INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

RESEARCH ARTICLE

Teng *et al.*, *Int. J. Syst. Evol. Microbiol.* 2023;73:005950
DOI 10.1099/ijsem.0.005950

MICROBIOLOGY SOCIETY

OPEN ACCESS

Allocoprobacillus halotolerans gen. nov., sp. nov and *Coprobacter tertius* sp. nov., isolated from human gut microbiota

Nancy M. Y. Teng[1,2], Raymond Kiu[1], Rhiannon Evans[1], David J. Baker[1], Christian Zenner[3], Stephen D. Robinson[1,4,*] and Lindsay J. Hall[1,2,3,*]

Abstract

Two novel bacterial isolates were cultured from faecal samples of patients attending the Breast Care clinic at the Norwich and Norfolk University Hospital. Strain LH1062^T was isolated from a 58-year-old female diagnosed with invasive adenocarcinoma with ductal carcinoma *in situ*. Strain LH1063^T was isolated from a healthy 51-year-old female. Isolate LH1062^T was predicted to be a potential novel genus most closely related to *Coprobacillus*, whilst LH1063^T was predicted to be a novel species belonging to *Coprobacter*. Both strains were characterized by polyphasic approaches including 16S rRNA gene analysis, core-genome analysis, average nucleotide identity (ANI) comparisons and phenotypic analysis. Initial screening of the 16S rRNA gene of LH1062^T returned a nucleotide identity of 93.4% to *Longibaculum muris*. For LH1063^T, nucleotide identity was a 92.6% to *Coprobacter secundus*. Further investigations showed that LH1062^T had a genome size of 2.9 Mb and G+C content of 31.3 mol%. LH1063^T had a genome size of 3.3Mb and G+C content of 39.2 mol%. Digital DNA-DNA hybridization (dDDH) and ANI values of LH1062^T with its closest relative, *Coprobacillus cateniformis* JCM 10604^T, were 20.9 and 79.54%, respectively. For LH1063^T, the dDDH and ANI values with its closest relative, *Coprobacter secundus* 177^T, were 19.3 and 77.81%, respectively. Phenotypic testing confirmed that LH1062^T could not be matched to a known validly published isolate in any database; thereby indicating a novel genus for which the name *Allocoprobacillus* gen. nov. is now proposed with LH1062^T (=DSM 114537^T=NCTC 14686^T) being the type strain of the proposed novel species *Allocoprobacillus halotolerans* sp. nov. Strain LH1063^T (=DSM 114538^T=NCTC 14698^T) fits within the genus *Coprobacter* and, it being the third species within this genus, the name *Coprobacter tertius* sp. nov. is proposed.

INTRODUCTION

The metagenomic era has allowed researchers to delve into the microbiota diversity of the human gut and associate certain taxa with disease status. However, in order to understand underlying mechanisms and develop new microbiota-based therapies, pure and well-characterized isolates are required. Certain taxa can be problematic to culture due to their fastidious nature, including their acute sensitivity to oxygen, which has limited progress in this area. Thus, we sought to apply a culturing approach to our observational trial. In the Breast hEalth And Microbiota (BEAM) study, we isolated two bacterial strains that did not have a match to the Type Strain Genome Server (TYGS) database. Further genomic investigations suggested the novelty of these isolates. Herein we describe two novel bacterial strains: *Allocoprobacillus halotolerans* LH1062^T and *Coprobacter tertius* LH1063^T.

Author affiliations: [1]Gut Microbes & Health, Quadram Institute Bioscience, Norwich Research Park, Norwich, UK; [2]Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, UK; [3]Intestinal Microbiome, School of Life Sciences, ZIEL—Institute for Food & Health, Technical University of Munich, Freising, Germany; [4]School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK.
***Correspondence:** Lindsay J. Hall, Lindsay.hall@quadram.ac.uk; Stephen D. Robinson, Stephen.robinson@quadram.ac.uk
Keywords: faecal culturing; human gut microbiota.
Abbreviations: ANI, average nucleotide identity; BEAM, breast health and microbiota; BHI, brain heart infusion; dDDH, digital DNA-DNA hybridization; MIDI, microbial identification system; PoCP, percentage of conserved proteins; YCFA, yeast-casitone-fatty acid.
The GenBank/EMBL/DDBJ accession number for the 16S rRNA gene sequence of strain LH1062T is ON553235 and for LH1063T is ON553234; for the draft genome sequences (genome assemblies) of strains LH1062T and LH1063T they are GCA_024399475.1 and GCA_024330105.1, respectively. Strain LH1062T and LH1063T have been deposited at DSMZ (accession numbers: 114537 and 114538 respectively) and NCTC (accession numbers: 14686 and 14698 respectively).

