Article

Novel bile acid biosynthetic pathways are enriched in the microbiome of centenarians

https://doi.org/10.1038/s41586-021-03832-5

Received: 24 November 2020

Accepted: 16 July 2021

Published online: 29 July 2021

Check for updates

Yuko Sato[1,2,3,20], Koji Atarashi[1,2,3,20], Damian R. Plichta[4,20], Yasumichi Arai[5], Satoshi Sasajima[1,3], Sean M. Kearney[1,2], Wataru Suda[2], Kozue Takeshita[1,3], Takahiro Sasaki[6], Shoki Okamoto[1], Ashwin N. Skelly[1], Yuki Okamura[1], Hera Vlamakis[4], Youxian Li[2], Takeshi Tanoue[1,2,3], Hajime Takei[7], Hiroshi Nittono[7], Seiko Narushima[1,2], Junichiro Irie[8], Hiroshi Itoh[8], Kyoji Moriya[9], Yuki Sugiura[10], Makoto Suematsu[10], Nobuko Moritoki[11], Shinsuke Shibata[11], Dan R. Littman[12,13], Michael A. Fischbach[14], Yoshifumi Uwamino[15], Takashi Inoue[16], Akira Honda[17], Masahira Hattori[2,18], Tsuyoshi Murai[6], Ramnik J. Xavier[4,19,20] ✉, Nobuyoshi Hirose[5] & Kenya Honda[1,2,3] ✉

Centenarians have a decreased susceptibility to ageing-associated illnesses, chronic inflammation and infectious diseases[1–3]. Here we show that centenarians have a distinct gut microbiome that is enriched in microorganisms that are capable of generating unique secondary bile acids, including various isoforms of lithocholic acid (LCA): iso-, 3-oxo-, allo-, 3-oxoallo- and isoallolithocholic acid. Among these bile acids, the biosynthetic pathway for isoalloLCA had not been described previously. By screening 68 bacterial isolates from the faecal microbiota of a centenarian, we identified Odoribacteraceae strains as effective producers of isoalloLCA both in vitro and in vivo. Furthermore, we found that the enzymes 5α-reductase (5AR) and 3β-hydroxysteroid dehydrogenase (3β-HSDH) were responsible for the production of isoalloLCA. IsoalloLCA exerted potent antimicrobial effects against Gram-positive (but not Gram-negative) multidrug-resistant pathogens, including *Clostridioides difficile* and *Enterococcus faecium*. These findings suggest that the metabolism of specific bile acids may be involved in reducing the risk of infection with pathobionts, thereby potentially contributing to the maintenance of intestinal homeostasis.

The microbiome has long been recognized as a key player in determining the health status of ageing individuals through its role in controlling digestive functions, bone density, neuronal activity, immunity and resistance to pathogen infection[3–7]. Microbial data from older individuals often show increased interindividual variability and reduced diversity, and are thus being linked to immunosenescence, chronic systemic inflammation and frailty. By contrast, centenarians (individuals aged 100 years and older) are less susceptible to age-related chronic diseases and have survived several bouts of infectious diseases[1–3]. It has been postulated that there are centenarian-specific members of the gut microbiota that, rather than representing a mere consequence of ageing, could actively contribute to resistance against pathogenic infection and other environmental stressors[3–7]. In this study, we aimed to identify such beneficial bacteria in the gut microbiota of centenarians.

Microbiome signature of centenarians

We recruited a cohort consisting of three age groups: centenarian (average age, 107 years old; $n = 160$), older (85–89 years old; $n = 112$) and young (21–55 years old, $n = 47$) (all Japanese individuals). All centenarians were recruited as part of the Japan Semi-supercentenarian Study[1], most of whom lived in nursing homes (85.0%) with the remainder living at home (9.4%) or in hospitals (5.6%) (Supplementary Table 1). Centenarians generally reported reduced activities of daily living and mini-mental state examination scores, along with reduced red-blood-cell counts and serum albumin (Extended Data Fig. 1a–c and Supplementary Table 1). Consistent with the paradigm that ageing is accompanied by chronic inflammation secondary to decreased barrier integrity and immunosenescence[4], a subset of centenarians showed signs of low-grade

[1]Department of Microbiology and Immunology, Keio University School of Medicine, Tokyo, Japan. [2]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [3]JSR-Keio University Medical and Chemical Innovation Center, Tokyo, Japan. [4]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Centre for Supercentenarian Medical Research, Keio University School of Medicine, Tokyo, Japan. [6]School of Pharmaceutical Sciences, Health Sciences University of Hokkaido, Hokkaido, Japan. [7]Junshin Clinic Bile Acid Institute, Tokyo, Japan. [8]Department of Internal Medicine, Division of Endocrinology, Metabolism and Nephrology, Keio University School of Medicine, Tokyo, Japan. [9]Department of Infection Control and Prevention, The University of Tokyo, Tokyo, Japan. [10]Department of Biochemistry, Keio University School of Medicine, Tokyo, Japan. [11]Electron Microscope Laboratory, Keio University School of Medicine, Tokyo, Japan. [12]The Kimmel Center for Biology and Medicine of the Skirball Institute, New York University School of Medicine, New York, NY, USA. [13]Howard Hughes Medical Institute, New York, NY, USA. [14]Department of Bioengineering, Stanford University, Stanford, CA, USA. [15]Department of Laboratory Medicine, Keio University School of Medicine, Tokyo, Japan. [16]Department of Marmoset Biology and Medicine, Central Institute for Experimental Animals, Kawasaki, Japan. [17]Division of Gastroenterology and Hepatology, Tokyo Medical University Ibaraki Medical Center, Ibaraki, Japan. [18]Graduate School of Advanced Science and Engineering, Waseda University, Tokyo, Japan. [19]Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA, USA. [20]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA. [21]These authors contributed equally: Yuko Sato, Koji Atarashi, Damian R. Plichta. ✉e-mail: xavier@molbio.mgh.harvard.edu; kenya@keio.jp

inflammation as evidenced by elevated serum C-reactive protein and faecal lipocalin (Extended Data Fig. 1c, d). Nevertheless, the majority of centenarians were free of chronic diseases such as obesity, diabetes, hypertension and cancer, and the prevalence of these diseases was not significantly increased compared to the group of older individuals (Extended Data Fig. 1e, f and Supplementary Table 1). We collected faecal samples from the three groups of participants and characterized the microbiome (excluding three centenarians and one older participant undergoing antibiotic treatment). Principal coordinate analysis revealed significant differences in the composition of the microbiome between centenarians and both control groups (Extended Data Fig. 2a). At the phylum level, we observed a significant enrichment of Proteobacteria and Synergistetes, a moderate enrichment of Verrucomicrobia and a depletion of Actinobacteria in centenarians compared to both control groups (Extended Data Fig. 2b, c), which is in partial agreement with previous centenarian studies, including that of a Sardinian cohort[3]. Such expansions of Proteobacteria are a frequent finding in patients with inflammatory bowel disease (IBD); however, in contrast to the reduced microbial α-diversity commonly observed in patients with IBD, centenarians had, on average, a higher Shannon index compared to young control participants (Extended Data Fig. 2d).

Several taxa showed differential relative abundances in centenarians versus the control groups (Extended Data Fig. 2e–g), which we categorized into three signatures based on the trajectory with age: (1) the first signature included taxa, the abundance of which was increased or decreased with age; (2) the second signature included taxa, the abundance of which was similar in centenarians and young control participants, but distinct from the older participants; and (3) the third signature included centenarian-specific taxa, the abundance of which was significantly different between centenarians and both the older and young control groups, but not between these two control groups. Notably, *Alistipes*, *Parabacteroides*, *Bacteroides*, *Clostridium* and *Methanobrevibacter* species were included in the third signature and were specifically enriched in centenarians (Extended Data Fig. 2g). One of the most-enriched species in centenarians was *Clostridium scindens*, which is known to possess the relatively rare 7α-dehydroxylation capacity needed to convert primary into secondary bile acids[8,9]. By contrast, key butyrate producers such as *Faecalibacterium prausnitzii* and *Eubacterium rectale* were selectively depleted in centenarians (Extended Data Fig. 2g). Some of these observations are in agreement with the Sardinian study, in which centenarians exhibited a decreased relative abundance of *F. prausnitzii* and *E. rectale* and an increase in *Methanobrevibacter smithii*[3].

We also analysed stool samples from the lineal descendants and siblings of centenarians by 16S ribosomal RNA (rRNA) sequencing (n = 22 relatives of 14 centenarians; 48–95 years old) (Extended Data Fig. 3). Some bacterial species—such as *Alistipes putredinis* and *Odoribacter splanchnicus*—were more abundant in centenarians and their family members compared to the other groups (Extended Data Fig. 3d). Enrichment of these taxa may be due to host genetics, lifestyle and diet.

Centenarians have a unique bile acid profile

We next assessed faecal metabolites. The levels of short-chain fatty acids (SCFAs) such as propionic and butyric acid were decreased, whereas branched SCFAs such as isobutyric and isovaleric acid, as well as ammonium, were elevated in centenarians (Extended Data Fig. 4a, b and Supplementary Table 2). These metabolic alterations may be attributable to the simultaneous depletion of SCFA producers and enrichment of protein-fermenting organisms, such as *A. putredinis*[10] (Extended Data Fig. 2g). The increase in amino-acid-using bacteria is probably a consequence of reduced upper intestinal proteolytic capacity. Moreover, faecal pH was significantly higher in centenarians than both groups of control participants (Extended Data Fig. 4c), which may be due in part to the lower SCFA concentrations and reduced gastric juice production that is characteristic of ageing.

Metagenomic analysis identified an increase in the relative abundance of bile-acid-inducible (*bai*) operon genes in centenarians, although this trend was not apparent in the Sardinian cohort[3] (Extended Data Fig. 2h, i). Therefore, we next focused on the distribution of bile acids in faeces. Although the total bile acid load was not significantly different between groups, centenarians showed a unique distribution with lower levels of primary bile acids and increased levels of chenodeoxycholic acid (CDCA) metabolites (Fig. 1a–d, Extended Data Fig. 5 and Supplementary Table 3). In particular, the levels of isoLCA, 3-oxoLCA, alloLCA, 3-oxoalloLCA and isoalloLCA were significantly elevated in centenarians, whereas they were comparably low in the older and young control groups (Fig. 1a, d and Extended Data Fig. 5e). Furthermore, faecal pH was positively associated with the concentrations of these bile acids (Extended Data Fig. 4d), potentially implying that such an intestinal milieu may promote the expansion of certain bacterial species and/or the expression and activity of enzymes involved in the production of these bile acids.

Identification of isoalloLCA-producing bacterial strains

We set out to identify the bacterial strains and enzymes responsible for the biosynthesis of bile acids unique to centenarians. 3-Oxodeoxycholic acid (3-OxoDCA) can be generated from 3-oxo-Δ^4-DCA through the action of 5β-reductase (5BR), which is encoded by the *baiCD* gene in a small number of commensals including *C. scindens* and *Clostridium hylemonae*[8,9]. It has additionally been reported that *Eggerthella lenta* and *Ruminococcus gnavus* can generate 3-oxoDCA from DCA, and isoDCA from 3-oxoDCA, by the actions of 3α-hydroxysteroid dehydrogenase (3α-HSDH) and 3β-HSDH, respectively[11]. Thus, we hypothesized that 3-oxoLCA and isoLCA are produced in a similar manner, through the actions of 5BR, 3α-HSDH and 3β-HSDH (Fig. 1e and Extended Data Fig. 6a). On the other hand, the biosynthetic pathway leading to 3-oxoallo-, allo- and isoallo-LCA generation had not been determined previously. We predicted that 3-oxoalloLCA might be generated from 3-oxo-Δ^4-LCA by a 5α-reductase (5AR) homologue (Fig. 1e), analogous to the 5AR-mediated conversion of testosterone into 5α-dihydrotestosterone[12] (Extended Data Fig. 6b). We also predicted that the subsequent transformation of 3-oxoalloLCA to alloLCA or isoalloLCA may use 3α-HSDH or 3β-HSDH, respectively (Fig. 1e and Extended Data Fig. 6a), mirroring the previously characterized conversion of 3-oxoDCA to DCA or isoDCA[11] (Extended Data Fig. 6c).

To validate our pathway predictions and identify unique bile-acid-producing bacterial strains, we followed up on a supercentenarian (CE91, who is more than 110 years old) who showed high faecal levels of these bile acids (Fig. 1a). We cultured faecal samples from CE91 and isolated 68 unique strains, which roughly recapitulated the microbiota structure of participant CE91 (Fig. 2a and Supplementary Table 4). We then incubated individual isolates with CDCA, LCA or 3-oxo-Δ^4-LCA as starting substrates. Whereas the incubation with CDCA did not result in the production of target bile acids in any of the cultures, incubation with LCA or 3-oxo-Δ^4-LCA afforded 12 out of 68 strains the capacity to generate 3-oxoLCA and 8 to generate isoLCA, suggesting that these strains possess 3α-HSDH, 3β-HSDH and/or 5BR (Fig. 2b and Extended Data Fig. 7). Notably, after incubation with 3-oxo-Δ^4-LCA, a marked accumulation of isoalloLCA was observed in the cultures of *Parabacteroides merdae* strain number (St)3, *Odoribacter laneus* St19, Odoribacteraceae spp. St21–St24 and, to a lesser degree, *Bacteroides dorei* St6, St7 (Fig. 2b and Extended Data Fig. 7c), suggesting that these strains contain both 5AR and 3β-HSDH. Additionally, *Parabacteroides goldsteinii* St1, St2, *Bacteroides thetaiotaomicron* St9, *Bacteroides uniformis* St10–St13, *Alistipes finegoldii* St15, St16, *Alistipes onderdonkii* St17, St18, and *O. laneus* St20 cultures all showed the substantial accumulation of 3-oxoalloLCA, but little to no isoalloLCA (Fig. 2b), which is

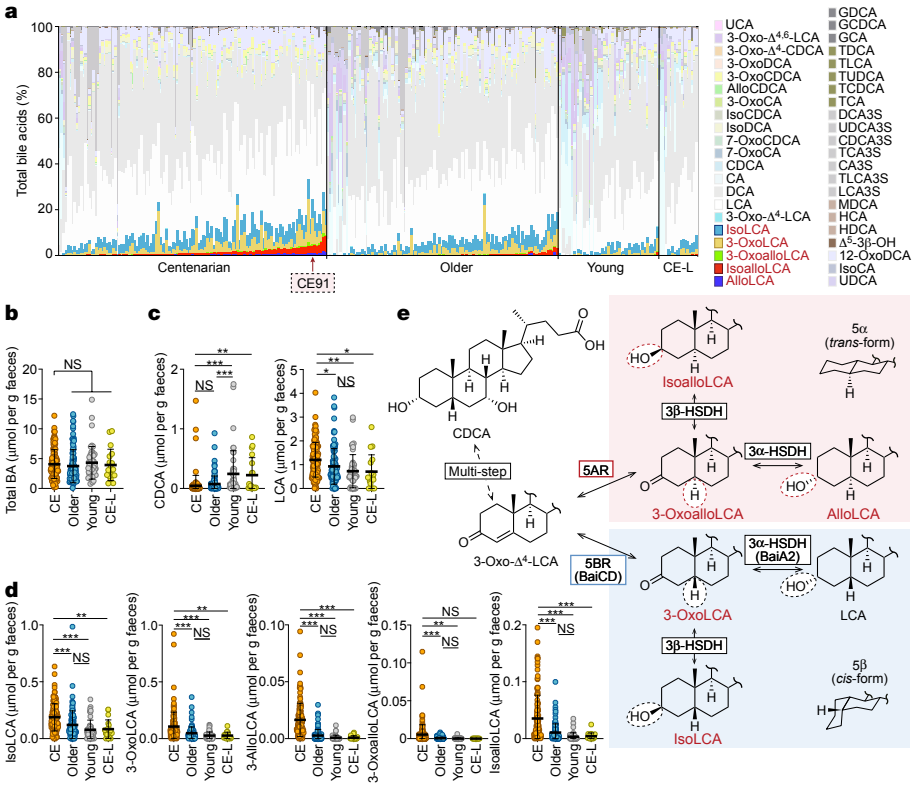


Fig. 1 | Centenarians have significantly elevated levels of faecal isoLCA, 3-oxoLCA, alloLCA, 3-oxoalloLCA and isoalloLCA. a–**d**, Faecal bile acid concentrations of centenarians (CE, $n = 125$), older participants ($n = 107$), young participants ($n = 47$) and lineal relatives of centenarians (CE-L, $n = 18$) were quantified by liquid chromatograpy–tandem mass spectrometry (LC–MS/MS). **a**, Ratios of each bile acid were calculated by dividing its concentration by the sum of the 43 bile acids measured within a sample. CE91 is labelled. CA, cholic acid; CDCA, chenodeoxycholic acid; GCA, glyco-cholic acid; GDCA, glyco-deoxycholic acid; GCDCA, glyco-chenodeoxycholic acid; HCA, hyocholic acid; MDCA, murideoxycholic acid; TCA, tauro-cholic acid; TDCA, tauro-deoxycholic acid; TCDCA, tauro-chenodeoxycholic acid; TLCA, tauro-lithocholic acid; TUDCA, tauro-ursodeoxycholic acid; UDCA, ursodeoxycholic acid. **b**–**d**, Quantification of total bile acids (BA; **b**) and the indicated bile acids (**c**, **d**). Each circle represents an individual. **e**, Simplified biosynthetic pathway predicted for the metabolism of CDCA by the gut microbiota. Responsible enzymes are indicated within boxes. 5AR, 5α-reductase; 5BR, 5β-reductase; HSDH, hydroxysteroid dehydrogenase. **b**–**d**, Data are mean ± s.d. $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; one-way ANOVA with Tukey's test. NS, not significant.

probably due to the carriage of 5AR but lack or insufficient activity of 3β-HSDH in these culture conditions. Collectively, a total of 20 Bacteroidales strains were found to be capable of transforming 3-oxo-Δ^4-LCA into 3-oxoalloLCA, 8 of which were able to generate isoalloLCA (Fig. 2b and Extended Data Figs. 7c, 8a). AlloLCA was consistently below the detection limit, suggesting that our 5AR-carrying isolates either lack or have insufficient 3α-HSDH activity in the tested culture conditions. Of note, incubation at pH 9 (representative of the gut environment of centenarians) enhanced the accumulation of 3-oxoalloLCA, isoalloLCA, isoLCA and 3-oxoLCA compared to that at pH 7 (Supplementary Fig. 1), suggesting that alkaline pH stress may promote the activity or expression of enzymes involved in the production of these bile acids.

