7
C <sub>12:0</sub> FAME	ND	1.17	1.75	1.99	10.55	0.61	20.93
C14:0 FAME	20.00	24.10	18.31	15.01	12.75	8.33	6.56
C14:0 DMA	7.11	3.65	4.55	0.55	0.66	2.62	2.33
C16:0 ALDE	2.35	2.64	3.41	0.61	10.19	3.81	8.11
C15:0 FAME	0.66	0.34	0.55	0.45	0.24	0.25	ND
C <sub>16:1</sub> cis7 FAME	ND	0.17	ND	1.24	0.43	0.13	ND
C <sub>16:1</sub> cis9 FAME	1.21	1.54	7.86	2.50	1.51	1.92	ND
C <sub>16:1</sub> cis11 FAME	0.30	0.26	0.25	ND	0.17	ND	ND
C16:0 FAME	22.60	19.09	17.37	32.26	10.91	9.17	7.35
C <sub>16:1</sub> cis9 DMA	5.71	1.65	3.51	0.54	0.73	7.76	ND
C <sub>16:0</sub> DMA	9.18	14.35	16.54	3.32	20.37	10.69	19.56
C <sub>18:0</sub> ALDE	0.45	0.40	0.65	ND	3.61	1.40	1.47
C <sub>18:2</sub> <i>cis</i> 9, 12	0.37	0.43	ND	0.23	0.20	ND	0.53
FAME							
C <sub>18:1</sub> cis9 FAME	4.48	9.34	3.91	16.25	2.82	6.49	5.78
C18:0 FAME	1.19	1.76	0.94	8.08	0.84	1.53	0.58
C <sub>18:1</sub> cis9 DMA	7.93	8.17	5.35	5.36	5.89	9.22	11.47
C <sub>18:1</sub> cis11 DMA	4.10	3.37	4.62	0.96	3.67	13.23	0.90
C <sub>18:0</sub> DMA	0.63	0.82	1.21	0.58	4.25	2.45	2.75
Summed							
features*							
1	1.37	0.88	1.07	ND	0.17	0.60	0.87
4	1.23	0.40	0.79	ND	0.41	2.21	ND
6	0.60	0.28	ND	ND	0.41	0.47	ND
7	2.36	1.50	1.10	1.40	2.90	3.59	5.14
8	1.22	0.62	1.02	ND	1.82	5.35	0.37
10	2.91	2.38	4.15	7.33	2.56	6.17	0.45

\*Summed features represent two or three fatty acids that cannot be separated by GLC using the MIDI system. Summed feature 1 comprised  $C_{13:1}cis12$  fatty acid methyl ester (FAME) and/or  $C_{14:0}$ alde; summed feature 4 comprised unknown 14.762 and/or  $C_{15:2}$ FAME; summed feature 6 comprised  $C_{15:0}$  anteiso 3-OH FAME and/ or  $C_{16:1}cis7$  DMA; summed feature 7 comprised  $C_{17:2}$  FAME and/or  $C_{17:1}cis8$  FAME; summed feature 8 comprised  $C_{17:1}cis9$  FAME and/ or  $C_{17:2}$  FAME; summed feature 10 comprised  $C_{18:1}cis11/trans9/$ *trans*6 FAME and/or unknown 17.834.

2008) and pretreated as described by Guerrant *et al.* (1982). A Thermo Surveyor HPLC (TSP-0299) system fitted with a Thermo PDA detector (Thermo Scientific) and an Alltech Prevail Organic Acid Column (Grace Davison Discovery Sciences) were used to identify the end products of glucose fermentation. In general, members of the genus *Blautia* produce acetate, ethanol, hydrogen, lactate or succinate as the end products of glucose fermentation (Liu *et al.*, 2008). Consistent with other members of the genus *Blautia*, strain M25<sup>T</sup> produced

lactate and acetate as the major end products of glucose fermentation.

For analysis of DNA G+C content, genomic DNA was extracted from strain M25<sup>T</sup> using an UltraClean Microbial DNA Isolation kit (Mo Bio Laboratories) and the G+C content of the DNA was analysed fluorometrically using SYBR dye and real-time PCR (Gonzalez & Saiz-Jimenez, 2002). All experiments were performed in triplicate. The data were calibrated against genomic DNA from three reference strains: *Escherichia coli* K-12 (Gonzalez & Saiz-Jimenez, 2002), *Bifidobacterium adolescentis* ATCC 15703<sup>T</sup> and *R. obeum* ATCC 29174<sup>T</sup> (for which complete genome analyses were performed). Strain M25<sup>T</sup> had a genomic DNA G+C content of 41.6 mol%, which is within the range 37–47 mol% reported for other members of the genus *Blautia* (Furuya *et al.*, 2010; Liu *et al.*, 2008).

Recognised species of the genus *Blautia* are Gram-positive coccoid-shaped bacteria. They do not have oxidase and catalase activities and do not produce indole (Liu *et al.*, 2008; Park *et al.*, 2012). Strain  $M25^{T}$  had similar phenotypic characteristics to other species of the genus *Blautia*, but there were enzymological and metabolic differences among members of the genus *Blautia* (Table 1). Consequently, the phenotypic, biochemical, genotypic and phylogenetic data suggest that strain  $M25^{T}$  is different from other established members of the genus *Blautia*, and therefore, strain  $M25^{T}$  represents a novel species of the genus *Blautia*, with the name *Blautia faecis* sp. nov.

#### Description of Blautia faecis sp. nov.

*Blautia faecis* (fa'e.cis. L. gen. n. *faecis* of faeces, from which the organism was isolated).