RESULTS

Isolation and ecology

Faecal samples were donated by breast cancer patients as part of the BEAM study in partnership with the Norwich and Norfolk University Hospital. This study gained favourable ethical approval by the Faculty of Medicine and Health Ethics board at the University of East Anglia (201819-092HT). Patients who were aged between 30–60 years old, with first time diagnosis of invasive breast cancer and did not have any antibiotics 3 months prior to consenting were eligible to partake. LH1062^T was isolated from a 58-year-old patient diagnosed with adenocarcinoma with ductal carcinoma *in situ*. The tumour was stage 1C with no presence of invasion into lymph nodes. Isolate LH1063^T was isolated from a healthy 51-year-old female. The protocol for faecal collection was laid out by the Norwich Research Park (NRP) Biorepository (Norwich, UK), and was in accordance with the terms of the Human Tissue Act 2004 (HTA) and approved with license number 11208 by the HTA.

Approximately 1 g faecal sample was transferred into preservation medium (20% glycerol in sterile PBS) and stored at −80 °C until further use; 100 mg was taken from the glycerol frozen aliquot and homogenized in sterile reduced PBS. Consequently, a serial dilution was prepared and 200 µl spread on a 14 cm agar plate with yeast–casitone–fatty acid (YCFA) medium supplemented with carbohydrates (glucose, maltose and cellobiose) and brain heart infusion (BHI) medium [1]. The plates were incubated anaerobically at 37 °C in an atmosphere containing N_2, CO_2, H_2 (85, 5 and 10%, respectively) for 72 h before colonies were picked and purified by re-streaking at least three times with 48 h growth periods in between. A pure liquid culture was prepared for long-term storage using 20% glycerol solution.

Genomic characterization

A. halotolerans LH1062^T was grown in BHI media and *C. tertius* LH1063^T in YCFA, both for 48 h. Genomic DNA was extracted using the FastDNA Spin Kit for Soil (MP Biomedicals, 116560200), following the manufacturer's protocol with an amendment of 3 min bead-beating procedure as described in [2]. The genome for LH1062^T, *Allocoprobacillus halotolerans*, was sequenced using the Nanopore MinION sequencing platform. Sequencing for LH1062^T was performed following Oxford Nanopore's native barcoding genomic DNA protocol (SQK-LSK109). The sequence reads were initially filtered through Filtlong version 0.2.1 [3], with only the top 90% quality of reads remained for subsequent genome assembly. Consequently, the genome was assembled using Flye version 2.9 [4], resulting in one contig of 2.93Mbp and a G+C content of 31.26 mol%. Sequencing for LH1063^T was carried out on an Illumina NextSeq500 instrument; libraries were prepared using a novel modified Illumina DNA prep tagmentation approach [5]. The genome was quality-filtered using Fastp version 0.20.0 with option -p 20 [6]. *De novo* assembly was performed using Spades version 3.11 [7]. BactspeciesID version 1.2 was used to check for contamination and provided a preliminary identity [8]. Using the output of BactspeciesID, if available, the type strain was downloaded and FastANI version 1.33 used to compare the query whole genome to the type strain whole genome to confirm identity [9]. For isolates that did not provide a BactspeciesID output, the 16S rRNA gene sequence was extracted *in silico* using BactspeciesID and run through BLASTn [10] and type strain whole genomes downloaded to use for final confirmation. The two query genomes were screened using GTDB (release R207) [11], not resulting in any matches. The assembled draft genome for LH1063^T had 25 contigs, a genome size of 3.3Mbp and a G+C content of 39.23 mol%. Using Protologger [12], we identified LH1062^T as representing a novel genus whilst LH1063^T as representing a novel species [12]. The genomes were also run on the TYGS [13] suggesting two potential novel species for both isolates.

Allocoprobacillus halotolerans LH1062^T

The full-length 16S rRNA gene sequences (1.5 Kb) of 35 species representing 35 genera within the family Erysipelotrichaceae were downloaded from the List of Prokaryotic names with Standing in Nomenlature (LSPN: June 2022) [14]. The 16S rRNA gene sequence for *Coprobacillus cateniformis* was also included after Protologger suggested it was the closest relative based on ANI results. The 16S rRNA gene sequences were aligned using MUSCLE version 3.8.31 [15] prior to the reconstruction of a maximum-likelihood phylogenetic tree using IQ-TREE version 2.0.5 [16] with the TEST model at 1000 bootstrap replications and subsequent visualization using iTOL version 6 [17]. LH1062^T was placed next to *Massiliomicrobiota timonensis* SN16 (Fig. 1a), but according to LSPN this has yet to be validated as an official new genus and species. The 16S rRNA gene sequence similarity between LH1062^T and *M. timonensis* SN16 was 96.49%. According to the Fig. 1a, *Intestinibaculum porci* KCTC 15725^T seems to be the closest relative to LH1062^T; however, the 16S rRNA percentage identity with *I. porci* KCTC 15725^T is only 89.32%, whereas that with *L. muris* DSM 29487^T is 93.04%. We also compared the 16S rRNA gene sequence of LH1062^T with that of *C. cateniformis* JCM 10604^T, as suggested by the Protologger result, which had a nucleotide identity of 90.80%. We reconstructed a phylogenomic tree using PhyloPhlan version 3.0.51 after downloading the genomes of the species used in Fig. 1a. No whole genome sequences could be found for *Breznakia pachnodae* Pei061^T and *Absiella argi* N6H1-5^T. The configuration file specified the use of DIAMOND version 0.9.19 and MAFFT version 7.515 as the aligner. Sequences were trimmed using TRIMAL version 2.4.rev15 and the tree reconstructed using IQ-TREE version 2.1.4. The tree was reconstructed with the PhyloPhlAn options –diversity medium and –accurate. Fig. 1d shows the genomic tree. As suggested by PhyloPhlAn, LH1062^T is placed amongst the family Coprobacillaceae and is closely related to *L. muris* DSM 29487^T and *C. cateniformis* JCM 10604^T. However, based on the 16S rRNA gene sequences, it has a higher identity percentage to *L. muris* DSM 29487^T than *C. cateniformis* JCM 10604^T. Further genomic investigation between LH1062^T and *L. muris* DSMZ 29487^T indicated dDDH was estimated at 21.7% (TYGS), whilst