5AR- and 3β-HSDH-mediated isoalloLCA production

To evaluate our predicted bile acid biosynthetic pathway and, in particular, the hypothesis that 3-oxo-Δ^4-LCA conversion to isoalloLCA is mediated by 5AR and 3β-HSDH, we sequenced the genomes of all 68 isolates (Supplementary Table 4). Sequences orthologous to human 5AR (*SRD5A1*) were identified in 21 Bacteroidales strains with more than 30% amino acid sequence similarity (Fig. 2c, d, Extended Data Fig. 8b and Supplementary Fig. 2a). In all 21 strains, we found clusters of genes that are functionally related to bile acid metabolism, including sequences annotated as NADH:flavin oxidoreductase, which we predicted to be 5BR (Fig. 2c, d, Extended Data Fig. 8b and Supplementary Fig. 2b). We also identified sequences annotated as short-chain dehydrogenase, which we predicted to be 3β-HSDH. These putative 3β-HSDH sequences comprised two groups: group I sequences (3β-HSDH-I) showed high similarity (>40%) to *P. merdae* St3 3β-HSDH, whereas group II sequences (3β-HSDH-II) were closely related to one another but not to *P. merdae* 3β-HSDH (Fig. 2c, d, Extended Data Fig. 8b and Supplementary Fig. 2c, d). We found that carriage of putative 5AR and 3β-HSDH gene clusters was clearly related to 3-oxoalloLCA and isoalloLCA production from 3-oxo-Δ^4-LCA in vitro (Fig. 2b, d). Querying the genome sequences also revealed carriage of putative 3α-HSDH genes by *Parabacteroides distasonis*, *Raoultibacter* and *Eggerthella* strains (Fig. 2d).

To further validate the relevant pathways, we incubated 24 Bacteroidales isolates, including 21 putative 5AR encoders, with 3-oxoalloLCA, 3-oxoLCA or isoLCA (Extended Data Fig. 9a–c). The observed patterns of bile acid transformation were largely consistent with our predicted pathway, although there was substantial substrate specificity and strain-to-strain variation in transformation efficiency. For instance, *P. merdae* St3, *P. distasonis* St4, St5 and Odoribacteraceae St21 exhibited strong 3β-HSDH activity, reflected by both high isoalloLCA production from 3-oxoalloLCA and isoLCA production from 3-oxoLCA, whereas other strains showed less efficient biotransformation or substrate specificity despite carriage of putative 3β-HSDH genes (Extended Data Fig. 9a–c). To examine cooperative activity of isolates, we co-cultured *E. lenta* St34 or *P. distasonis* St4 with *P. merdae* St3 or Odoribacteraceae St21 in the presence of LCA. All combinations resulted in production of isoalloLCA, with the co-culture of *E. lenta* St34 and *P. merdae* St3 giving the highest yield (Extended Data Fig. 9d). Furthermore, *C. scindens* St59

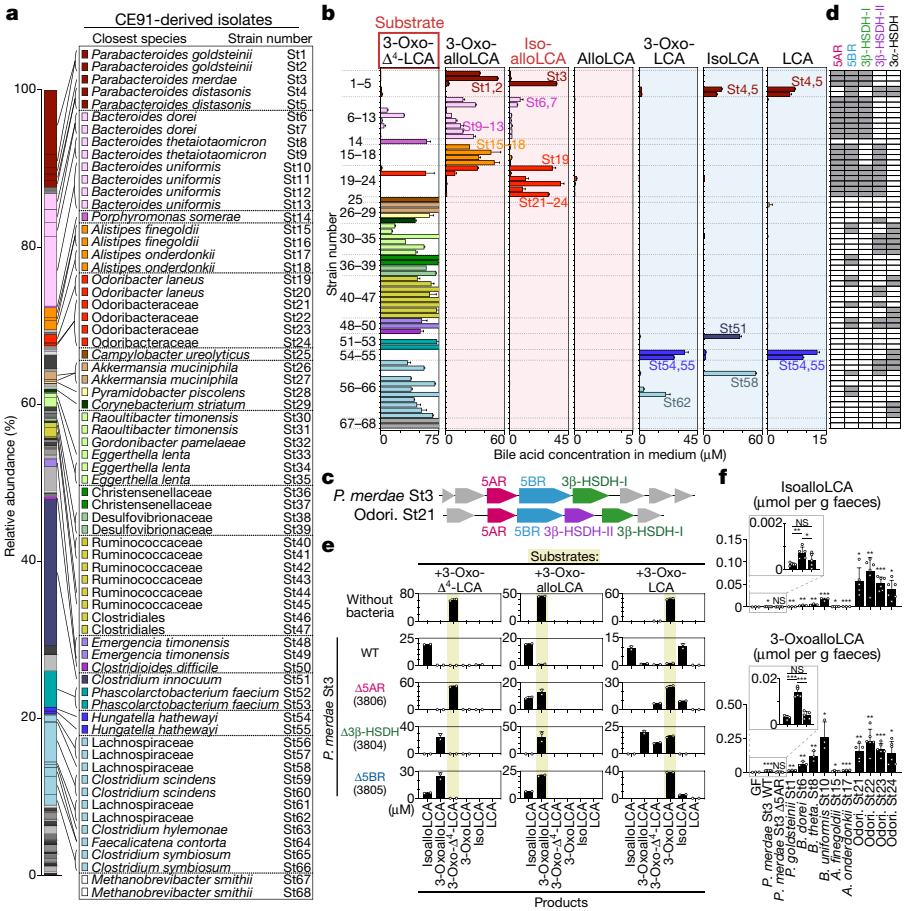


Fig. 2 | Identification of bacterial strains and genes involved in the generation of isoalloLCA and related bile acids. a, Faecal microbial composition of CE91 aligned with a list of the 68 isolated strains. **b**, In vitro bile acid metabolism by the 68 isolated strains using 50 μM 3-oxo-Δ^4-LCA as a starting substrate. Graphs with red backgrounds indicate *trans*-bile acids and blue backgrounds indicate *cis*-bile acids. **c**, Predicted bile-acid-metabolizing enzyme genes in *P. merdae* St3 and Odoribacteraceae St21. **d**, The presence of genes homologous to 5AR, 5BR, 3β-HSDH (3β-HSDH-I, 3β-HSDH-II) and 3α-HSDH in the corresponding strains. **e**, Bile acid metabolism of Δ5AR, Δ3β-HSDH or Δ5BR *P. merdae* St3 mutant strains. We added 50 μM 3-oxo-Δ^4-LCA, 3-oxoalloLCA or 3-oxoLCA as starting substrates. WT, wild type. **f**, Germ-free C57BL/6N mice were colonized with 1 of 12 Bacteroidales strains and fed a 0.1% 3-oxo-Δ^4-LCA diet for 7 days. Faecal isoalloLCA and 3-oxoalloLCA levels were determined. Each dot represents an individual ($n = 3–6$). Data are mean ± s.d. $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; one-way ANOVA with Tukey's test. $^{*}P < 0.05$; unpaired *t*-test for the leftmost panel in **f**. Data in **b**, **e** are representative of two independent experiments conducted in duplicate.

together with *P. merdae* St3 or *P. distasonis* St4 and Odoribacteraceae St21 produced isoalloLCA from CDCA (Extended Data Fig. 9e). Collectively, the Bacteroidales gene clusters identified above can cooperatively produce isoalloLCA and related bile acids (Extended Data Fig. 8a).

We next generated *P. merdae* St3 mutants lacking genes that encode putative 5AR, 3β-HSDH or 5BR (Supplementary Fig. 3). (A lack of genetic tools hampered the introduction of targeted mutations in Odoribacteraceae strains.) As expected, when incubated with 3-oxo-Δ^4-LCA, *P. merdae* Δ5AR did not produce either 3-oxoalloLCA or isoalloLCA, whereas *P. merdae* Δ3β-HSDH was able to generate 3-oxoalloLCA, but not isoalloLCA (Fig. 2e). Consistently, when incubated with 3-oxoalloLCA, *P. merdae* Δ5AR generated isoalloLCA in a manner similar to the wild-type strain, whereas *P. merdae* Δ3β-HSDH did not (Fig. 2e). *P. merdae* Δ3β-HSDH additionally failed to convert 3-oxoLCA into isoLCA, confirming that 3β-HSDH can utilize both *trans*- and *cis*-bile acids as substrates. *P. merdae* Δ5BR produced isoalloLCA from 3-oxo-Δ^4-LCA or 3-oxoalloLCA but showed a defect in transforming 3-oxoLCA into 3-oxo-Δ^4-LCA (Fig. 2e). Together, these results corroborate the involvement of 5AR, 3β-HSDH and 5BR in the production of 3-oxoLCA, isoLCA, 3-oxoalloLCA and isoalloLCA by the human gut microbiota.

We next tested whether our isolates could produce appreciable quantities of isoalloLCA in vivo. Germ-free C57BL/6N mice were monocolonized with one of 12 selected Bacteroidales isolates and fed a 3-oxo-Δ^4-LCA-containing diet. Upon faecal bile acid profiling, only Odoribacteraceae strains were found to induce robust accumulation of isoalloLCA (Fig. 2f). Although *P. merdae* St3 efficiently transformed 3-oxo-Δ^4-LCA to isoalloLCA in vitro, only marginal levels of isoalloLCA were detected in vivo. Mice monocolonized with *P. merdae* St3—but not *P. merdae* Δ5AR—exhibited significant faecal accumulation of 3-oxoalloLCA (Fig. 2f), demonstrating that *P. merdae* 5AR retains its activity in vivo, albeit weakly. *B. uniformis*, *B. thetaiotaomicron* and *B. dorei* strains were capable of producing 3-oxoalloLCA but were not able to effectively convert it into isoalloLCA. Overall, these in vivo data suggest that the activity of commensal 5AR and 3β-HSDH is context-dependent and that the intestinal environment (at least in mice) favours isoalloLCA production by Odoribacteraceae.

Antimicrobial actions of isoalloLCA against pathogens

Secondary bile acids are known to have important roles in several biological contexts, such as the modulation of host metabolic and immune responses (including the regulation of T cells)[13–17] and prevention of intestinal pathogen expansion[18–20]. In particular, DCA, LCA and isoLCA have been implicated in inhibiting the growth of *C. difficile*[18–20]. We

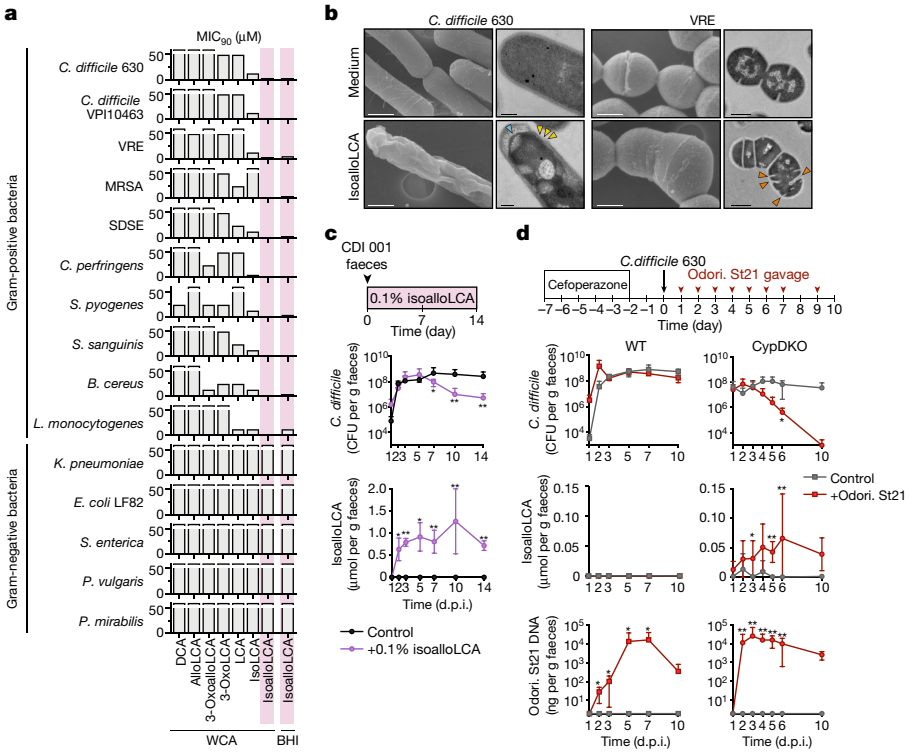


Fig. 3 | IsoalloLCA exerts potent antimicrobial activity against Gram-positive pathogens. a, MIC_{90} of bile acids against Gram-positive and Gram-negative pathogens in WCA and BHI media. MRSA, methicillin-resistant *Staphylococcus aureus*; SDSE, *Streptococcus dysgalactiae* subsp. *equisimilis*. **b**, Representative scanning and transmission electron microscopy images of *C. difficile* 630 and VRE grown with or without 8 μM isoalloLCA for 5 h. Arrows indicate morphological alternations after isoalloLCA treatment. Scale bars, 500 nm (main and right magnified images (VRE)) and 200 nm (left magnified images (*C. difficile* 630)). **c**, Germ-free C57BL/6N mice were colonized with the faecal microbiome of a patient with a *C. difficile* infection (CDI001) and the mice were then placed on a 0.1% isoalloLCA diet ($n = 5$ animals each). Faecal *C. difficile* colony-forming units (CFUs) and isoalloLCA levels were determined. **d**, SPF *Cyp2a12*$^{-/-}$*Cyp2c70*$^{-/-}$ double-knockout (CypDKO) and wild-type C57BL/6N mice were pretreated with antibiotics, colonized with *C. difficile* 630, and then treated by oral gavage with Odoribacteraceae St21 ($n = 3$–6 animals each). Odoribacteraceae St21 colonization, faecal *C. difficile* CFUs and isoalloLCA levels are shown. d.p.i., days post-infection. **c**, **d**, Data are mean ± s.d. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; Mann–Whitney *U*-test (two-tailed) with Welch's correction.

therefore incubated *C. difficile* 630 with various concentrations of bile acids and tracked bacterial growth over time in vitro (Extended Data Fig. 10a, b). Notably, isoalloLCA potently inhibited the growth of *C. difficile* 630. The minimal inhibitory concentration required to prevent ≥90% growth (MIC_{90}) was 2.0 μM, far below that of the other bile acids tested (Fig. 3a, Extended Data Fig. 10a–c and Supplementary Table 5a). IsoalloLCA also potently inhibited growth of toxigenic *C. difficile* VPI10463 and vancomycin-resistant *E. faecium* (VRE) (Fig. 3a and Extended Data Fig. 10a–c). Electron microscopy revealed that isoalloLCA was bactericidal, inducing morphologic alterations such as collapse, swelling and multiple cross walls in *C. difficile* and VRE (Fig. 3b). These patterns of damage are reminiscent of those induced by β-lactam antibiotics. Co-culturing with Odoribacteraceae St21 in conjunction with 3-oxo-Δ^4-LCA supplementation resulted in significant growth inhibition of *C. difficile* and VRE, similar to that observed with isoalloLCA treatment (Extended Data Fig. 10d). By contrast, growth inhibitory effects were not observed when co-culturing was performed with *Clostridium innocuum* St51 (an isoLCA producer) or *P. distasonis* St4 (an isoLCA and LCA producer).

We then examined the effect of isoalloLCA on other Gram-positive, as well as Gram-negative, pathogens. IsoalloLCA strongly inhibited growth of all Gram-positive pathogens tested with MIC_{90} values ranging from 0.5 to 3 μM in Wilkins–Chalgren Anaerobe (WCA) medium and from 3 to 6.25 μM in brain heart infusion (BHI) medium. By contrast, all members of our Gram-negative pathogen panel were resistant to isoalloLCA (Fig. 3a, Extended Data Fig. 10a–c and Supplementary Table 5a). Taken together, these data show that isoalloLCA has strong bactericidal and/or bacteriostatic effects specifically on Gram-positive pathogens. The median absolute concentration of isoalloLCA in faeces of centenarians was 19.5 μM (7.33 μg ml^{-1}) (Extended Data Fig. 5f), which is sufficiently above the in vitro MIC_{90} values, suggesting that it is present at physiologically relevant concentrations in vivo.