Cells are Gram-stain-positive, strictly anaerobic, nonmotile coccobacilli that are  $1.0-2.3 \times 0.5-0.8 \ \mu m$  in size. After 72 h at 37 °C on GAM agar under N<sub>2</sub>/H<sub>2</sub>/CO<sub>2</sub> (90:5:5, by vol.) gas phase, cells are non-spore-forming, circular, umbonate, pale grey opaque colonies with intact edges. Catalase and oxidase activity test results are negative. In Rapid ID 32A system tests there are positive results for arginine dihydrolase,  $\alpha$ -galactosidase,  $\beta$ -galactosidase,  $\alpha$ glucosidase,  $\beta$ -glucosidase,  $\alpha$ -arabinosidase,  $\alpha$ -fucosidase, leucine arylamidase, glycine arylamidase and serine arylamidase activities, and weakly positive results for  $\beta$ galactosidase-6-phosphate and alkaline phosphatase, but negative results for urease,  $\beta$ -glucuronidase, N-acetyl- $\beta$ glucosaminidase, mannose fermentation, raffinose fermentation, glutamic acid decarboxylase, reduction of nitrates, indole production, arginine arylamidase, proline arylamidase, leucyl glycine arylamidase, phenylalanine arylamidase, pyroglutamic acid arylamidase, tyrosine arylamidase, alanine arylamidase, histidine arylamidase and glutamyl glutamic arylamidase activities. The API ZYM system shows positive reactions for leucine arylamidase, acid phosphatase,  $\alpha$ -galactosidase,  $\beta$ -galactosidase,  $\alpha$ glucosidase,  $\beta$ -glucosidase, and  $\alpha$ -fucosidase activities, but negative reactions for alkaline phosphatase, esterase (C4),

esterase lipase (C8), lipase (C14) valine arylamidase, cysteine arylamidase, trypsin, α-chymotrypsin, naphthol-AS-BI-phosphohydrolase,  $\beta$ -glucuronidase, N-acetyl- $\beta$ glucosaminidase and  $\alpha$ -mannosidase activities. In the API 50CHB/E system there were positive reactions for Darabinose, ribose, aesculin, D-tagatose and 5-keto-gluconate, but negative reactions for glycerol, erythritol, Larabinose, D-xylose, L-xylose, adonitol, methyl  $\beta$ -D-xyloside, galactose, fructose, mannose, sorbose, rhamnose, dulcitol, inositol, mannitol, sorbitol, methyl a-D-mannoside, methyl α-D-glucoside, N-acetylglucosamine, amygdalin, arbutin, salicin, cellobiose, maltose, lactose, melibiose, sucrose, trehalose, inulin, melezitose, raffinose, starch, glycogen, xylitol, gentiobiose, turanose, D-lyxose, D-fucose, L-fucose, D-arabitol, L-arabitol, gluconate and 2-ketogluconate. Straight-chain saturated and monounsaturated cellular fatty acids are present and are predominantly C14:0, C16:0 and C16:0 DMA acids. The end products of glucose fermentation are mainly lactic acid and acetic acid.

The type strain,  $M25^{T}$  (=KCTC 5980<sup>T</sup>=JCM 17205<sup>T</sup>) was isolated from human faces. The G+C content of the genomic DNA of the type strain is 41.6 mol%.

# Acknowledgements

This work was supported by grants from the National Institute of Biological Resources (NIBR) and the Mid-career Researcher Program (2011-0028854) through the National Research Foundation of Korea (NRF).

## References

Baker, G. C., Smith, J. J. & Cowan, D. A. (2003). Review and reanalysis of domain-specific 16S primers. *J Microbiol Methods* 55, 541– 555.

Cavalli-Sforza, L. L. & Edwards, A. W. (1967). Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet* 19, 233–257.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17, 368–376.

**Furuya, H., Ide, Y., Hamamoto, M., Asanuma, N. & Hino, T. (2010).** Isolation of a novel bacterium, *Blautia glucerasei* sp. nov., hydrolyzing plant glucosylceramide to ceramide. *Arch Microbiol* **192**, 365–372.

**Gonzalez, J. M. & Saiz-Jimenez, C. (2002).** A fluorimetric method for the estimation of G + C mol% content in microorganisms by thermal denaturation temperature. *Environ Microbiol* **4**, 770–773.

**Guerrant, G. O., Lambert, M. A. & Moss, C. W. (1982).** Analysis of short-chain acids from anaerobic bacteria by high-performance liquid chromatography. *J Clin Microbiol* **16**, 355–360.

Kim, O. S., Cho, Y. J., Lee, K., Yoon, S. H., Kim, M., Na, H., Park, S. C., Jeon, Y. S., Lee, J. H. & other authors (2012). Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int J Syst Evol Microbiol* **62**, 716–721.

Liu, C., Finegold, S. M., Song, Y. & Lawson, P. A. (2008). Reclassification of *Clostridium coccoides*, *Ruminococcus hansenii*, *Ruminococcus hydrogenotrophicus*, *Ruminococcus luti*, *Ruminococcus productus* and *Ruminococcus schinkii* as *Blautia coccoides* gen. nov., comb. nov., *Blautia hansenii* comb. nov., *Blautia hydrogenotrophica* comb. nov., *Blautia luti* comb. nov., *Blautia producta* comb. nov., *Blautia schinkii* comb. nov. and description of *Blautia wexlerae* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* 58, 1896– 1902.

MIDI (1999). Sherlock Microbial Identification System Operating Manual, version 3.0. Newark, DE: MIDI.

Park, S. K., Kim, M. S., Roh, S. W. & Bae, J. W. (2012). Blautia stercoris sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* **62**, 776–779.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406–425.

Suau, A., Bonnet, R., Sutren, M., Godon, J. J., Gibson, G. R., Collins, M. D. & Doré, J. (1999). Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl Environ Microbiol* **65**, 4799–4807.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28, 2731–2739.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25, 4876–4882.