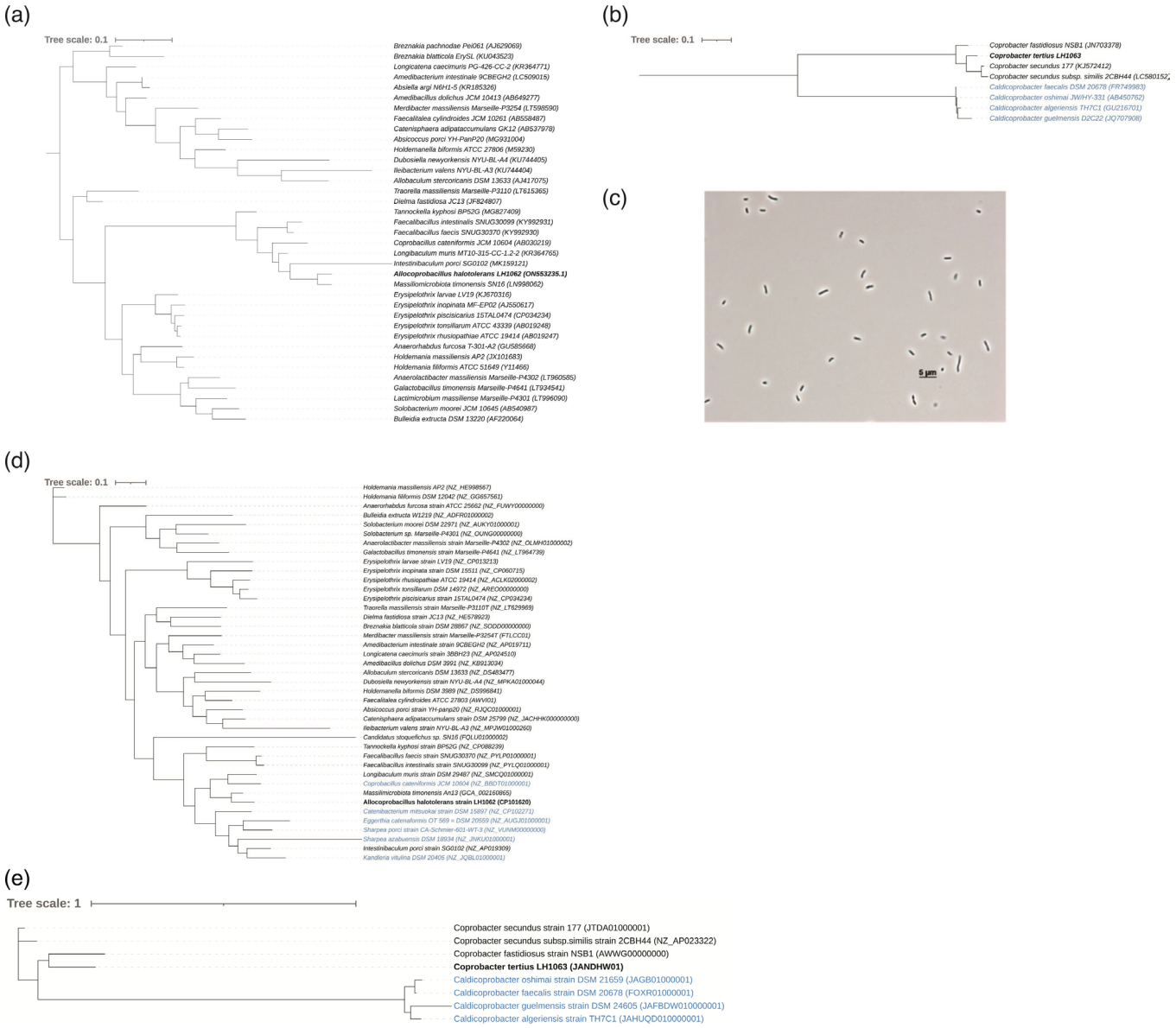

Fig. 1. (a) A mid-point rooted maximum-likelihood phylogenetic tree of *Allocoprobacillus halotolerans* LH1062^T relative to 16S rRNA gene sequences of the genera in the family Erysipelotrichaceae. (b) Mid-point rooted maximum-likelihood phylogenetic tree of Coprobacter *tertius* LH1063^T, relative to the 16S rRNA gene sequences of the species in the genus *Coprobacter*. The outgroups are members of the genus *Caldicoprobacter*, belonging to the family Caldicoprobacteraeceae. (c) Phase contract microscopy image of LH1063^T. (d) Genome tree reconstructed using PhyloPhlAn for LH1062^T. In blue are outgroup genomes, which are the species in the genus *Coprobacillus*, family Coprobacillaceae. (e) Genome tree reconstructed using PhyloPhlAn for LH1063^T. In blue are outgroup genomes, which are the species belonging to the genus *Caldicoprobacter*, family Caldicoprobacteraceae.