We next examined the effect of isoalloLCA on *C. difficile* infection in vivo. Specific-pathogen-free (SPF) C57BL/6N mice were infected with *C. difficile* 630 and given an 0.1% isoalloLCA-containing diet. IsoalloLCA administration resulted in elevated faecal and serum levels of isoalloLCA without causing systemic toxicity, and induced a significant reduction in *C. difficile* shedding (Extended Data Fig. 11a–c). We next colonized germ-free C57BL/6N mice with faecal microbiota from a patient with a *C. difficile* infection and, as above, subsequent isoalloLCA administration significantly reduced *C. difficile* shedding by these mice (Fig. 3c). Next, we examined whether colonization with Odoribacteraceae St21 could help to fend off *C. difficile* in vivo. To this end, we used *Cyp2c70*$^{-/-}$*Cyp2a12*$^{-/-}$ double-knockout mice[21], which mimic the human bile acid profile with high levels of LCA derivatives and lack muricholic acids (Extended Data Fig. 11d–g). SPF *Cyp2c70*$^{-/-}$*Cyp2a12*$^{-/-}$ double-knockout mice were infected with *C. difficile* 630, followed by repeated oral gavage with Odoribacteraceae St21. Odoribacteraceae administration significantly increased faecal isoalloLCA levels and markedly reduced *C. difficile* shedding to below the limit of detection (Fig. 3d). By contrast, wild-type mice undergoing the same treatment regimen did not produce isoalloLCA nor showed inhibition of *C. difficile* shedding, despite a similar extent of Odoribacteraceae engraftment (Fig. 3d). Together, these results suggest that Odoribacteraceae and

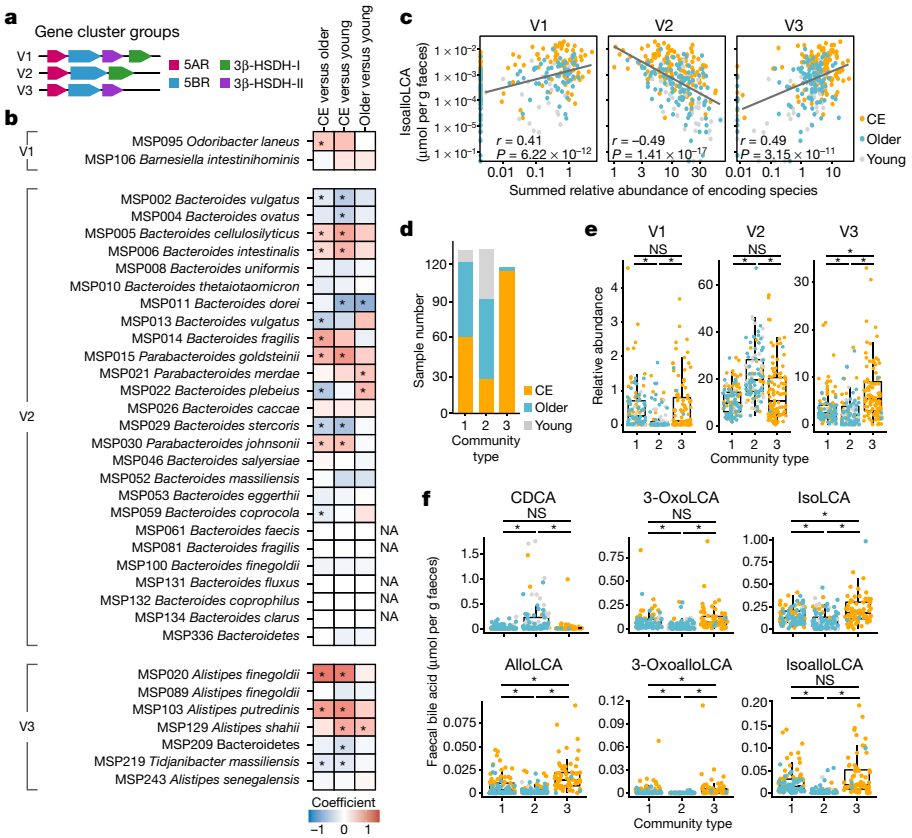


Fig. 4 | Association of gut microbiome structure with bile acid profile. **a**, Categorization of 5AR and 3β-HSDH carriage into clusters V1, V2 and V3. **b**, Coefficient from the linear model indicates enrichment (red) or depletion (blue) of metagenome species (MSPs) in the respective comparisons: centenarian versus older participants, centenarian versus young participants, and older versus young participants; in each case, the latter group is used as a reference. NA, not applicable. **c**, Spearman's coefficient (*r*) and significance (*P*) between faecal isoalloLCA level and summed relative abundance of cluster V1-, V2- or V3-encoding species. **d**, Stratification of gut microbiomes into microbial community types 1, 2 and 3 using Dirichlet multinomial mixtures. **e**, **f**, Relative abundance of species carrying each gene cluster (**e**) and faecal bile acid levels (**f**) within each community type. **b**, **e**, **f**, *False-discovery rate (FDR)-adjusted $P < 0.05$; linear model. Horizontal lines indicate the median; box boundaries indicate the interquartile range; whiskers represent values within 1.5× the interquartile range of the first and third quartiles. Metagenomics: centenarian ($n = 176$ (153 individuals)), older ($n = 110$) and young ($n = 44$) participants. Metagenomics and metabolomics: centenarian ($n = 119$), older ($n = 107$) and young ($n = 39$) participants. Each dot represents an individual.

their derivative bile acids, including isoalloLCA, contribute to the inhibition of Gram-positive pathogen expansion in vivo.

Effects of isoalloLCA on the commensal gut microbiota

We next investigated whether isoalloLCA affects common members of the human microbiota. A total of 42 prevalent gut microbiota members were selected from our culture collection, and each was incubated with isoalloLCA. IsoalloLCA did not appreciably affect the growth of most Gram-negative commensals such as *Bacteroides*. By contrast, it substantially interfered with the growth of Gram-positive commensals (Extended Data Fig. 12a, b and Supplementary Table 5b). However, MIC_{90} values for commensal strains were generally higher than those for pathogens, and electron microscopy revealed that commensal morphology (*Clostridium sporogenes*, *Clostridium indolis* and *Clostridium* HGF2) was preserved when incubated with 2.5 μM isoalloLCA (1.25× MIC_{90} for *C. difficile*) (Extended Data Fig. 12c). Moreover, culturing commensal strains in peptone- and amino-acid-rich BHI medium conferred increased resistance to isoalloLCA compared to culturing in WCA medium (Extended Data Fig. 12b, d and Supplementary Table 5a, b), whereas pathogens generally remained sensitive to isoallo LCA irrespective of the medium used (Fig. 3a). These results indicate that although the concentration at which isoalloLCA exerts antimicrobial effects varies substantially depending on environmental conditions, Gram-positive pathogens consistently remain more sensitive to it than commensals.

To further evaluate the effects of isoalloLCA within the context of a complex gut flora, we incubated human faecal microbiota from healthy volunteers with bile acids. After incubation with isoalloLCA, we observed a pronounced reduction in Gram-positive species such as *Faecalibacterium*, *Bifidobacterium* and *Streptococcus*, along with a corresponding increase in Gram-negative species such as *Bacteroides* and *Alistipes* (Extended Data Fig. 13a, b), consistent with the enrichment of *Bacteroides* and *Alistipes* and depletion of *Streptococcus* species seen in the microbiomes of centenarians (Extended Data Figs. 2g, 3c–e, 13c–e). We additionally examined the effect of isoalloLCA on human microbiota in the context of *C. difficile* infection by incubating stool samples from three patients who had a *C. difficile* infection with bile acids. IsoalloLCA altered the microbial community structure and exerted the strongest inhibitory effect on *C. difficile* in all three samples (Extended Data Fig. 13f–h). These results are consistent with the notion that isoalloLCA can directly affect the structure of intestinal microbial communities and protect against potential pathogens.

Association of gut microbiome with bile acid profile

We next conducted a more in-depth metagenomic analysis of the participants. 5AR and 3β-HSDH gene clusters were identified in 35 species (all Bacteroidales). These clusters can be categorized into three groups: V1, V2 and V3 (Fig. 4a). Cluster V1 contains 5AR together with both 3β-HSDH-I and 3β-HSDH-II, and is found in the genomes of

Odoribacter, whereas cluster V3 contains 5AR and 3β-HSDH-II and is found in the genomes of *Alistipes*. By contrast, cluster V2 comprises 5AR and 3β-HSDH-I and is carried by diverse *Bacteroides* species (Fig. 4b). We found that clusters V1 and V3 were significantly associated with the levels of faecal isoalloLCA, 3-oxoalloLCA, alloLCA, 3-oxoLCA and isoLCA in our cohort (Fig. 4c and Extended Data Fig. 14a, b). Moreover, V1- or V3-carrying *Odoribacter* and *Alistipes* species were more abundant in centenarians than other groups (Fig. 4b and Extended Data Fig. 14a). By contrast, the abundance of cluster V2 was negatively associated with the concentration of these bile acids (Fig. 4c and Extended Data Fig. 14b). These results suggest that carriage of 5AR together with 3β-HSDH-II (rather than 3β-HSDH-I) is required to produce isoalloLCA and related bile acids. Moreover, bacterial species carrying cluster V2 may compete with those carrying V1 or V3 (Extended Data Fig. 14c). Notably, several species such as *Anaerotruncus* spp. and *M. smithii*—which do not encode 5AR—were strongly positively associated with isoalloLCA and related bile acid concentration (Extended Data Fig. 14d), suggesting that bile acid production by 5AR- and 3β-HSDH-II-carrying species may be modulated by other microbial community members.

Given the context-dependent activity of 5AR gene clusters and the need to better account for interindividual microbial community variation, we stratified the microbiomes of our cohorts into three microbial community types (termed community types 1, 2 and 3)[22] (Extended Data Fig. 15a–c). The community structure of the young group was predominantly type 2, whereas the community structure of the centenarians was predominantly type 3 followed by type 1 (Fig. 4d). Notably, individuals with community types 1 and 3 tended to have microbiomes with higher abundances of species carrying V1 and V3 relative to those carrying V2 (Fig. 4e and Extended Data Fig. 15d). Moreover, higher levels of 3-oxoLCA, isoLCA, alloLCA, 3-oxoalloLCA and isoalloLCA were observed in individuals with community types 1 and 3 than those with type 2, whereas this trend was reversed for CDCA (Fig. 4f). Notably, both the faecal bile acid profile and community type of centenarians were generally stable between longitudinal samples (Extended Data Fig. 15e, f). Overall, these results further support the notion that the underlying microbial community structure affects the intestinal bile acid profile, and that community types 1 and 3 are associated with Japanese centenarians and may promote an environment that facilitates the expansion and function of isoalloLCA-producing species.

Discussion

In the present study, we identified centenarian-specific gut microbiota signatures and defined bacterial species as well as genes and/or pathways that promote the generation of isoLCA, 3-oxoLCA, 3-oxoalloLCA and isoalloLCA. Several reports have demonstrated that bile acids contribute to the protection against enteropathogenic infection[18–20]. To our knowledge, isoalloLCA is one of the most potent antimicrobial agents that is selective against Gram-positive microorganisms, including multidrug-resistant pathogens, suggesting that it may contribute to the maintenance of intestinal homeostasis by enhancing colonization-resistance mechanisms. There are a number of limitations to this study. In particular, there are several confounding factors in our cohorts, making it difficult to discern the major driving forces behind the observed enrichment of isoalloLCA-producing organisms in centenarians. In addition, the causal relationship between the unique bile acids of centenarians and longevity needs to be validated with longitudinal surveys, additional participants and long-term analyses of animal models. Regardless, we may be able to exploit the unique bile-acid-metabolizing abilities of the bacterial strains identified in this study to rationally manipulate the bile acid pool and combat diseases caused by Gram-positives such as antibiotic-resistant *C. difficile* and VRE.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03832-5.

1. Hirata, T. et al. Associations of cardiovascular biomarkers and plasma albumin with exceptional survival to the highest ages. *Nat. Commun.* **11**, 3820 (2020).
2. Sebastiani, P. et al. Meta-analysis of genetic variants associated with human exceptional longevity. *Aging* **5**, 653–661 (2013).
3. Wu, L. et al. A cross-sectional study of compositional and functional profiles of gut microbiota in Sardinian centenarians. *mSystems* **4**, e00325-19 (2019).
4. Franceschi, C., Garagnani, P., Parini, P., Giuliani, C. & Santoro, A. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nat. Rev. Endocrinol.* **14**, 576–590 (2018).
5. Bárcena, C. et al. Healthspan and lifespan extension by fecal microbiota transplantation into progeroid mice. *Nat. Med.* **25**, 1234–1242 (2019).
6. Biagi, E. et al. Gut microbiota and extreme longevity. *Curr. Biol.* **26**, 1480–1485 (2016).
7. Rampelli, S. et al. Shotgun metagenomics of gut microbiota in humans with up to extreme longevity and the increasing role of xenobiotic degradation. *mSystems* **5**, e00124-20 (2020).
8. Funabashi, M. et al. A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature* **582**, 566–570 (2020).
9. Ridlon, J. M., Harris, S. C., Bhowmik, S., Kang, D. J. & Hylemon, P. B. Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes* **7**, 22–39 (2016).
10. David, L. A. et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
11. Devlin, A. S. & Fischbach, M. A. A biosynthetic pathway for a prominent class of microbiota-derived bile acids. *Nat. Chem. Biol.* **11**, 685–690 (2015).
12. Nixon, M., Upreti, R. & Andrew, R. 5α-Reduced glucocorticoids: a story of natural selection. *J. Endocrinol.* **212**, 111–127 (2012).
13. Hang, S. et al. Bile acid metabolites control T_H17 and T_{reg} cell differentiation. *Nature* **576**, 143–148 (2019).
14. Campbell, C. et al. Bacterial metabolism of bile acids promotes generation of peripheral regulatory T cells. *Nature* **581**, 475–479 (2020).
15. Song, X. et al. Microbial bile acid metabolites modulate gut $ROR\gamma^+$ regulatory T cell homeostasis. *Nature* **577**, 410–415 (2020).
16. Paik, D. et al. Human gut bacteria produce T_H17-modulating bile acid metabolites. Preprint at https://doi.org/10.1101/2021.01.08.425913 (2021).
17. Li, W. et al. A bacterial bile acid metabolite modulates T_{reg} activity through the nuclear hormone receptor NR4A1. Preprint at https://doi.org/10.1101/2021.01.08.425963 (2021).
18. Buffie, C. G. et al. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* **517**, 205–208 (2015).
19. Begley, M., Gahan, C. G. M. & Hill, C. The interaction between bacteria and bile. *FEMS Microbiol. Rev.* **29**, 625–651 (2005).
20. Thanissery, R., Winston, J. A. & Theriot, C. M. Inhibition of spore germination, growth, and toxin activity of clinically relevant *C. difficile* strains by gut microbiota derived secondary bile acids. *Anaerobe* **45**, 86–100 (2017).
21. Honda, A. et al. Regulation of bile acid metabolism in mouse models with hydrophobic bile acid composition. *J. Lipid Res.* **61**, 54–69 (2020).
22. Holmes, I., Harris, K. & Quince, C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS ONE* **7**, e30126 (2012).

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Human sample collection

Faecal samples and blood tests from Japanese young and older particiapatns, centenarians, and lineal relatives of centenarians were obtained following a protocol approved by the Institution Review Board of Keio University School of Medicine (code 20150075 for young healthy donors; 20160297 for older cohorts (as part of the Kawasaki Ageing and Wellbeing project); and 20022020 for centenarians and lineal relatives of centenarians (as part of The Japan Semi-supercentenarian Study[1]). The microbiome dataset of Japanese patients with IBD was obtained from previous studies[23]. Informed consent was obtained from each donor before participation. All experiments adhered to the regulations mandated by these Review Boards. All study procedures were performed in compliance with the relevant ethical regulations. The Japan Semi-supercentenarian Study[1] and Kawasaki Ageing and Wellbeing project are registered in the University Hospital Medical Information Network Clinical Trial Registry as observational studies (ID: UMIN000040447 and UMIN000026053, respectively). Faecal sample collection of patients with a *C. difficile* infection was conducted under a protocol approved by the Institution Review Board of Keio University School of Medicine (code 20150075).

Metagenomic sequencing and 16S rRNA gene pyrosequencing of human stool samples

Faecal samples were suspended in an equal volume of PBS containing 20% glycerol and 10 mM EDTA and stored at −80 °C until use. After thawing, 100 µl of faecal suspension was gently mixed and incubated in 800 µl TE10 (10 mM Tris-HCl, 10 mM EDTA) buffer containing RNase A (final concentration of 100 µg ml^{-1}; Invitrogen) and lysozyme (final concentration of 15 mg ml^{-1}; Sigma) for 1 h at 37 °C. Purified achromopeptidase (final concentration of 2,000 U ml^{-1}; Wako) was added and further incubated for 30 min at 37 °C. SDS (final concentration of 1%) and proteinase K (final concentration of 1 mg ml^{-1}; Roche) was further added to the mixture and incubated for 1 h at 55 °C. High-molecular-mass DNA was extracted with phenol:chloroform:isoamyl alcohol (25:24:1 at pH 7.9), precipitated with isopropanol (equal volume to the aqueous phase), washed with 1 ml 70% ethanol and gently resuspended in 30 µl of TE buffer.