the ANI is 79.3% (fastANI v1.3) [9, 13]. The dDDH comparison between LH1062^T and *C. cateniformis* JCM 10604^T is 21% and ANI 78.8%. The highest ANI (79.3%) and dDDH (21.7%) values were significantly below the intra-species thresholds of 95 and 70% for ANI and dDDH, respectively. We used EzAAI version 1.2.1 [18] to calculate average amino acid identity (AAI) using the genomes shown in Fig. 1a. The highest percentage was matched to *C. cateniformis* JCM 10604^T at 73.25%. This was followed by *L. muris* DSM 29487^T at 71.22% and *I. porci* SG0102^T (KCTC15725^T) at 62.29%. AAI with *M. timonensis* SN16 was only 50.65%. Using Protologger [12], the percentage of conserved proteins (PoCP) analysis assigned LH1062^T to *Clostridium* with a value of 50.08%, which is borderline to be suggestive of a novel genus. However, using BLASTn and limiting the search to the 'Bacillus/Clostridium group', the highest 16S rRNA gene sequence from the genus *Clostridium* was 89%, which is even lower than to *L. muris* DSM 29487^T and to *C. cateniformis* JCM 10604^T. Taken together and given the inconsistencies within the genus *Clostridium*, a novel genus *Allocoprobacillus* is proposed, with *Allocoprobacillus halotolerans* sp. nov. as the type species and LH1062^T representing the type strain.

Phenotypic investigations were carried out by DSMZ Services, Leibniz-Institute DSMZ–Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany. This involved: cell and colony morphology, salt and temperature tolerance, fermentation profiles of different carbohydrates, catalase activity, oxidase activity, and fatty acid analysis. Despite the strain being isolated in BHI, it was found to grow better in PY+X medium (DSMZ: 104b), which was subsequently used for the phenotypic testing. LH1062^T cells were found to grow in rods and in chains and were Gram-positive. They were negative for catalase and oxidase activity but positive for haemolytic activity. The bacterium grew in a relatively broad range of salt conditions (1–20%), with growth delayed between 7–20%. It failed to grow at temperatures below 25 °C, but grew normally up to a maximum of 40 °C; however, weak growth was observed at 45 °C. Optimum growth was observed between 30–40 °C. Biochemical characteristics were observed using API50CHB strips. Weak activity for D-arabinose, L-arabinose, D-xylose, L-xylose, fructose, mannose, sucrose, turanose and D-lyxose and positive activity for glucose, sorbose, aesculin, gentiobiose, D-tagatose and 5-ketogluconate was observed. The isolate was also incubated in the Gen III Biolog MicroPlate using Medium A. The inoculum was grown to a transmission of 93% turbidity before anaerobically incubating with the substrates for 48 h at 37 °C; after which, the plate was read using Biolog's Microbial Identification Systems software. The bacterium was positive for gentiobiose, D-fructose, D-fucose, L-fucose, L-rhamnose, D-serine, D-fructose-6-PO4, minocycline, L-galactonic acid lactone, D-glucuronic acid, glucuronamide and sodium butyrate (Table 1). We noted inconsistencies between the API strips and the Gen III Biolog MicroPlate results, which may be due to the fact that the API50CHB strip was incubated for 12 days at 37 °C aerobically (covered in paraffin), whereas the Biolog MicroPlate assay was performed anaerobically. Comparing the reaction patterns with the not validly published isolate *M. timonensis* SN16, there is a distinct difference in the ability of LH1062^T to react with these substrates. Although the closest relative *L. muris* DSM 29487^T was negative for the acidification of carbohydrates, LH1062^T was not (Table 1). *C. cateniformis* JCM 10604^T was found to have acidification of glucose, mannose, galactose, fructose, sucrose, maltose, cellobiose, lactose and trehalose, following a similar profile to LH1062^T. Cellular fatty acids were detected after converting them into fatty acid methyl esters (FAMEs) following a modified protocol [19]. The FAME mixture was separated by gas chromatography and detected by a flame ionization detector using Sherlock Microbial Identification System (MIDI) based on the TSBA6 database. $C_{16:0}$ was the most abundant fatty acid for LH1062^T at 19.08%. This was also the major fatty acid for the not validly published isolate *M. timonensis* SN16, at 41% and for *L. muris* DSM 29487^T at 30.1%.

Based on the genomic and phenotypic results presented above, we propose LH1062^T as the type strain of a new genus *Allocoprobacillus* gen. nov., naming it like its closest genomic relative based on 16S rRNA gene sequence results, i.e. *Coprobacillus*. Strain LH1062^T is suggested reperesent a novel species named *Allocoprobacillus halotolerans* sp. nov.

Coprobacter tertius LH1063^T

For *Coprobacter tertius* LH1063^T, the 16S rRNA gene sequences representative of six *Coprobacter* species and two *Coprobacter secundus* subspecies type strains were downloaded from LSPN (LSPN: June 2022) [14]. The maximum-likelihood phylogenetic tree was generated as aforementioned for *A. halotolerans* LH1062^T. Strain LH1063^T was placed next to *Coprobacter fastidiosus* and *Coprobacter secundus* (Fig. 1b), based on the 16S rRNA gene sequences. From the phylogenomic tree (Fig. 1e), strain LH1063^T is more closely related to *C. fastidiosus* NSB1^T than *C. secundus* species. Based on 16S rRNA gene sequences comparison (Protologger), the closest relative was *C. secundus* with a nucleotide identity of 91.5%. *C. secundus* was also the closest match based on ANI at 77.81% (Table 2). We note that the ANI values for *C. secundus* 177^T reported by Protologger and OrthoANI [20] are different, being 77.81 and 72.8%, respectively. This discrepancy is explained due to Protologger using fastANI while OrthoANI uses USearch. OrthoANI was used as opposed to fastANI as fastANI would not provide an output if the ANI<80%, which it was for each species in the genus *Coprobacter*. The dDDH values between LH1063^T and *C. fastidiosus* DSMZ 26242^T, *C. secundus* 177^T and *C. secundus* subsp. *similis* 2CBH44^T were 20.1, 19.4 and 19.3% respectively. LH1603^T had a genome size of 3.3 Mbp and a G+C content of 39.23 mol%, whilst the genome size and G+C content for *C. secundus* 177^T are 4.1 Mbp and 37.8 mol%, respectively.

Phenotypic investigations were also carried out by DSMZ Services, Leibniz-Institute DSMZ–Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany. This involved: cell morphology, salt, bile and temperature tolerance, fermentation profiles of different carbohydrates and fatty acid analysis. *C. tertius* LH1063^T cells grow as rods in pairs measuring roughly 3 µm long (Fig. 1c). The strain was Gram-stain-negative and negative for catalase, oxidase and haemolytic activity. The bacterium tolerated up to 3% salinity and was shown to grow well between 30–40 °C, with weak growth at 25 °C. Unlike *C. fastidiosus* and *C. secundus*, LH1063^T failed to grow in any concentration of ox gall [21, 22]. Biochemical characteristics were observed using API20A strips and the inoculation was grown anaerobically at 37 °C for 24 h before the test was performed. The strain produced acid from glucose, lactose, maltose, mannose, raffinose and trehalose. The strain could hydrolyse gelatin and aesculin, and was weakly positive for acid production from mannitol, sucrose and melezitose. In APIrID32A assays, the strain was positive for α-galactosidase, β-galactosidase, α-glucosidase, *N*-acetyl-β-glucosaminidase, raffinose and glutamic acid decarboxylase fermentation, alkaline phosphatase, arginine phosphatase, leucyl-glycine arylamidase, phenylalanine arylamidase, leucine arylamidase, tyrosine arylamidase, alanine arylamidase, glycine arylamidase, histidine arylamidase, glutamyl-glutamic acid arylamidase, and serine arylamidase. The bacterium also had a weak reaction for mannose fermentation. Strain LH1063^T was additionally run on the Gen III Biolog MicroPlate with the same conditions described previously for strain LH1062^T. It was positive

Table 1. Comparison of analytical profile indexes of strain LH1062^T, *Massiliomicobiota timonensis* SN16 and *Coprobacillus cateniformis* JCM 10604^T

Two tests were used, API50CHB and Biolog GenIII Microplate, to test for a broader range of substrates for *A. halotolerans* LH1062^T. Blank cells indicate that the substrate was not present in the test. The *M. timonensis* SN16 profile was determined using previously published literature [23]. +, Positive; –, negative; –/+, borderline/weak.

	API50CHB for LH1062^T	*M. timonensis* SN16	Biolog III microplate for LH1062^T	C.cateniformis JCM 10604
Methyl α-D-glucoside	–	–		
Methyl α-D-mannoside	–	+		
2-Ketogluconate	–	–		
5-Ketogluconate	+	+		
Adonitol	–	+		
Amygdalin	–	+		–
Arbutin	–	+		
Cellobiose	–	+		+
D-Arabinose	–/+	–		–
D-Arabitol	–	–	–	
D-Fucose	–	+	+	
D-Lyxose	–/+	+		
D-Tagatose	+	+		
Turanose	–/+	+	–/+	
D-Xylose	–/+	–		–
Dulcitol	–	+		
Erythritol	–	+		–
Aesculin	+	+		–
Fructose	–/+	+	+	+
Galactose	–	+	–/+	+
Gentibiose	+	+	+	
Gluconate	–	+		
Glucose	+	+	–/+	+
Glycerol	–	–	–	
Glycogen	–	+		–
Inositol	–	–	–	–
Inulin	–	+		
L-Arabinose	–/+	–		
L-Arabitol	–			
L-Fucose	–	+	+	
L-Xylose	–/+	–		–
Lactose	–	+	–	+
Maltose	–	+	–	+
Mannitol	–	–	–	–
Mannose	–/+	+	–	+