We collected faecal samples from the three groups and characterized the microbiome by both 16S rRNA amplicon and whole-metagenome shotgun sequencing, excluding samples from participants undergoing antibiotic treatment (three centenarians and one older participant). The 16S rRNA sequencing was performed using MiSeq according to the Illumina protocol. PCR was performed using 27Fmod 5′-AGRGTTTGATYMTGGCTCAG-3′and 338R 5′-TGCTGCCTCCCGTAGGAGT-3′ to the V1–V2 region of the 16S rRNA gene. Amplicons generated from each sample (around 330 bp) were purified using AMPure XP magnetic beads (Beckman Coulter). DNA was quantified using a Quant-iT Picogreen dsDNA assay kit (Invitrogen) and Infinite M Plex plate reader (Tecan) and then stored at 4 °C. The pooled amplicon library was sequenced using a MiSeq Reagent Kit v2 (500 cycles) and Miseq sequencer (Illumina; 2 × 250-bp paired-end reads). After demultiplexing the 16S sequence reads based on the sample-specific index, primer sequences were trimmed by Cutadapt v.1.15. The trimmed reads were uploaded to the DADA2 R package v.1.18.0 to construct amplicon sequence variants (ASVs) using the filterAndTrim function with standard parameters (maxN = 0, truncQ = 2 and maxEE = 2). Possible chimeric reads were removed with the removeBimeraDenovo function of the DADA2. The taxonomic assignment of each ASV was determined by similarity searching against the NCBI RefSeq database using the GLSEARCH program.

Metagenomic sequencing libraries were prepared from 2 ng of input DNA using the Nextera XT DNA Library Preparation kit (Illumina) according to the manufacturer's recommended protocol. Libraries were pooled by equal volume, and insert sizes and concentrations for each pooled library were determined using an Agilent Bioanalyzer DNA 1000 kit (Agilent Technologies). Sequencing was performed on an Illumina NovaSeq 6000 with 151-bp paired-end reads to yield around 10 million paired-end reads per sample. Data were analysed using the Broad Picard Pipeline, which includes demultiplexing and data aggregation (https://broadinstitute.github.io/picard). The quality control for the metagenomic data was conducted using Trim_Galore! to detect and remove sequencing adapters (minimum overlap of 5 bp) and KneadData v.0.7.2 to remove human DNA contamination and trim low-quality sequences (HEADCROP:15, SLIDINGWINDOW:1:20), retaining reads that were at least 50 bp long. Metagenomic reads were assembled individually for each sample into contigs using MegaHIT[24], followed by an open-reading-frame prediction with Prodigal[25] and retaining predicted genes that had both a start and a stop codon. A non-redundant gene catalogue was constructed by clustering predicted genes based on sequence similarity at 95% identity and 90% coverage of the shorter sequence using CD-HIT[26,27]. Reads were mapped to the gene catalogue with BWA requiring a unique, strong mapping with at least 95% sequence identity over the length of the read[28], counted (count matrix) and normalized to transcripts per kilobase per million (TPM matrix) using in-house scripts. The count matrix served as an input for binning genes into metagenomic species pan-genomes (core and accessory genes) using MSPminer with default settings[29]. We represented the abundance of each metagenomic species (MSP) in a sample as a median TPM for the 30 top representative core genes reported by MSPminer. Assembled genes were annotated at species, genus and phylum levels with NCBI RefSeq (version May 2018) as previously described[30]. To annotate phylogenetically MSPs that had no match to any species from NCBI RefSeq, we used PhyloPhlan with default settings[31]. The α-diversities were calculated using the Shannon index and the β-diversity was calculated using Bray–Curtis dissimilarity based on relative abundances at the species levels (Vegan package in R). The non-redundant gene catalogue was queried using USEARCH ublast[30] with proteins in the *bai* operon of *C. scindens*[8] or proteins in the bacterial isolates reported here as 5AR, 5BR, 3β-HSDH-I or 3β-HSDH-II, to identify and annotate homologous proteins with at least 40% identity and 80% coverage to the query sequence. An identical processing pipeline has been applied to a dataset describing the gut microbiome in Sardinian centenarians[3].

In the subsequent analysis, we only used samples with at least 4 million reads after the quality-control step. Additionally, we discarded samples that were collected while the participant was undergoing any antibiotic treatment. To test the differential abundance of species or phyla and differences in the Shannon diversity index, we used linear random-effects modelling (centenarians versus young or older controls) or fixed-effects modelling (older versus young controls), as implemented in the lmer and lm functions in R. Furthermore, for the analysis of species differential abundance, we restricted the analysis to MSPs present in at least 10% of samples, zeros were replaced by half of the smallest non-zero measurement on a per-feature basis and $\log_{10}$ transformation was applied to the relative abundances for normality. Linear modelling included fixed-effect covariates: sex (male or female) and cohort information (centenarian, older or young); random effects included participant information to account for more than one sample among a few centenarians. The permutational multivariate analysis of variance (PERMANOVA) analysis as implemented in the adonis function in the R package vegan was applied to the Bray–Curtis dissimilarity to identify the correlation between age group (centenarian, older or young) and sex information and the composition of the gut microbiome as a whole. To determine the clustering of samples based on

their metagenomes into community types, we applied the Dirichlet multinomial mixtures algorithm[22] to MSPs abundance matrix (median reads per killobase for 30 top representative core genes reported by MSPminer). The appropriateness of community types was determined based on the lowest Laplace approximation score.

Faecal bile acid quantification using LC–MS/MS

Accurately weighed freeze-crushed faecal samples were resuspended in 20 times the volume of water (for example, 12.5 mg of faece in 250 µl water). Then, 250 µl of faecal suspension was homogenized in 747 µl of 0.27 N NaOH by ultrasonication for 1 h in a screw-cap glass vial containing 3 µl of deuterium-labelled internal standards (d_4-CA, d_4-GCDCA, d_4-TCDCA, d_4-CDCA-3S and d_4-LCA; 50 µM each). After incubation for 1 h at room temperature, the pH was adjusted to 8.0 using 8 N HCl, mixed with 110 µl of 0.5 M EDTA/0.5 M Tris-HCl. The solution was centrifuged at 15,000 rpm and the supernatant was transferred onto a solid-phase extraction cartridge (Agilent Bond Elut C18, 100 mg per 3 ml, preconditioned with 1 ml of methanol and 3 ml of water, three times). The cartridge was washed with 1 ml of water and captured bile acids were eluted with 300 µl 90% ethanol. Then, 2 µl was used for LC/electrospray ionization (ESI)–MS/MS injected to a Triple Quad 6500+ tandem mass spectrometer, equipped with an ESI probe and Exion LC AD ultra-high pressure liquid chromatography system (SCIEX). A separation column, InertSustain C18 (150 mm × 2.1 mm inner dimension, 3 µm particle size; GL Sciences), was used at 40 °C. A mixture of 10 mM ammonium acetate and acetonitrile was used as the eluent and the separation was carried out by linear gradient elution at a flow rate of 0.2 ml min^{-1}. The mobile phase composition was gradually changed as follows: ammonium acetate:acetonitrile (86:14, v/v) for 0.5 min; (78:22, v/v) for 0.5–5 min; (72:28, v/v) for 5–28 min; (46:54, v/v) for 28–55 min; (2:98, v/v) for 55–66 min; and (2:98, v/v) for 4 min. The total run time was 70 min. To operate the LC/ESI–MS/MS, the following MS parameters were used for the positive-ion multiple reaction monitoring (MRM) mode: ion spray voltage, 5,500 V; interface temperature, 500 °C; curtain gas, 30 psi; collision gas (nitrogen), 9 psi; nebulizing gas, 80 psi; and heater gas, 80 psi. For the negative-ion MRM mode: ion spray voltage, −4,500 V; interface temperature, 500 °C; curtain gas, 30 psi; collision gas (nitrogen), 9 psi; nebulizing gas, 80 psi; and heater gas, 80 psi. Samples were obtained using Analyst software v.1.71 and analysed using SCIEX OS-MQ software v.1.7.0.36606. The detailed set-up for each of the 43 bile acid compounds is listed in Supplementary Table 6.

Faecal SCFA, pH and ammonia measurements

The faecal SCFA concentration was determined by gas chromatography–mass spectrometry (GC–MS) (Shimadzu QP2020 system with a flame ionization detector), equipped with PAL RTC autosampler (CTC Analytics). Using the PAL RTC autosampler, 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4methylmorpholinium and *n*-octylamine (10 µl of each reagent at a concentration of 80 µM) were added to each faecal sample and reacted for 9 h before injection and analysis by GC–MS. Fused silica capillary columns (30 m × 0.25 mm, coated with 0.25-µm film thickness) were used with helium as the carrier gas. The injection port temperature was set to 250 °C. The initial oven temperature was held at 60 °C for 2 min and then ramped to 330 °C at a rate of 15 °C per min. MS parameters were set to: ion source temperature at 200 °C, interface temperature at 280 °C and loop time of 0.3 s. For the GC–MS measurement, 50 µl of faecal samples with a concentration of 0.5 µg $µl^{-1}$ and 20 µg $µl^{-1}$ prepared in ethanol were mixed with 10 µl of acetic acid-d_4 (80 µM). Samples were analysed and quantified using LabSolutions Insight GC–MS software (Shimadzu).

Faecal pH was measured from the supernatant from 0.1 mg $µl^{-1}$ of faecal suspension in distilled water using a pH meter (Horiba). From the same faecal suspension, the faecal ammonia level was quantified using an enzymatic ammonia ELISA assay kit (Abcam) according to the manufacturer's protocol.

Isolation of bacterial strains from a centenarian

A faecal sample from a supercentenarian (CE91, Japanese woman, age over 110 years) was suspended in an equal volume (w/v) of PBS containing 20% glycerol, snap-frozen in liquid nitrogen, and stored at −80 °C until use. Then, 200 µl of thawed faecal suspension was serially diluted with PBS and 100 µl was seeded onto nonselective (Brucella agar with haemin, vitamin K1, lysed rabbit blood and defibrinated sheep blood (BHK-RS), Kyokuto) and selective (for Gram-negative bacteria, paromomycin and vancomycin supplemented BHK agar, Kyokuyo; for Clostridial bacteria, oxoid-reinforced Clostridial agar, ThermoFisher) plates and grown inside an anaerobic chamber (Coy Laboratory Products) under anaerobic conditions (80% N_2, 10% H_2 and 10% CO_2) at 37 °C. Individual colonies emerged after 72 h and up to 10 days of incubation were picked. Isolated strains were identified by PCR amplification of the 16S rRNA gene region with universal primers (27Fmod 5′-AGRGTTTGATYMTGGCTCAG-3′; 1492R 5′-GGYTACCTTGTTACGACTT-3′) for Sanger sequencing and using the NCBI genome database. Individual isolates in the culture collection were given species name with >98.0% of 16S rRNA sequence homology, family name with >94.5% similarity and order name with >86.5% similarity. Bacterial isolates were cryo-preserved in 20% glycerol in optimal culture broth at −80 °C.

In vitro screening of microbial bile acid metabolism

Under anaerobic conditions, isolated bacteria strains were cultured together with 50 µM CDCA, 3-oxo-Δ^4-LCA or LCA to screen for their bile acid metabolism in a 96-deep-well plate (Treff Lab) covered with a gas-permeable membrane (Breathe-easier, Diversified Biotech). In brief, 20 µl of bacterial culture in exponential to stationary phase was inoculated into 1 ml of Wilkins–Chalgren Anaerobe (WCA) media (ThermoFisher) adjusted to pH 7 (using MOPS buffer solution; Dojindo) or pH 9 (using TAPS buffer solution; Dojindo). Several bacterial strains required growth in reinforced Clostridial or WCA media supplemented with additional nutrients. Ruminococcaceae St42, St43 and St45, *Clostridiales* St47 and Lachnospiraceae St56 were cultured in reinforced Clostridial medium, whereas *Phascolarctobacterium faecium* St52 and St53 were cultured in reinforced Clostridial medium supplemented with sodium succinate (20 mmol l^{-1}). For *Akkermansia muciniphila* St26 and St27, WCA medium supplemented with ammonium chloride (1.0 g l^{-1}), L-cysteine (1.0 g l^{-1}), vitamin K (0.5 mg l^{-1}), haemin (5 mg l^{-1}) and 0.29% volatile fatty acid solution (based on DSMZ 1611 YCFA modified medium) was used. *A. finegoldii* St16, *Campylobacter ureolyticus* St25, Christensenellaceae St36 and St37, and Ruminococcaceae St44 were cultured in WCA medium supplemented with 4% salt solution (0.2 g l^{-1} calcium chloride, 0.2 g l^{-1} magnesium sulfate, 1g l^{-1} dipotassium hydrogen phosphate, 1 g l^{-1} potassium dihydrogen phosphate, 10 g l^{-1} sodium hydrogen carbonate, and 2 g l^{-1} sodium chloride), ammonium chloride (1.0 g l^{-1}), L-cysteine (1.0 g l^{-1}), vitamin K (0.5 mg l^{-1}), haemin (5 mg l^{-1}), sodium acetate (1.0 g l^{-1}), sodium formate (0.15 g l^{-1}), sodium fumarate (0.15 g l^{-1}), sodium thioglycolate (0.3 g l^{-1}), 1% ATCC vitamin solution, and 1% ATCC Trace element solution. For *M. smithii* St67 and St68, the above modified WCA medium was further supplemented with sodium bicarbonate (0.25 g l^{-1}), sodium sulfide (0.05 g l^{-1}) and sodium formate (1.36 g l^{-1})[32]. After 48 h of anaerobic incubation at 37 °C, culture supernatants were collected and stored at −20 °C until sample preparation for analysis.

For sample preparation, 100 µl of the culture supernatant was transferred into a screw-cap glass vial containing 10 µl of deuterium-labelled internal standards (d_4-CA, d_4-CDCA and d_4-LCA; 1 nmol ml^{-1} each). Then, 400 µl water was added and the sample was sonicated for 10 min and applied onto the solid-phase extraction cartridge (Agilent Bond Elut C18, 100 mg per 1 ml; preconditioned with 1 ml methanol and 3 ml water). The cartridge was washed with 1 ml water and captured bile acids were eluted with 1 ml 90% ethanol. After solvent evaporation, the

remaining residue was dissolved in 100 µl 50% ethanol, of which 2 µl of the solution was injected to LC/ESI–MS/MS spectrometer (LC-MS8040 tandem mass spectrometer, equipped with an ESI probe and Nexera X2 ultra-high-pressure liquid chromatography system; Shimadzu). A separation column, InertSustain C18 (150 mm × 2.1 mm inner diameter, 2 µm particle size; GL Sciences), was used at 40 °C. Mixture A (10 mM ammonium acetate, 0.01% formic acid and 20% acetonitrile) and mixture B (30% acetonitrile and 70% methanol) were used as the eluent, and the separation was carried out by linear gradient elution at a flow rate of 0.2 ml min^{-1}. The mobile phase composition was gradually changed as follows: mixture A:B (80:20, v/v) for 0.1 min; (48:52, v/v) for 0.1–1 min; (30:70, v/v) for 1–27 min; (0:100, v/v) for 27–27.1 min; (0:100, v/v) for 27.1–33 min; (80:20, v/v) for 33–33.1 min; and (80:20 v/v) for 33.1–38 min. The total run time was 38 min. To operate the LC/ESI–MS/MS, the following MS parameters were used: heating block temperature, 400 °C; nebulizing gas flow, 3 l min^{-1}; drying gas flow, 15 l min^{-1}; interface temperature, 300 °C; collision gas (argon) pressure, 230 kPa; collision energy, negative (11 to 31 eV) and positive (−16 to −19 eV) ion modes. Samples were analysed and quantified using LabSolutions Insight LC–MS software (Shimadzu). The detailed set-up for each of the bile acid compounds is listed in Supplementary Table 7.

Bacterial whole-genome sequencing

The extracted genomic DNA of 68 isolated strains was sheared to yield DNA fragments. The genome sequences were determined by the whole-genome shotgun strategy using PacBio Sequel and Illumina MiSeq sequencers. The library of the Illumina Miseq 2 × 300-bp paired-end sequencing was prepared using the TruSeq DNA PCR-Free kit (target length = 550 bp) and all of the MiSeq reads were trimmed and filtered with a more than 20 quality value using FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The library of the PacBio Sequel sequencing was prepared using the SMRTbell template prep kit 2.0 (target length = 10–15 kb) without DNA shearing. After the removal of the internal control and adaptor trimming by Sequel, error correction of the trimmed reads was performed using Canu v.1.8 with additional options (corOutCoverage = 10,000, corMinCoverage = 0, corMhapSensitivity = high). De novo hybrid assembly of the filter-passed MiSeq reads and the corrected Sequel reads were performed using Unicycler v.0.4.8, which contained checks for overlapping and circularization to generate circular contigs. The gene prediction and annotation of the generated contigs were performed using the Rapid Annotations based on Subsystem Technology (RAST) server[33] and Prokka software tool[34]. Default parameters were used unless otherwise specified.