Continued

Table 1. Continued

	API50CHB for LH1062^T	*M. timonensis* SN16	Biolog III microplate for LH1062^T	C.cateniformis JCM 10604
Melibiose	−	+	−/+	
Melizitose	+	+		−
N-Acetylglucosamine	−	−	−	
Raffinose	−	+	−	−
Rhamnose	−	−	+	−
Ribose	+	−		−
Salicin	−	+	−	+
Sorbitol	−	−	−	−
Sorbose	+	−		
Methyl β-D-xyloside	−			
Starch	−	+		−
Sucrose	−/+	+	−/+	
Trehalose	−	+	−	+
Xylitol	−	+		

after 48 h incubation for gentiobiose, melibiose, α-D-glucose, D-mannose, D-fructose, D-galactose, 3-methyl glucose, D-fucose, L-fucose, L-rhamnose, D-glucose 6-phosphate, D-fructose 6-phosphate, minocycline, D-galacturonic acid and D-glucuronic acid. The acidification of the various carbohydrates including α-glucose, D-mannose and D-fructose to name a few was also confirmed using the Gen III Biolog MicroPlate. The biochemical profile of strain LH1063^T is quite similar to that of *C. fastidiosus* NSB1^T and slightly different from *C. secundus* 177^T (Table 3). Cellular fatty acids were detected using the same method for LH1062^T. The major fatty acid produced by LH1603^T was anteiso-$C_{15:0}$ at just 25%, followed closely by iso-$C_{15:0}$ at 20%. This was similar to *C. fastidiosus* NSB1^T and *C. secundus* 177^T. For *C. fastidiosus* NSB1^T it was 23–27% and 26–27% and for *C. secundus* 177^T it was 0.24–0.34% and 0.59–0.70% for anteiso-$C_{15:0}$ and iso-$C_{15:0}$, respectively.

Based on the genomic and phenotypic results presented above, we propose LH1063^T as representing a novel species within the genus *Coprobacter*. We propose the epithet *tertius*, as this is the third species of this genus.

DESCRIPTION OF *ALLOCOPROBACILLUS* GEN. NOV.

Allocoprobacillus (Al.lo.co.pro.ba.cil.lus. Gr. masc. adj. *allos*, another, other, different; Gr. fem. n. *kopros*, excrement, ordure, faeces; L. masc. n. *bacillus*, a small rod; N.L. masc. n. *Allocoprobacillus*, another small rod isolates from faeces). The genus is placed into the

Table 2. Digital DNA–DNA hybridization (dDDH) and average nucleotide identity (ANI) percentages between species in *Coprobacter* compared to Coprobacter *tertrius* LH1063^T

dDDH was determined using TYGS [13]. ANI was determined using the EzBioCloud ANI calculator [20].

Species	dDDH	ANI
Coprobacter fastidiosus DSM 26242^T	20.1	71.9
Coprobacter secundus 177^T	19.4	72.8
Coprobacter secundus subsp. *similis* 2CBH44^T	19.3	72.0
Caldicoprobacter oshimai DSM 21659^T	18.7	63.3
Caldicoprobacter guelmensis DSM 24605^T	18.6	63.3
Caldicoprobacter faecalis DSM 20678^T	18.5	61.8
Caldicoprobacter algeriensis DSM 22661^T	18.2	60.6

Table 3. Comparison of analytical profile indexes of strain LH1063^T and other members of the genus *Coprobacter*

Three tests were used: API 20A, API rID 32A and Biolog Gen III MicroPlate. The profiles of *C. fastidiosus* and *C. secundus* were based on previous published literature [21, 22]. +, Positive; –, negative; –/+, borderline/weak.

API 20A	*C. tertius* LH1063^T	*C. fastidiosus* NSB1^T	*C. secundus* 177^T	API rID 32A/Biolog GenIII MicroPlate	*C. tertius* LH1063^T	*C. fastidiosus* NSB1^T	*C. secundus* 177^T
Acid from arabinose	–	–	–	α-Arabinosidase	–	–	–
Acid from cellobiose	–	–	+	α-Fucosidase	–	–	+
Acid from glucose	+	+	+	α-Galactosidase	+	+	+
Acid from glycerol	–	–	–	α-Glucosidase	+	+	+
Acid from lactose	+	+	+	Alanine arylamidase	+	+	+
Acid from maltose	+	+	+	Alkaline phosphatase	+	+	+
Acid from mannitol	–/+	–	–	Arginine arylamidase	+	–	–
Acid from mannose	+	+	+	Arginine dihydrolase	–	–	–
Acid from melezitose	–/+	–	–	β-Galactosidase	+	+	+
Acid from raffinose	+	+	+	β-Galactosidase 6-phosphate	–	–	–
Acid from rhamnose	–	–	+	β-Glucosidase	–	–	+
Acid from salicin	–	–	-/+	β-Glucuronidase	–	–	+
Acid from sorbitol	–	–	–	Glutamic acid decarboxylase	+	+	–
Acid from sucrose	–/+	–	+	Glutamyl-glutamic acid arylamidase	+	–	–
Acid from trehalose	+	–	+	Glycine arylamidase	+	–	–
Acid from xylose	–	–	–	Histidine arylamidase	+	–	–
Aesculin hydrolysis	+	–	–	Indole production	–	–	–
Gelatin hydrolysis	+	+	–	Leucine arylamidase	+	–	–
Indole production	–	–	–	Leucyl-glycine arylamidase	+	+	+
Urease	–	–	–	Mannose fermentation	–/+	+	–/+
				N-Acetyl-β-Glucosaminidase	+	+	+
				Nitrate reduction	–	–	–
				Phenylalanine arylamidase	+	–	–
				Proline arylamidase	–	–	–
				Pyroglutamic acid arylamidase	–	–	–
				Raffinose fermentation	+	+	+
				Serine arylamidase	+	–	–
				Tyrosine arylamidase	+	–	–
				Urease	–	–	–

family *Coprobacillaceae* (phylum Bacillota) based GTDB-Tk comparison, but placed into the family Erysipelotrichaceae (phylum Firmicutes) based on 16S rRNA gene analysis. The ANI value of the type strain LH1062^T with *Coprobacillus cateniformis* JCM 10604^T, which was suggested to be the closest relative based on GTDB-Tk comparison, was 78.8%. The closest relative based on 16S rRNA gene analysis is *Longibaculum muris* DSM 29487^T with a nucleotide identity of 91.84%. The higher nucleotide identity to *L. muris* suggests that LH1062^T belongs to the family *Erysipelotrichaceae*. PoCP was 50.08% with *Clostridium*, furthermore suggesting that this could be novel. Based on phenotypic characterisation, *L. muris* DSM29487^T was negative for carbohydrate acidification while LH1062^T was not. LH1062^T had more in common with the metabolic profile of *C. cateniformis* JCM 10604^T. A

novel genus, *Allocoprobacillus*, is proposed within the currently validly named family *Erysipelotrichaceae* to accommodate isolate LH1062^T, with the type species being *Allocoprobacillus halotolerans*.

DESCRIPTION OF *ALLOCOPROBACILLUS HALOTOLERANS* SP. NOV.

Allocoprobacillus halotolerans [ha.lo.to'le.rans. Gr. masc. n. *hals* (gen. *halos*), salt; L. pres. part. *tolerans*, tolerating, enduring; N.L. part. adj. *halotolerans*, salt-tolerating].

Description is based on a single strain. Cells are Gram-positive, facultative anaerobic, haemolytic, rod-shaped and grows in chains. The bacterium grows well in a temperature range of 30–40 °C, with weak growth at 45 °C. It tolerates a range of NaCl concentrations (1–20%), with delayed growth between 7–20% in PY-X medium. Colonies on BHI medium after 48 h are circular, smooth, shiny with entire margins roughly 0.3–0.4 mm in diameter, which are positive for ribose, glucose, sorbose, aesculin, melizitose, gentibiose, D-tagatose and 5-ketogluconate. The major fatty acid produced is $C_{16:0}$. The type strain, LH1062^T (DSM 114537^T=NCTC 14686^T), was isolated from a donated faecal sample from a 58-year-old breast cancer patient. The genome size is 2.92Mbp with a G+C content of 31.26 mol%.

DESCRIPTION OF *COPROBACTER TERTIUS* SP. NOV.

Coprobacter tertius (ter.ti.us. L. masc. adj. *tertius* third, referring to the fact that this is the third species to be described within the genus *Coprobacter*).

Description is based on a single strain. Cells are Gram-negative, facultative anaerobic, non-haemolytic and absent for catalase and oxidase. LH1063^T cells are rod shaped and grow in pairs. The bacterium tolerates NaCl concentrations between 1–3% and a temperature range of 30–40 °C with weak growth observed at 25 °C when grown in YCFA. Colonies on YCFA medium after 48 h are circular, shiny, convex with no clear margins roughly 0.1 mm in diameter and does not tolerate any concentration of ox bile. The strain is positive for gelatine and aesculin hydrolysis, α-galactosidase, β-galactosidase, α-glucosidase, *N*-acetyl-β-glucosaminidase, glutamic acid decarboxylase, alkaline phosphatase, arginine arylamidase, leucyl-glycine arylamidase, phenylalanine arylamidase, leucine arylamidase, tyrosine arylamidase, alanine arylamidase, histidine arylamidase, glutamyl-glutamic acid arylamidase and serine arylamidase. The strain also produces acid from glucose, lactose, maltose, mannose, raffinose and trehalose. The major fatty acid produced is anteiso-$C_{15:0}$. The type strain, LH1063^T (DSM 114538^T=NCTC 14698^T), was isolated from a faecal sample from healthy 51-year-old female. The genome size is 3.3Mbp with a G+C content of 39.23 mol%.

Funding information
This work was funded by BigC grant 17-16R to L.J.H. and S.D.R. (studentship for N.M.Y.T.). L.J.H. and S.D.R. are also supported by the Biotechnology and Biological Sciences Research Council (BBSRC), Institute Strategic Programme Gut Microbes and Health BB/R012490/1, and its constituent projects BBS/E/F/000PR10353 and BBS/E/F/000PR10356. L.J.H. is also supported by Wellcome Trust Investigator Awards 100974/C/13/Z and 220876/Z/20/Z.