Mutant generation

The deletion mutants (Δ5AR, Δ5BR and Δ3β-HSDH) of *P. merdae* St3 were generated by conjugation-mediated plasmid transfection and selection of double-crossover resolvents with a rhamnose-inducible ssBfe1 cassette[35]. Approximately 2-kb sequences flanking the coding region were amplified by PCR (PCR primers used in this study are listed in Supplementary Table 8) and assembled into the PstI and SalI sites of the suicide vector pLGB30 using the HiFi DNA Assembly (NEB) as per the manufacturer's protocol. Then, 1 µl aliquots of each reaction were transformed into electro-competent *Escherichia coli* MFDpir[36]. Transformants were conjugated with *P. merdae* St3 as follows. The donor (*E. coli* MFDpir) and recipient (*P. merdae* St3) strains were cultured in LB and BHI media, respectively, to an optical density at 600 nm (OD_{600}) of 0.5 and mixed at a ratio of 1:1. The mixture was dropped onto a BHI agar plate and incubated anaerobically at 37 °C for 16 h. Transconjugants were selected on BHI agar plates containing tetracycline (6 µg ml^{-1}). Subsequently, to select for loss of the plasmid from the genome by a second crossover, transconjugants were plated on M9 agar supplemented with 0.25% (w/v) glucose, 50 mg l^{-1} L-cysteine, 5 mg l^{-1} haemin, 2.5 µg l^{-1} vitamin K1, 2 mg l^{-1} $FeSO_4 \cdot 7H_2O$, 5 µg l^{-1} vitamin B12 and 10 mM rhamnose. Successful deletions were confirmed by PCR and Sanger sequencing.

Bacterial growth inhibition assays

Clostridioides difficile strain 630 (ATCC BAA-1382), *C. difficile* VPI 10463 (ATCC 43255), vancomycin-resistant *E. faecium* (ATCC 700221), methicillin-resistant *S. aureus* (JCM 16555), *S. dysgalactiae* subsp. *equisimilis* (ATCC 12394), *Clostridium perfringens* (JCM 1290^T), *Streptococcus pyogenes* (JCM 5674^T), *Streptococcus sanguinis* (JCM 5708^T), *Bacillus cereus* (JCM 2152^T), *Listeria monocytogenes* strain EGD, carbapenemase-resistant *Klebsiella pneumoniae* (ATCC BAA-1705), adherent invasive *E. coli* LF82 (a gift from N. Banich), *Salmonella enterica* serovar Typhimurium SL1344, *Proteus mirabilis* (JCM 1669^T) and *Proteus vulgaris* (JCM 20013) were tested for inhibition assays.

For gut commensals, *C. scindens* (ATCC 35704^T), *Clostridium sporogenes* (ATCC 15579), *Dorea formicigenerans* (ATCC 27755^T), *Ruminococcus lactaris* (ATCC 29176^T), *Bacteroides fragilis* (ATCC 25285^T), *C. indolis* (JCM 1380^T), *Clostridium hiranonis* (JCM 10541^T), *C. hylemonae* (JCM 10539^T), *Clostridium nexile* (JCM 31500^T), *Clostridium butyricum* (JCM 1391^T), *Dorea longicatena* (JCM 11232^T), *Eubacterium hallii* (JCM 31263), *Streptococcus thermophilus* (JCM 17834^T), *R. gnavus* (JCM 6515^T), *Anaerotruncus colihominis* (JCM 15631^T), *Blautia producta* (JCM 1471^T), *Blautia obeum* (JCM 31340), *Bifidobacterium bifidum* (JCM 1254), *Bifidobacterium breve* (JCM 1192^T), *Bifidobacterium longum* subsp. *longum* (JCM 1217^T), *Lactobacillus casei* (JCM 1134^T), *Lactobacillus paragasseri* (JCM 1130), *Lactobacillus reuteri* (JCM 1112^T), *Collinsella aerofaciens* (JCM 10188^T), *Roseburia intestinalis* (JCM 17583^T), *E. lenta* (JCM 9979^T), *Bacteroides caccae* (JCM 9498^T), *Bacteroides finegoldii* (JCM 13345^T), *Bacteroides intestinalis* (JCM 13265^T), *Bacteroides ovatus* (JCM 5824^T), *Bacteroides stercoris* (JCM 9496^T), *Parabacteroides johnsonii* (JCM 13406^T) and *Prevotella copri* (JCM 13464^T) were obtained from ATCC and JCM. Previously described regulatory T-cell-inducing strains[37] from our laboratory were also included in the commensal panel: *Clostridium symbiosum* (VE202-16), *Clostridium ramosum* (VE202-18), *Clostridium bolteae* (VE202-7) and *Flavinofractor plautii* (VE202-3). In addition, *Hungatella hathewayi*, *E. rectale* and *A. putredinis* isolated from human faeces in our laboratory were used. *Clostridium* HGF2 (*innocuum*) HM287 was obtained through BEI Resources, NIAID, NIH as part of the Human Microbiome Project: *Clostridium* sp., strain HGF2, HM-287.

From fresh colonies grown on BHK blood agar plates (Kyokuto), a primary suspension adjusted to an OD_{600} of 0.63 was prepared in WCA or BHI medium (the composition of media is presented in Supplementary Table 9). Subsequently, the secondary suspension was prepared by diluting 100 µl of the primary suspension into a total of 2.4 ml of medium. Then, 10 µl of the secondary suspension was inoculated to a total of 200 µl of medium containing varying concentrations (3.175, 6.25, 12.5, 25 or 50 µM) of bile acids; DCA, LCA, 3-oxoLCA, 3-oxoalloLCA, isoLCA, alloLCA or isoalloLCA. The growth of bacteria was monitored every 0.5–1 h by OD_{600} measurement using a microplate reader (Sunrise Thermo, Tecan) set at 37 °C with 60 s shaking before each time point and a PLATEmanager v.5/S software for the data collection. To determine the MIC, 10 µl of the secondary suspension was inoculated into a total of 200 µl of medium containing 0.25–50 µM isoalloLCA.

Electron microscopy

Bacterial cultures incubated with or without isoalloLCA were collected after incubation for 5 h for electron microscopy samples. For scanning electron microscopy, 10–30 µL of culture was spotted on a Nano Percolator membrane (JEOL) and fixed in freshly prepared 2.5% glutaraldehyde solution. After overnight fixation at 4 °C, samples were washed in 0.1 M phosphate buffer (pH 7.4, Muto Pure Chemicals), fixed with 1.0% osmium tetroxide (TAAB Laboratories) for 2 h at 4 °C, and treated with a series of increasing concentrations of ethanol. Samples were dried up with a critical point dryer (CPD300, Leica Biosystems) and coated with about 2 nm thickness of osmium using a conductive

osmium coater (Neoc-ST, Meiwafosis). Scanning electron microscopy images were acquired using the SU6600 (Hitachi High Tech) at an electron voltage of 5 keV.

For transmission electron microscopy, microbial pellets were prepared by centrifugation (13,000 rpm, 2 min) from 25 ml bacterial cultures. Pellets were fixed with 2.5% glutaraldehyde solution overnight at 4 °C. After washing with 0.1 M phosphate buffer, samples were fixed with 1.0% osmium tetroxide for 2 h at 4 °C, washed in distilled water and embedded into low-gelling-temperature Type VII-A agarose (Sigma-Aldrich). Samples were dehydrated by a series of increasing concentrations of ethanol to absolute ethanol, soaked with acetone (Sigma-Aldrich), with *n*-butyl glycidyl ether (Okenshoji), graded concentration of Epoxy resin with *n*-butyl glycidyl ether and also with 100% Epoxy resin (100 g Epon was composed of 27.0 g MNA, 51.3 g EPOK-812, 21.9 g DDSA and 1.1 ml DMP-30, all from Okenshoji) for 48 h at 4 °C. Polymerization of pure Epoxy resin was completed for 72 h at 60 °C. The ultrathin sections (70 nm) were prepared on copper grids (Veco Specimen Grids, Nisshin-EM) with an ultramicrotome (Leica UC7, Leica Biosystems) and stained with uranyl acetate and lead citrate for 10 min each. Transmission electron microscopy images were obtained using the JEM-1400plus (JEOL) at an electron voltage of 80–100 keV.

Culturing human faeces with bile acids

Human faecal culture was conducted in 96-deep-well plates (Treff Lab) using stool samples obtained from young and healthy donors (filtered and resuspended in 20% glycerol for cryo-preservation). In brief, 5 mg of stool was inoculated in 1 ml of WCA medium supplemented with 4% salt solution (0.2 g l^{-1} calcium chloride, 0.2 g l^{-1} magnesium sulfate, 1 g l^{-1} dipotassium hydrogen phosphate, 1 g l^{-1} potassium dihydrogen phosphate, 10 g l^{-1} sodium hydrogen carbonate and 2 g l^{-1} sodium chloride), ammonium chloride (1.0 g l^{-1}), L-cysteine (1.0 g l^{-1}), vitamin K (0.5 mg l^{-1}), haemin (5 mg l^{-1}), sodium acetate (1.0 g l^{-1}), sodium formate (0.15 g l^{-1}), sodium fumarate (0.15 g l^{-1}), sodium thioglycolate (0.3 g l^{-1}), 1% ATCC vitamin solution and 1% ATCC Trace element solution based on media that have previously been used for human faecal batch culture[38,39]. Faecal cultures with a final concentration of 50 μM bile acids (LCA, 3-oxoLCA or isoalloLCA) were incubated anaerobically for 48 h at 37 °C. DNA was extracted from the faecal sample culture for 16S metagenomic sequencing as described above.

Animal experiments

Germ-free female C57BL/6N mice were purchased from CLEA Japan or bred and maintained within the gnotobiotic facilities of JSR-Keio University Medical and Chemical Innovation Center. All animal experiments were approved by the Keio University Institutional Animal Care and Use Committee. All animals were maintained on a 12-h light–dark cycle and received gamma-irradiated pellet food (50 kGy radiated CL2, CLEA Japan). For quantification of microbial bile acid metabolism, an overnight bacterial culture (200 μl) was orally administered to 8-week-old germ-free female mice and the mice were switched from pellet food to powder diet (50 kGy radiated AIN-93G with bacteria-free mineral acid casein; Research Diets) with or without 0.1% 3-oxo-Δ^4-LCA. Faeces samples were collected during the experiment to quantify the amount of bile acids using a SCIEX Triple Quad 6500+ system. The same method for human faecal sample preparation was applied to mouse faecal samples (diluted 20 times in 1× PBS).

For *C. difficile* infection, SPF C57BL/6N mice and *Cyp2c70*$^{-/-}$*Cyp2a12*$^{-/-}$ mice (C57BL/6J background)[21] were made susceptible to colonization by administration of cefoperazone (0.4–5 g l^{-1}) in drinking water for 4–5 days. After 2 days of pause from the antibiotic, mice were challenged orally with 6,000 *C. difficile* 630 spores. Infected mice were given an 0.1% isoalloLCA-containing diet or a daily administration of Odoribacteraceae St21 (200 μl of overnight bacterial culture) after 1 day of infection. Germ-free C57BL/6N mice colonized with the faecal microbiota from a patient with a *C. difficile* infection were obtained by administration of 200 mg μl^{-1} of faecal slurry followed by an 0.1% isoalloLCA-containing diet from the next day. The abundance of *C. difficile* was quantified by CFU after anaerobic culture of diluted faecal samples on BD BBL S-CCFA (Becton Dickinson)-selective agar plates at 37 °C for 24 h. Odoribacteraceae St21 was quantified by qPCR from DNA extracted from faeces (5′-ACGTGGACCATCAGTGAACT-3′, 5′-AGTTCTCCAAGTCCCTCTGC-3′).

Chemical synthesis of 3-oxo-Δ^4-LCA and isoalloLCA

3-oxo-Δ^4-LCA was prepared from 5 g of LCA through the following synthesis reactions. First, 35 ml of methanol and 1 l of HCl-infused methanol were added to LCA in a 300-ml round-bottom flask. The suspension was then stirred at 100 °C for 1 h for methyl esterification of the carboxylic arm of LCA. After cooling down the reaction mixture, the methanol was evaporated. The slurry was dissolved in 300 ml of chloroform and transferred to a separatory funnel containing 50 ml water. The layers were vigorously mixed and separated, and the aqueous layer containing HCl was discarded (3 × 50 ml). The combined organic layers were washed with saturated aqueous NaCl, dehydrated over Na_2SO_4 for 15 min and filtered, and the chloroform was evaporated. To generate methylated 3-oxoLCA, crystals were dissolved in 40 ml of acetone and then reacted dropwise with 3.5 ml of Jone's reagent (CrO_3 in aqueous H_2SO_4) while cooling with ice water. Then, 2 ml methanol was added to reduce toxic Cr^{6+} to Cr^{3+}. The reaction mixture was slowly poured into 600 ml of stirred ice water to crystalize the synthesized compound, then rinsed and collected by filtration. Crystals were dissolved in 300 ml of ethyl acetate and transferred to a separatory funnel with 10 ml of water to clean up H_2SO_4 and residual chromium. The organic layers were collected, washed and dehydrated as described above, followed by evaporation of ethyl acetate. To further purify the methylated 3-oxoLCA, the reaction product was dissolved in 1 l of toluene and 100 ml of ethyl acetate for fractionation by silica-based liquid chromatography. Upon verification of fractions containing the target compound by thin-layer chromatography, fractions were combined and the eluent was evaporated.

To brominate the C4-carbon atom on steroid ring A, 30 ml of acetic acid was added to the methylated 3-oxoLCA and dropwise reacted with the same molar amount of bromine solution. The dark-blue mixture was slowly poured into 600 ml of stirred ice water to crystalize the brominated compounds (yellowish crystalline solid), after which it was rinsed and collected by filtration. Crystals were dissolved in 100 ml of chloroform and transferred to a separatory funnel with 10 ml of water for liquid–liquid extraction, followed by evaporation of the chloroform. The compound with bromide at the C4 carbon was purified by silica-based liquid chromatography (toluene:ethyl acetate, 49:1). Collected fractions were combined and the eluent was evaporated. For the debromination reaction to generate the C4–C5 double bond, 10 ml of *N*,*N*-dimethylformamide was added to the crystals, followed by 3.5× the amount of $LiCO_3$ and LiBr. The mixture was stirred and heated at 100 °C overnight. Methylated 3-oxo-Δ^4-LCA was extracted by adding 50 ml of chloroform and filtering out the precipitates. The organic solution was cleaned up by liquid–liquid extraction, followed by the evaporation of the *N*,*N*-dimethylformamide and chloroform. Methylated 3-oxo-Δ^4-LCA was further purified by silica-based liquid chromatography (toluene:ethyl acetate, 10:1).

Methylated 3-oxo-Δ^4-LCA was also used for isoalloLCA synthesis. In brief, 3.6 g methylated 3-oxo-Δ^4-LCA was dissolved in 5 ml dehydrated pyridine, mixed with 1.9 g hexamethyldisilazane and 3.6 g bromotrimethylsilane for 40 min while cooling. The reaction solution was cleaned up by liquid–liquid extraction followed by solvent evaporation. The crude product was dissolved in 100 ml of isopropanol and then mixed with 780 mg sodium borohydride ($NaBH_4$) for 4 h. The reaction was quenched with acetic acid and evaporated, extracted with ethyl acetate, and cleaned up with liquid–liquid extraction. The reaction product (methylated Δ5-3β-ol) was purified by silica-based

liquid chromatography (benzene:ethyl acetate, 5:1). To generate methylated isoalloLCA by hydorgenation, methylated Δ5-3β-ol was dissolved in 20 ml ethyl acetate, followed by carefully adding 300 mg of 10% palladium-activated carbon and hydrogen gas. After 2 h of reaction, the mixture was filtered and methylated isoalloLCA was purified by silica-based liquid chromatography (benzene:ethyl acetate, 4:1). Finally, purified methylated 3-oxo-Δ^4-LCA and methylated isoalloLCA were mixed with the same molar amount of KOH (2 M solution in 90% methanol) at 50 °C for 1 h, and then with 2 M HCl to neutralize the solution overnight to remove the methyl group for the synthesis of 3-oxo-Δ^4-LCA and isoalloLCA. Crystals were dissolved in chloroform for the final clean-up by liquid–liquid extraction, followed by evaporation of chloroform. The purity of 3-oxo-Δ^4-LCA and isoalloLCA was determined by thin-layer chromatography and LC-MS/MS. 3-Oxo-Δ^4-LCA and isoalloLCA were also generated by Sundia MediTech.

Statistical analysis

A pairwise Wilcoxon rank-sum test (nonparametric) was used to evaluate the differences in the relative abundance of *bai* operon homologues and meta-16S rRNA analysis between different age groups. A Spearman's rank correlation was used to evaluate trends between the relative abundance of Bacteroidales species encoding 5AR, 5BR, 3β-HSDH-I or 3β-HSDH-II genes and the abundance of the secondary bile acids in stool samples. Overall nominal *P* values were adjusted for multiple testing using Benjamini–Hochberg correction and associations with FDR-adjusted $P < 0.05$ (unless stated differently) were considered to be significant. Statistical analyses were performed using GraphPad Prism software (GraphPad Software) and RStudio. One-way ANOVA with Tukey's test (parametric) and Kruskal–Wallis with Dunn's test (nonparametric) were used for multiple comparisons. A Mann–Whitney *U*-test (two-tailed) with Welch's correction (nonparametric) was used for all comparisons between two groups in the in vivo bile acid metabolism, co-culture pathogen inhibition, in vivo *C. difficile* inhibition experiments.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Shotgun sequencing data are deposited in NCBI under BioProject PRJNA675598. Genome sequences of the 68 strains isolated from a centenarian and 16S rRNA amplicon sequence data are deposited in the DNA Data Bank of Japan under BioProject PRJDB11902 and PRJDB11894, respectively. LC–MS/MS data are deposited in Metabolomics Workbench (https://www.metabolomicsworkbench.org/) under project ID PR001168 with study ID ST001851 for human faeces data and study ID ST001852 for in vitro data. Source data are provided with this paper.

Code availability

Code for all of the analyses is available on GitHub (https://gitlab.com/xavier-lab-computation/public/centenarianmicrobiome).

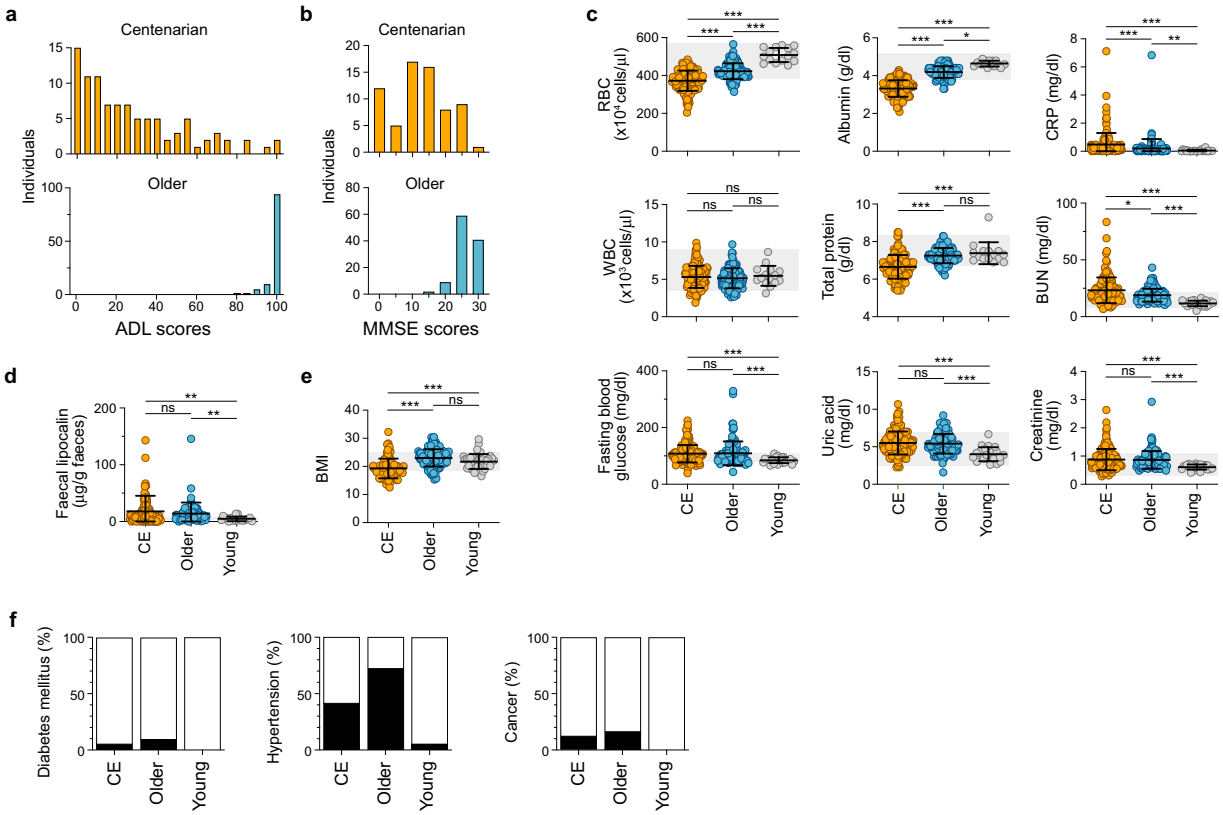
23. Atarashi, K. et al. Ectopic colonization of oral bacteria in the intestine drives T_H1 cell induction and inflammation. *Science* **358**, 359–365 (2017).
24. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
25. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
26. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
27. Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
28. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
29. Plaza Oñate, F. et al. MSPminer: abundance-based reconstitution of microbial pan-genomes from shotgun metagenomic data. *Bioinformatics* **35**, 1544–1552 (2019).
30. Li, J. et al. An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
31. Segata, N., Börnigen, D., Morgan, X. C. & Huttenhower, C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* **4**, 2304 (2013).
32. Khelaifia, S., Raoult, D. & Drancourt, M. A versatile medium for cultivating methanogenic archaea. *PLoS ONE* **8**, e61563 (2013).
33. Aziz, R. K. et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).
34. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
35. García-Bayona, L. & Comstock, L. E. Streamlined genetic manipulation of diverse *Bacteroides* and *Parabacteroides* isolates from the human gut microbiota. *mBio* **10**, e01762-19 (2019).
36. Ferrières, L. et al. Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J. Bacteriol.* **192**, 6418–6427 (2010).
37. Atarashi, K. et al. T_{reg} induction by a rationally selected mixture of *Clostridia* strains from the human microbiota. *Nature* **500**, 232–236 (2013).
38. Quinn, R. A. et al. Global chemical effects of the microbiome include new bile-acid conjugations. *Nature* **579**, 123–129 (2020).
39. McDonald, J. A. K. et al. Evaluation of microbial community reproducibility, stability and composition in a human distal gut chemostat model. *J. Microbiol. Methods* **95**, 167–174 (2013).
40. Claesson, M. J. et al. Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178–184 (2012).

Acknowledgements We thank L. Besse for project management and making data available through the SRA and the Broad Institute Genomics Platform and Microbial 'Omics Core for sample processing and sequencing data generation; A. Minowa, M. Asakawa, K. Sugita and members of the JSR-Keio University Medical and Chemical Innovation Center for their technical support; N. Hasegawa, K. Fukunaga, T. Kanai, K. Masaki, S. Azekawa, K. Sasahara, S. Hosomi, M. Shimura and Y. Abe for their assistance in collecting clinical samples; S. Atsushi and members of the Honda laboratory for their suggestions during the course of this studies. K.H. was funded through AMED LEAP under grant number JP20gm0010003, Grant-in-Aid for Specially Promoted Research from JSPS (no. 20H05627), Public/Private R&D Investment Strategic Expansion Program (PRISM) from Cabinet Office of the Government of Japan, the Naito Foundation and the Takeda Science Foundation. The centenarian study was funded by the Japan Ministry of Agriculture, Forestry and Fisheries (M.S.) and Keio Global Research Institute (Y.A. and N.H.). The older-participant recruitment study was funded by a Grant-in-Aid for Scientific Research (no. 18H03055) from the Japan Society for the Promotion of Science and JST Research Complex Program (JP15667051). D.R.P. and R.J.X. were funded by the Center for the Study of Inflammatory Bowel Disease (DK043351), Center for Microbiome informatics and Therapeutics (CMIT) and AT009708. Y. Sato. was supported by the Terumo Life Science Foundation.

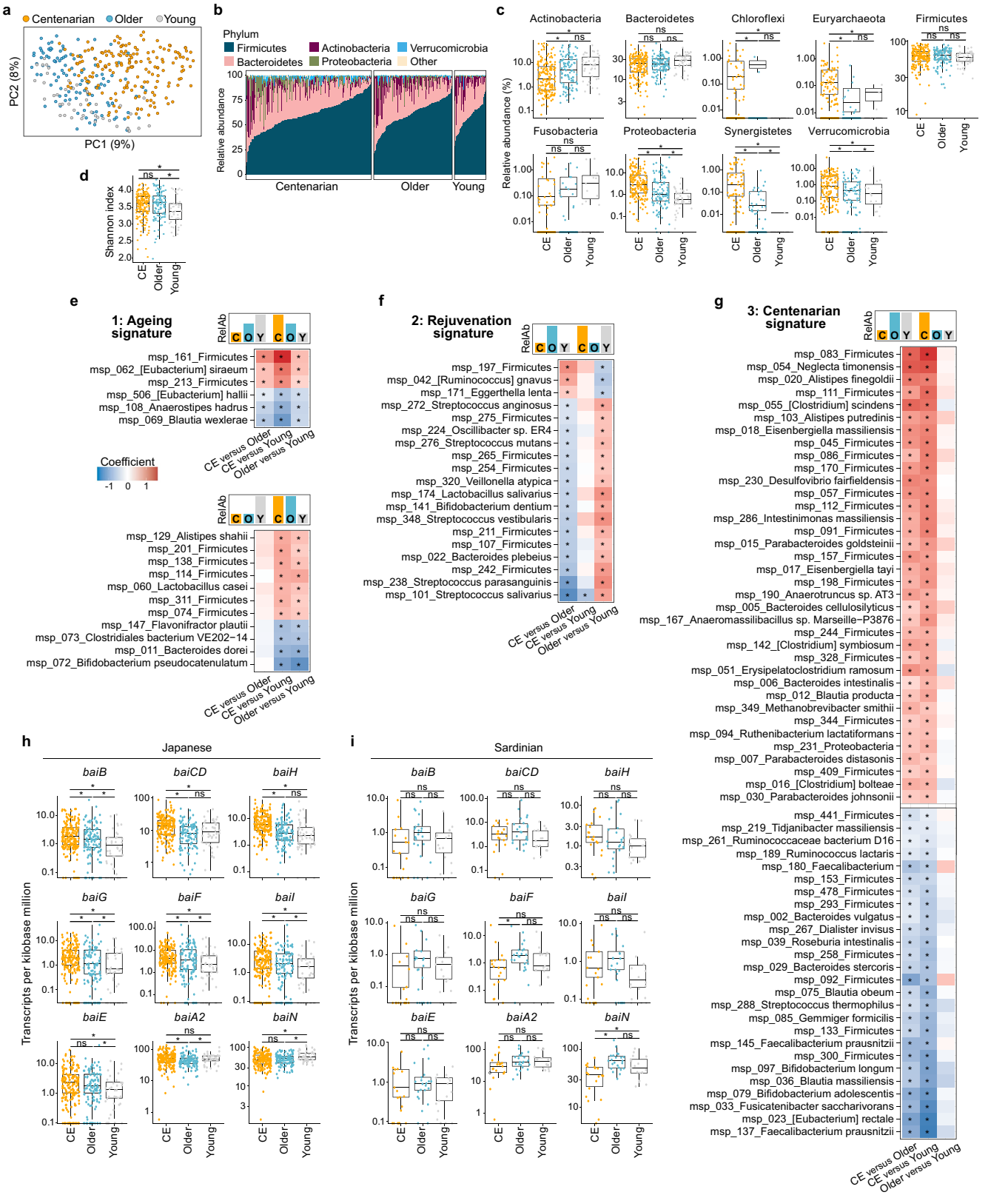
Author contributions K.H., Y. Sato and K.A. planned experiments, analysed data and wrote the paper together with D.R.P., R.J.X., A.N.S. and S.M.K.; Y.A., K.T. and N.H. collected clinical samples; Y. Sato and K.A. performed bacterial experiments; D.R.P., H.V. and R.J.X. performed metagenome analyses; K.A., S.M.K., Y.O., W.S. and M.H. performed whole-genome and meta-16S rRNA gene sequencing and analyses; S. Sasajima, Y. Sato, K.A., H.T., H.N., S.N., Y. Sugiura and M.S. performed metabolomic analysis; T.S., S.O., S. Sasajima and T.M. synthesized chemical compounds; N.M. and S. Shibata. performed electron microscopy imaging; A.H., Y.U., T.I., Y.L., T.T., J.I., H.I. and K.M. provided essential materials; D.R.L. and M.A.F. supervised bacterial experiments.

Competing interests K.H. is a scientific advisory board member of Vedanta Biosciences and 4BIO CAPITAL.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03832-5.
Correspondence and requests for materials should be addressed to R.J.X. or K.H.
Peer review information *Nature* thanks Pieter Dorrestein, Andrew Patterson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.

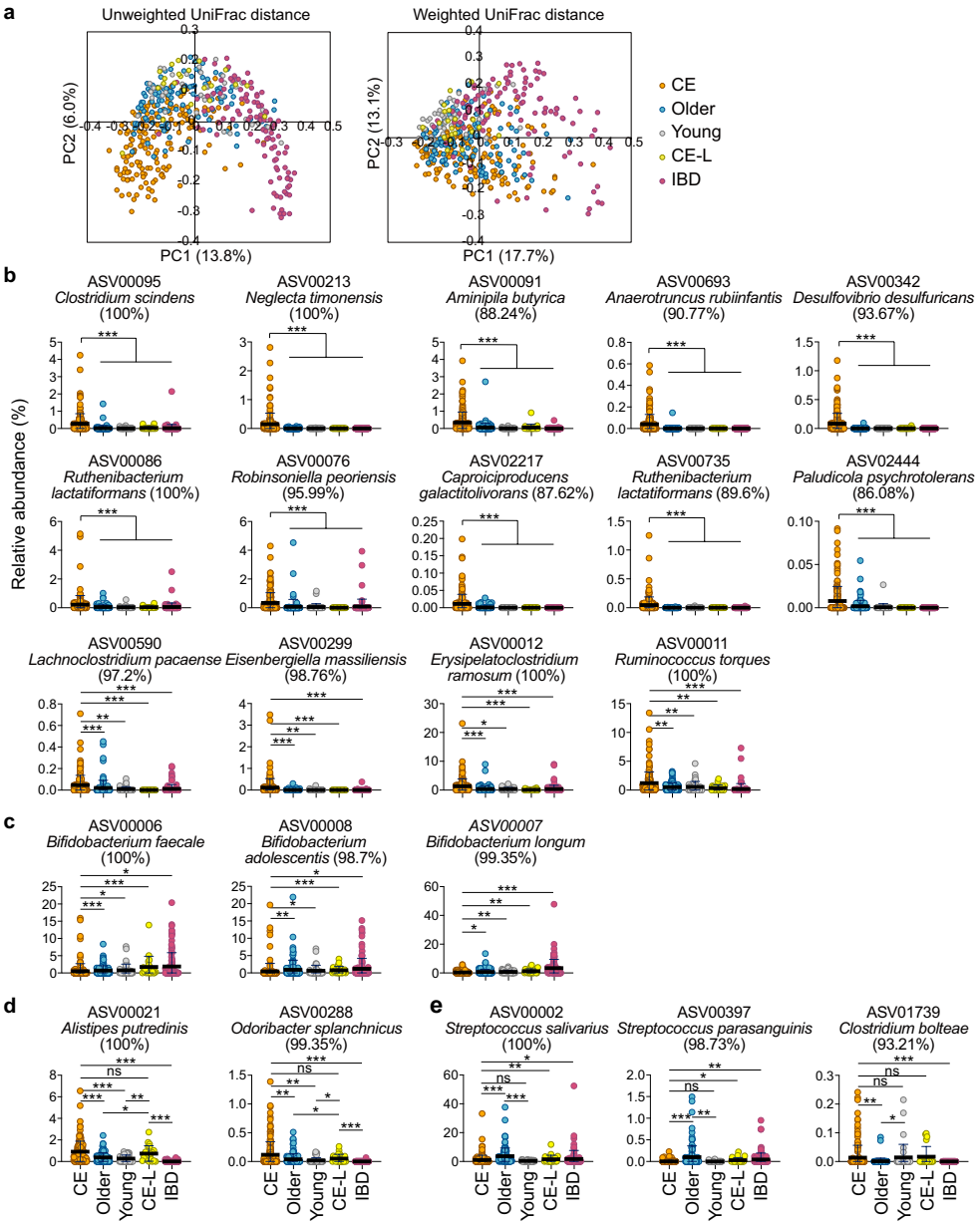


Extended Data Fig. 1 | Clinical characteristics of centenarians. **a**, **b**, Activities of daily living (ADL) scores assessed via the Barthel index (**a**) and mini-mental state exam (MMSE) scores (**b**) of centenarian (top, ADL, $n = 96$; MMSE, $n = 67$) and older (bottom, ADL, $n = 111$; MMSE, $n = 111$) groups. **c**, Blood tests from centenarian (CE, $n = 146$-150, orange), older ($n = 111$, blue), and young ($n = 15$-39, grey) participants; red blood cell (RBC) count; albumin; C-reactive protein (CRP); white blood cell (WBC) count; total protein; blood urea nitrogen (BUN); fasting blood glucose; uric acid; and creatinine. **d**, Faecal lipocalin level of centenarian ($n = 122$), older ($n = 67$), and young ($n = 19$) individuals quantified by ELISA. **e**, Body-mass-index (BMI) values of centenarian ($n = 89$), older ($n = 111$), and young ($n = 39$) individuals. **f**, Percentages of individuals with medical histories of diabetes mellitus, hypertension, or cancer. Black colour represents affected individuals. Medical history was surveyed from 155-156 centenarian, 110-111 older, and 39 young individuals. In **c**-**e**, data are mean ± s.d. $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; Kruskal–Wallis with Dunn's test. ns, not significant. Grey areas within graphs are ranges of normal values.