Acknowledgements
This research was supported in part by the Norwich Bioscience Institutes (NBI) Computing infrastructure for Science (CiS) group through the provision of a High-Performance Computing (HPC) Cluster. We would like to thank Prof. Mark Pallen and Dr. Rachel Gilroy, at QIB, for expert advice in naming these novel strains.

Author contributions
Conceptualization, N.M.Y.T., L.J.H. and S.D.R; methodology, N.M.Y.T, R.E. and D.J.B; software, N.M.Y.T., R.K.; validation, N.M.Y.T., R.K., C.Z. and L.J.H; formal analysis, N.M.Y.T. and R.K.; investigation, N.M.Y.T. and R.K.; resources, N.M.Y.T, R.K., R.E and D.J.B; data curation, N.M.Y.T. and R.K.; writing – original draft preparation, N.M.Y.T. and R.K.; writing – reviewing and editing, N.M.Y.T, R.K., L.J.H., C.Z., R.E. and S.D.R; visualisation, N.M.Y.T. and R.K; supervision, L.J.H. and S.D.R.; project administration, N.M.Y.T.; funding acquisition, L.J.H. and S.D.R.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
This study gained favourable ethical approval by the Faculty of Medicine and Health Ethics board at the University of East Anglia (FMH 201819-092HT). The patient provided signed informed consent to participate in this study. The protocol for faecal collection was laid out by the Norwich Research Park (NRP) Biorepository (Norwich, UK), and was in accordance with the terms of the Human Tissue Act 2004 (HTA) and approved with license number 11 208 by the HTA.

Consent to publish
Mandatory if personal details of an individual that may lead to their identification has been included in the article. Details include direct identifiers such as names, images and videos; or indirect identifiers that when used together may reveal the individual's identity (e.g., gender, age, location of treatment, rare disease, socioeconomic data). You will need to upload evidence of written consent for the publication of these details to the peer review system and you must include a sentence stating that this consent was obtained in the manuscript. For articles describing individuals under the age of 18, consent for publication must be obtained from their parent or legal guardian. If the person has died, consent must be obtained from their next of kin. You can use our consent form to obtain consent for publication, or a consent form from your own institution or region if appropriate.

References

1. Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, *et al*. Culturing of "Unculturable" human Microbiota reveals novel Taxa and extensive Sporulation. *Nature* 2016;533:543–546.
2. Alcon-Giner C, Dalby MJ, Caim S, Ketskemety J, Shaw A, *et al*. Microbiota supplementation with *Bifidobacterium* and *Lactobacillus* modifies the Preterm infant gut Microbiota and Metabolome: an observational study. *Cell Rep Med* 2020;1:100077.
3. Wick R, Menzel P. Filtlong: a tool for filtering long reads by quality. https://github.com/rrwick/Filtlong 2021.
4. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
5. Baker DJ, Aydin A, Le-Viet T, Kay GL, Rudder S, *et al*. CoronaHiT: high-throughput sequencing of SARS-Cov-2 Genomes. *Genome Med* 2021;13:21.
6. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–i890.
7. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
8. Kiu R. BACTspeciesID: identify microbial species and genome contamination using 16S rRNA gene approach; 2020. https://github.com/raymondkiu/bactspeciesID
9. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9:5114.
10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
11. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, *et al*. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 2022;50:D785–D794.
12. Hitch TCA, Riedel T, Oren A, Overmann J, Lawley TD, *et al*. Automated analysis of genomic sequences facilitates high-throughput and comprehensive description of bacteria. *ISME Commun* 2021;1:16.
13. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res* 2022;50:D801–D807.
14. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* 2020;70:5607–5612.
15. Edgar RC. MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. *bioRxiv* 2021.
16. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, *et al*. Corrigendum to: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 2020;37:2461.
17. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.
18. Kim D, Park S, Chun J. Introducing EzAAI: a pipeline for high throughput calculations of prokaryotic average amino acid identity. *J Microbiol* 2021;59:476–480.
19. Miller LT. Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxy acids. *J Clin Microbiol* 1982;16:584–586.
20. Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek* 2017;110:1281–1286.
21. Shkoporov AN, Khokhlova EV, Chaplin AV, Kafarskaia LI, Nikolin AA, *et al*. *Coprobacter fastidiosus* gen. nov., sp. nov., a novel member of the family *Porphyromonadaceae* isolated from infant faeces. *Int J Syst Evol Microbiol* 2013;63:4181–4188.
22. Shkoporov AN, Chaplin AV, Khokhlova EV, Shcherbakova VA, Motuzova OV, *et al*. *Alistipes inops* sp. nov. and *Coprobacter secundus* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* 2015;65:4580–4588.
23. Ndongo S, Khelaifia S, Fournier P-E, Raoult D. "*Massiliomicrobiota timonensis*," a new bacterial species isolated from the human gut. *New Microbes New Infect* 2016;13:25–26.