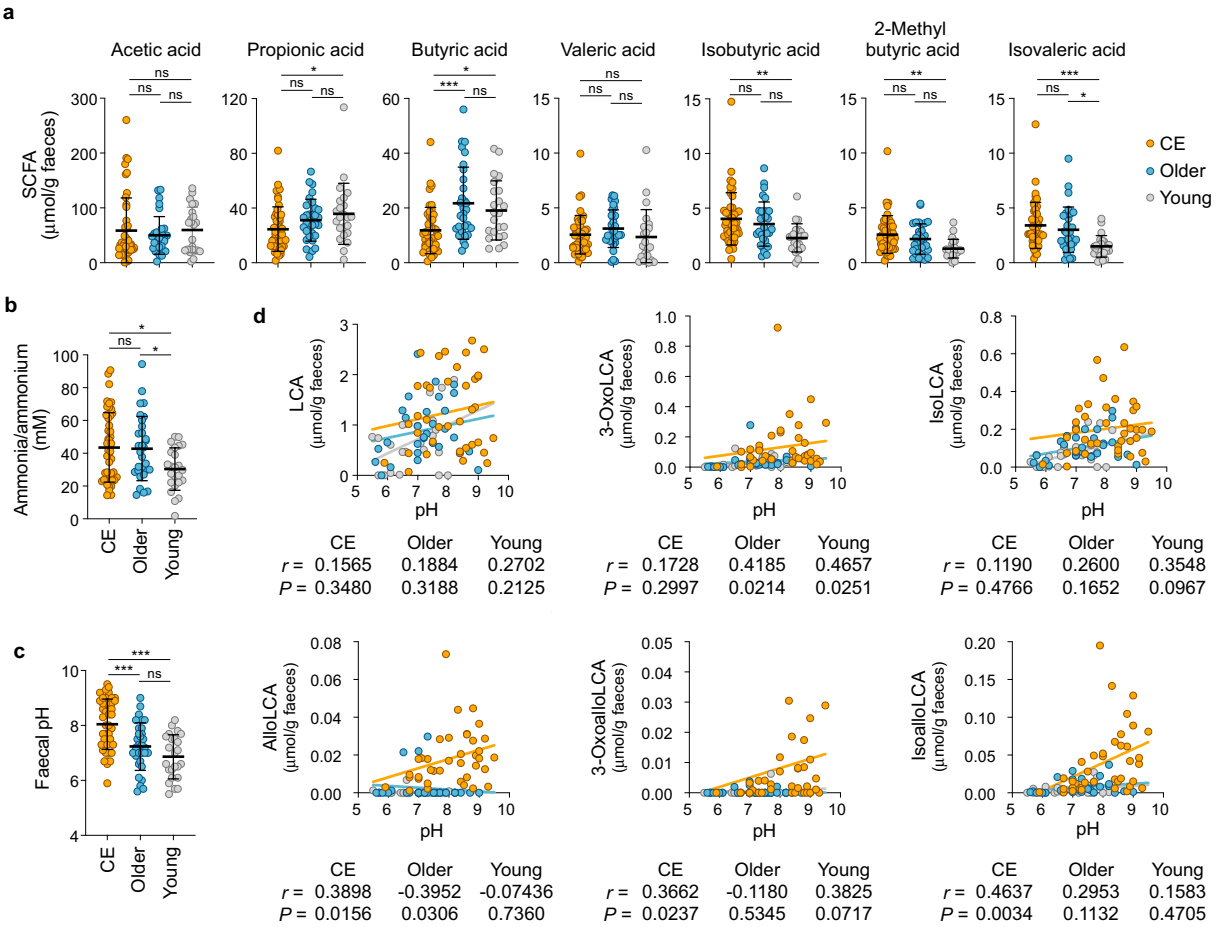


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Gut microbiome signatures in centenarian, older and young Japanese participants based on whole-metagenome shotgun sequencing and de novo assembly analysis. a, Principal coordinate analysis based on species-level Bray–Curtis dissimilarity from the assembled faecal metagenomes of centenarian [CE, $n = 176$, orange (154 individuals: 3 participants undergoing antibiotic treatment and 3 with insufficient bacterial DNA yield were excluded from the total 160 participants; analysis also included samples collected twice from 20 individuals and three times from 1 individual with an intervening one or two year interval)], older [$n = 110$, blue (110 individuals: 1 participant undergoing antibiotics treatment and 1 participant with insufficient bacterial DNA yield were excluded from 112 participants)], and young [$n = 44$, grey (44 individuals: 3 participants with insufficient bacterial DNA yield were excluded from 47 participants]. The microbiota composition of centenarians was significantly different from that of both control groups (PERMANOVA false-discovery rate (FDR)-adjusted $P < 0.05$). **b**, **c**, Relative abundance across phyla. **d**, Shannon diversity index (*$P < 0.05$, linear model). **e**-**g**, Changes in the relative abundance (RelAb) of gut metagenome species (MSPs) between centenarian, older, and young participants grouped according to the following signatures of differential abundance: 1) ageing signature, 2) rejuvenation signature, and 3) centenarian signature. Each signature is accompanied by models depicting microbial relative abundance patterns in centenarian (C), older (O), and young (Y) groups that would fall into the given signature. Colour scale represents the coefficient from the linear model and indicates enrichment (red) or depletion (blue) of a species in the respective comparisons: centenarian compared to older, centenarian compared to young, and older compared to young; in each case, the latter group is used as a reference in the model. Differentially abundant species that are significant at FDR $P < 0.05$ are indicated with asterisks. The first signature ('ageing signature') included taxa whose abundance was increased or decreased with age (**e**). For example, *Eubacterium siraeum* and undefined Firmicutes species (msp_161, 213) were most abundant in centenarians, followed by the older and then the young controls, whereas *Blautia wexlerae* displayed the opposite trend, being most abundant in young controls, followed by the older participants, and finally the centenarians. These findings are in alignment with previous studies that suggest the relative abundances of these taxa reflect adaptation to ageing, and may be related to physical activity and diet[3,40]. The second signature ('rejuvenation signature') included taxa whose abundance was similar in centenarians and young controls, but distinct from the older participants (**f**). These species might reflect the maintenance of youth or possess reverse-ageing effects. Notably, *R. gnavus* and *E. lenta* were part of this signature, as they were comparably abundant in both centenarians and young controls. Notably, these species have been implicated in bile acid metabolism, particularly the biosynthesis of iso-bile acids[11]. The third signature ('centenarian signature') included centenarian-specific taxa whose abundance was significantly different between centenarians and both the older and young control groups, but not between these two control groups (**g**). In the third signature, *Alistipes*, *Parabacteroides*, *Bacteroides* and *Clostridium* species, as well as *Methanobrevibacter*, a predominant archaeon in the human gut, were specifically enriched in centenarians compared with the other groups. One of the most abundant species in centenarians was *C. scindens*. By contrast, key butyrate producers such as *F. prausnitzii* and *E. rectale* were selectively depleted in centenarians. **h**, **i**, Abundance of genes homologous to the *C. scindens bai* operon in Japanese (**h**) and Sardinian (**i**) centenarian, older, and young age groups. Sardinian centenarians ($n = 19$), older controls ($n = 23$), and young controls ($n = 17$) from the European Nucleotide Archive (accession number PRJEB25514). **c**, **d**, **h**, **i**, Horizontal lines indicate the median; box boundaries indicate the interquartile range; whiskers represent values within 1.5 × the interquartile range of the first and third quartiles. Each circle represents one sample. In **c**, **e**-**i**, *FDR-adjusted $P < 0.05$; Wilcoxon rank-sum test. ns, not significant.

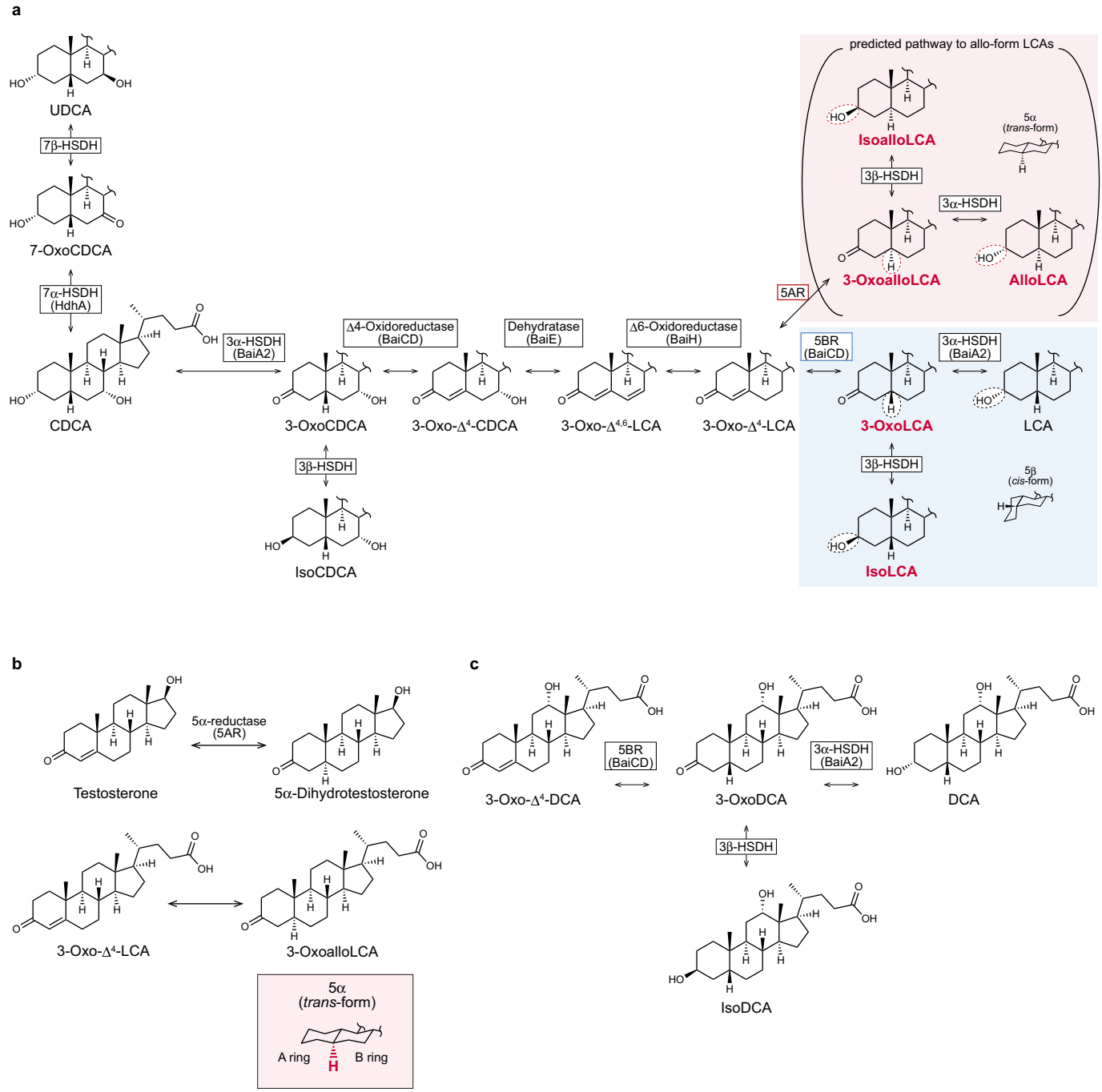


Extended Data Fig. 3 | The microbiota compositions of centenarians, their lineal descendants and patients with IBD. Faecal meta 16S rRNA gene sequencing of centenarians (CE, $n = 157$), older participants ($n = 111$), young participants ($n = 40$), lineal relatives of centenarians (CE-L, $n = 22$, 48-95 years old, average 74.7 years old), and patients with IBD (Crohn's disease; $n = 12$, ulcerative colitis; $n = 91$, 15-78 years old, average 49.0 years old) (all Japanese). **a**, PCoA plots based on unweighted UniFrac distance among faecal microbiome of centenarian (orange), older (blue), young (grey), CE-L (yellow), and IBD (red) groups. Note that the microbiota composition of centenarians was distinct from that of patients with IBD. **b**-**e**, Relative abundance of ASV (amplicon sequence variants) that are significantly enriched in centenarians (**b**); depleted in centenarians (**c**); commonly enriched in both centenarians and their lineal relatives (**d**); and commonly depleted or enriched in both centenarians and young participants (**e**). The closest species of ASVs were assigned using the National Center for Biotechnology Information Reference Sequence (NCBI RefSeq) database. Data are mean ± s.d. $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; Wilcoxon rank-sum test. ns, not significant. Each dot represents an individual.

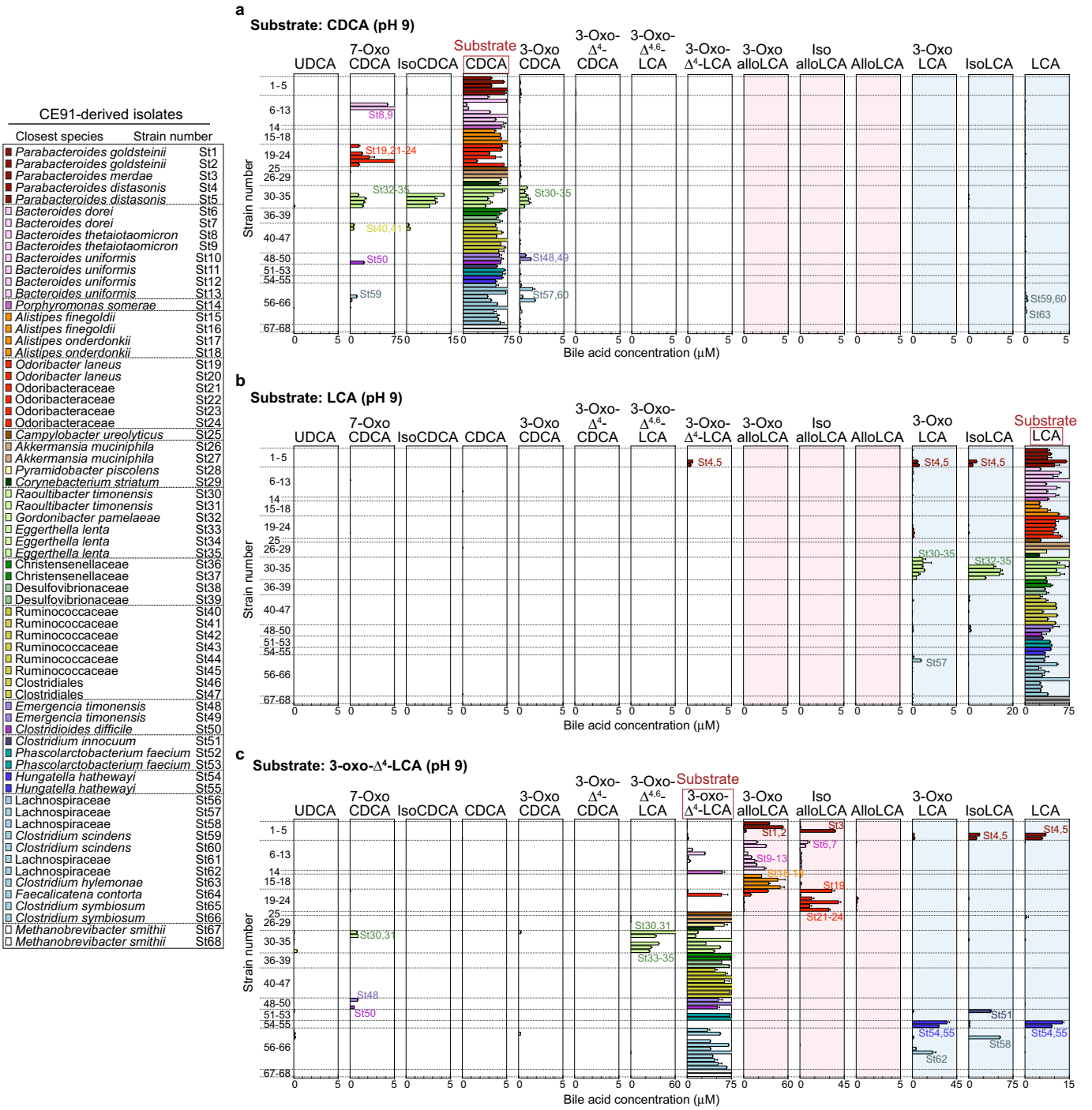


Extended Data Fig. 4 | Faecal short-chain fatty acids and pH in centenarians. a, GC-MS-based quantification of faecal short-chain fatty acids (SCFAs) from centenarian (CE, $n = 47$, orange), older ($n = 31$, blue), and young ($n = 23$, grey) individuals. Faecal SCFAs are shown in µmol/g wet weight faeces. **b**, **c**, Faecal ammonia/ammonium (**b**) and faecal pH levels (**c**) from individuals within each group. **d**, Correlation between the levels of secondary bile acids (LCA, 3-oxoLCA, isoLCA, alloLCA, 3-oxoalloLCA, and isoalloLCA) and pH. Each circle represents an individual. Spearman's coefficient (r) and significance (P) were calculated separately for each group. There is significant positive correlation between pH and faecal alloLCA ($P = 0.0156$), 3-oxoalloLCA ($P = 0.0237$), and isoalloLCA ($P = 0.0034$) in centenarians. In **a**-**c**, data are mean ± s.d. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; one-way ANOVA with Tukey's test. ns, not significant.

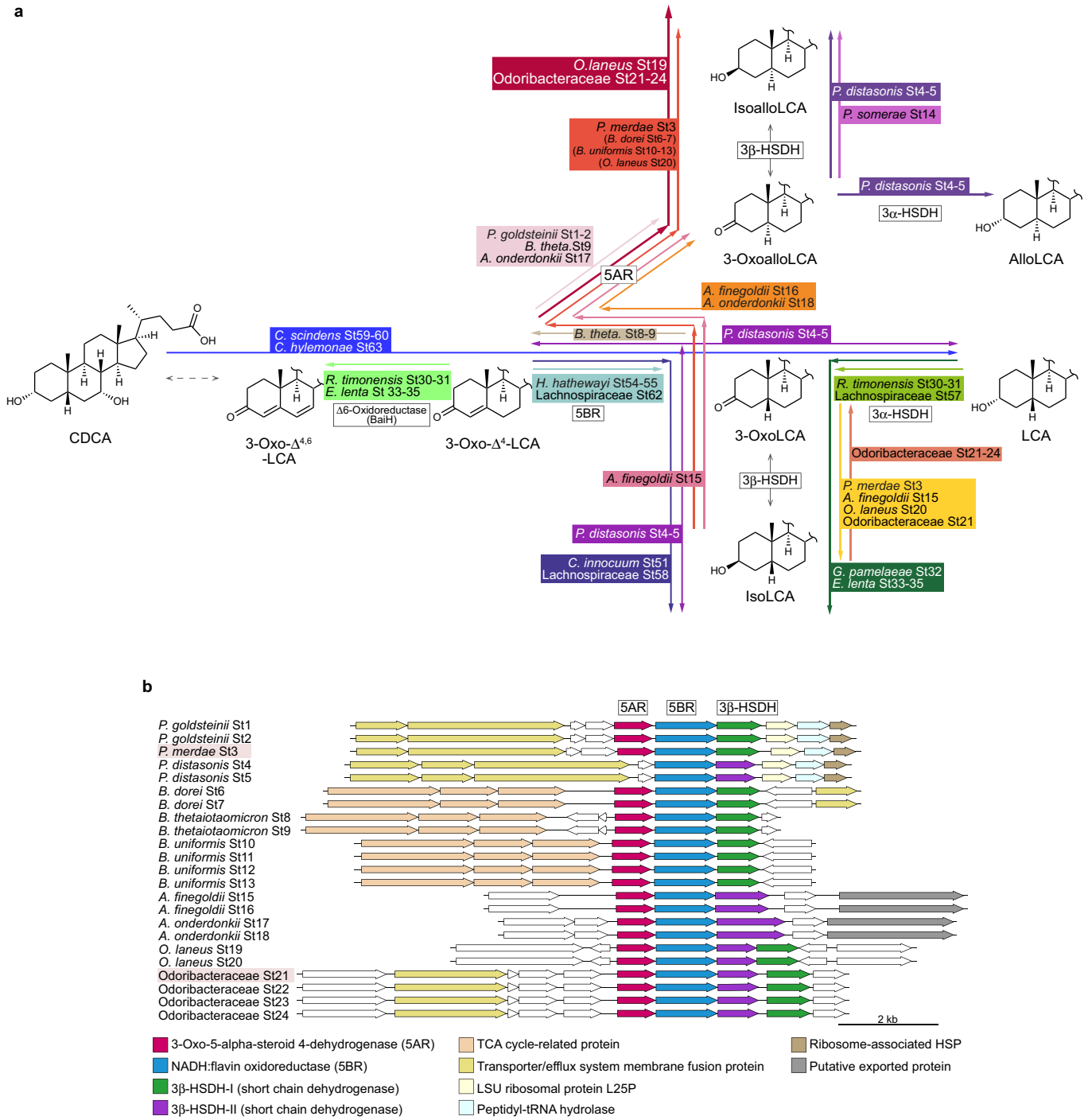
Extended Data Fig. 5 | Quantification of faecal bile acids by age group. Faecal bile acid composition of centenarian (CE, $n = 125$, orange), older ($n = 107$, blue), young ($n = 47$, grey), and lineal relatives of centenarians (CE-L, $n = 18$, yellow) were profiled and quantified by LC-MS/MS (µmol/g wet weight faeces). In pilot studies, we found that 94 of 137 examined bile acids were minor components of centenarians' faeces (see Supplementary Table 3). We thus selected the remaining 43 bile acid compounds for follow-up quantitative analysis (see also Fig. 1a). **a**, Multi-dimensional scaling plot using Spearman's correlation highlights differences among the four groups' bile acid profiles. Each circle represents an individual participant from the indicated age group. $P = 4.27 \times 10^{-9}$ for CE versus older; $P = 2.72 \times 10^{-12}$ for CE versus young; $P = 4.18 \times 10^{-6}$ for CE versus CE-L; $P = 0.00123$ for older versus young; Wilcoxon rank-sum test. **b**, **c**, Average ratio of total primary versus secondary (**b**) and CA- versus CDCA-derived bile acids (**c**). **d**, Sum of conjugated and unconjugated bile acids. **e**, Concentration of each individual bile acid. In **d**, **e**, each circle represents an individual. Data are mean ± s.d. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; one-way ANOVA with Tukey's test. ns, not significant. Faecal bile acids are shown in µmol/g wet weight faeces. **f**, Distribution of participants' faecal isoalloLCA concentrations. The median faecal isoalloLCA concentration in centenarians was 19.5 µM, meaning that 63 centenarian samples among 125 (50.4%) had >19.5 µM isoalloLCA. In contrast, only 17 older (15.9%) and 3 young (7.7%) participants were found to have >19.5 µM isoalloLCA.



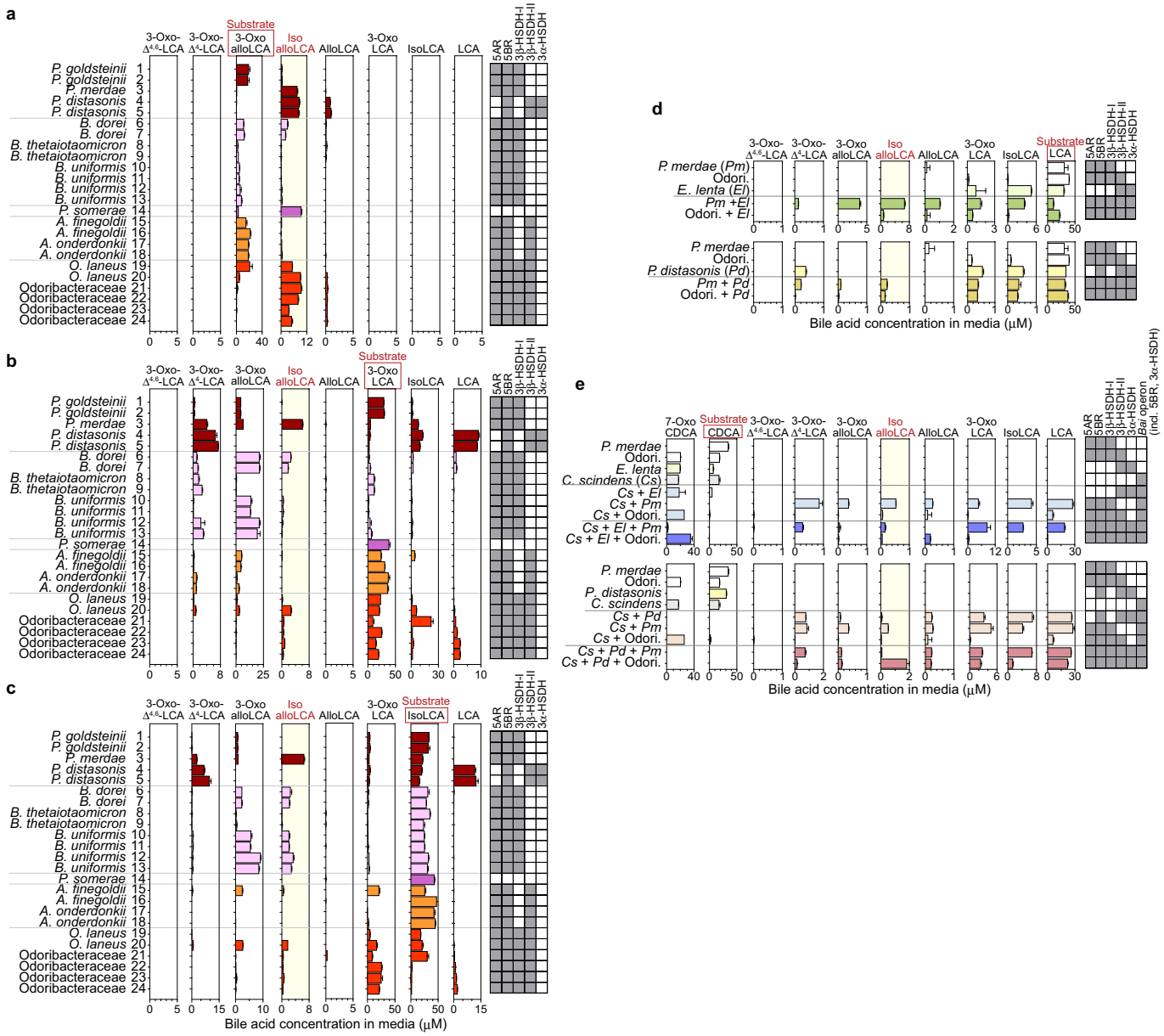
Extended Data Fig. 6 | Predicted bile acid biosynthesis by gut microorganisms. a, Biosynthetic pathway of secondary bile acids metabolized by the gut microbiota from primary bile acid chenodeoxycholic acid (CDCA). Responsible enzymes are indicated within boxes. The glycine or taurine conjugated primary bile acids are deconjugated (not depicted) and biotransformed into a variety of secondary bile acids by the gut microbiota. The predominant biotransformation is 7α-dehydroxylation of CDCA by *bai* operon genes, thereby converting it into lithocholic acid (LCA). In addition, bile acids can undergo oxidation and epimerization to generate oxo- (keto-), iso- (3β-hydroxy-), allo- (5α-H-), as well as *cis*- (indicated in a blue box) and *trans*-forms (indicated in a pink box). Brackets indicate predicted pathways to allo-form LCA production. Chemical structures are simplified by depicting only A and B steroid rings. **b**, 3-Oxo-Δ^4-LCA (also termed 3-oxo-4,5-dehydro-LCA) and 3-oxoalloLCA are structurally similar to testosterone and 5α-dihydrotestosterone (DHT), respectively. Both DHT and 3-oxoalloLCA have A and B steroid rings in a planar (*trans*) conformation (indicated in a pink box). We predicted that alloLCA and isoalloLCA might be generated from 3-oxo-Δ^4-LCA by the sequential action of a 5α-reductase (5AR) homologue and 3α-HSDH (for alloLCA) or 3β-HSDH (for isoalloLCA), through a 3-oxoalloLCA intermediate (see **a**), analogous to the 5AR-mediated conversion of testosterone into DHT by hydrogenating across the C4-C5 double bond, thereby forcing the A and B steroid rings into a planar conformation. **c**, Biosynthetic pathway of DCA and related bile acids by the gut microbiota based on ref. [11]. 3-OxoDCA can be generated from 3-oxo-Δ^4-DCA by hydrogenation across the C4-C5 double bond such that the C5 hydrogen is in the β position. This reaction is mediated by a 3-oxo-5β-steroid 4-dehydrogenase (also termed 5β-reductase, 5BR) encoded by the *BaiCD* gene. We predicted that LCA and isoLCA might be generated from 3-oxo-Δ^4-LCA by the sequential action of a 5BR homologue and 3α-HSDH (for LCA) or 3β-HSDH (for isoLCA), through a 3-oxoLCA intermediate (see panel **a**), mirroring the previously characterized conversion of 3-oxoDCA to DCA or isoDCA. In **a**–**c**, dashed-wedged lines indicate α-positions of -H and -OH groups, while bold-wedged lines indicate β-positions.



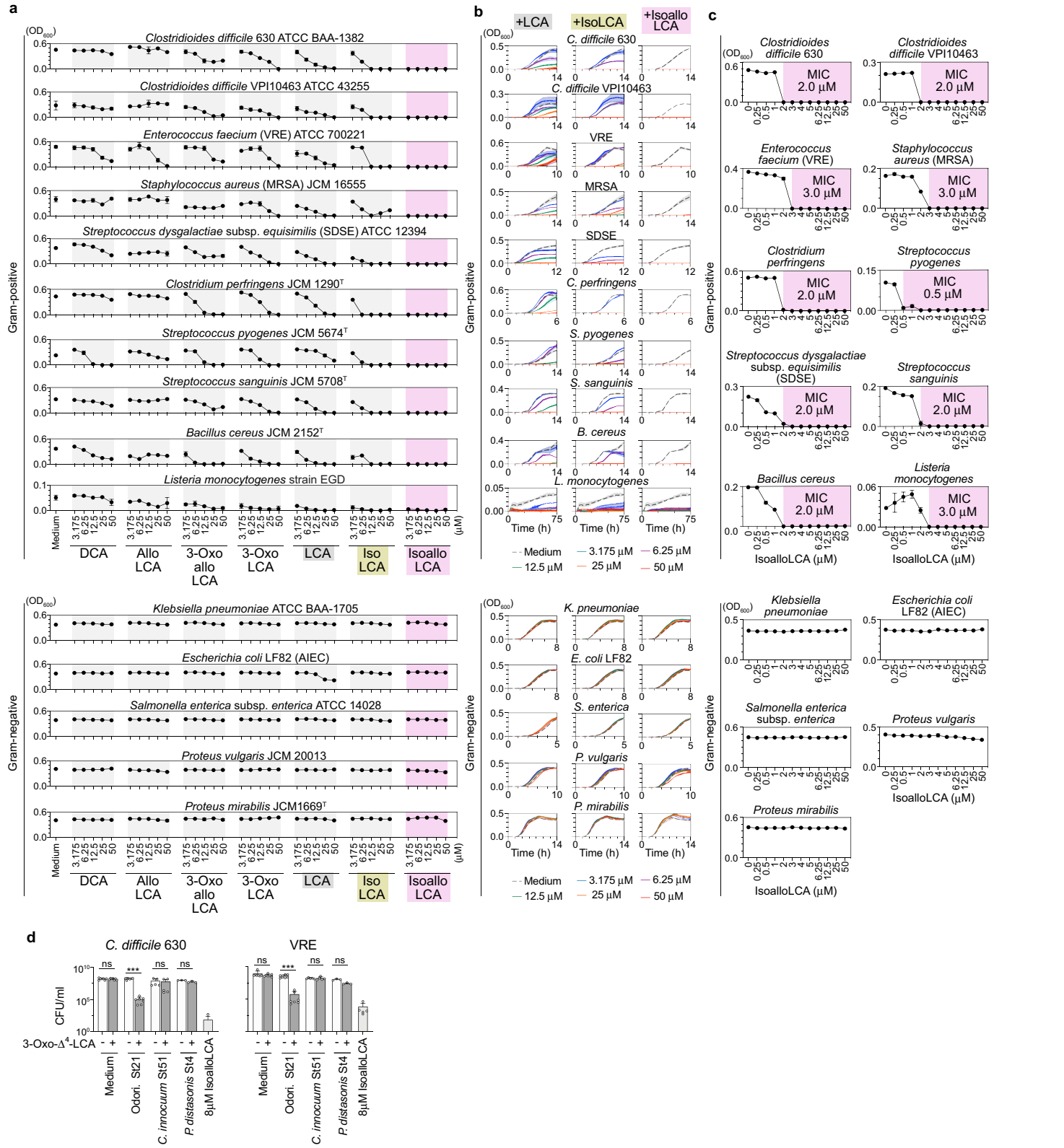
Extended Data Fig. 7 | Identification of bile acid-metabolizing bacterial strains isolated from the microbiota of a centenarian. In vitro bile acid metabolism by 68 CE91-derived isolates using 50 μM of CDCA (**a**), LCA (**b**), and 3-oxo-Δ^4-LCA (**c**) as starting substrates in pH 9-adjusted media. Data was obtained by LC-MS/MS analysis of 48 h culture supernatants. A list of the 68 isolated strains is shown in the left panel. The closest species to each isolate was identified based on 16S rRNA sequence similarity to the National Center for Biotechnology Information Reference Sequence (NCBI RefSeq) database. Graphs with red backgrounds indicate *trans*-bile acids and blue backgrounds indicate *cis*-bile acids. Bile acid profiles after culturing in pH 7-adjusted media are shown in Supplementary Fig. 1. **a**, Incubation with CDCA did not result in production of target bile acids in any of the cultures, though *C. scindens* strains 59-60 (St59-60) and *C. hylemonae* St63 were able to produce LCA, albeit at low levels. **b**, When cultured with LCA, *Gordonibacter pamelaeae* St32 and *E. lenta* St33-35 were found to produce 3-oxoLCA and isoLCA, implying carriage of 3α-HSDH and 3β-HSDH as predicted in a previous study[11]. In addition, *Raoulibacter timonensis* St30-31 and Lachnospiraceae spp. St57 were also capable of transforming LCA into 3-oxoLCA, suggesting their possession of 3α-HSDH, similar to *E. lenta*. **c**, When 3-oxo-Δ^4-LCA was used as a substrate, 3-oxoLCA accumulated to high levels in the supernatants of *H. hathewayi* St54-55 and Lachnospiraceae spp. St62 cultures, suggesting their carriage of 5BR. Similarly, isoLCA was generated from 3-oxo-Δ^4-LCA at high levels in *C. innocuum* St51 and Lachnospiraceae spp. St58 cultures, suggesting carriage of 5BR and 3β-HSDH. It is noteworthy that *P. distasonis* St4-5 converted LCA to 3-oxoLCA and further to 3-oxo-Δ^4-LCA, as well as 3-oxo-Δ^4-LCA to isoLCA and LCA, suggesting that these strains possess 3α-HSDH, 3β-HSDH, and 5BR. Data are mean + s.d. of duplicate samples and representative of two independent experiments.



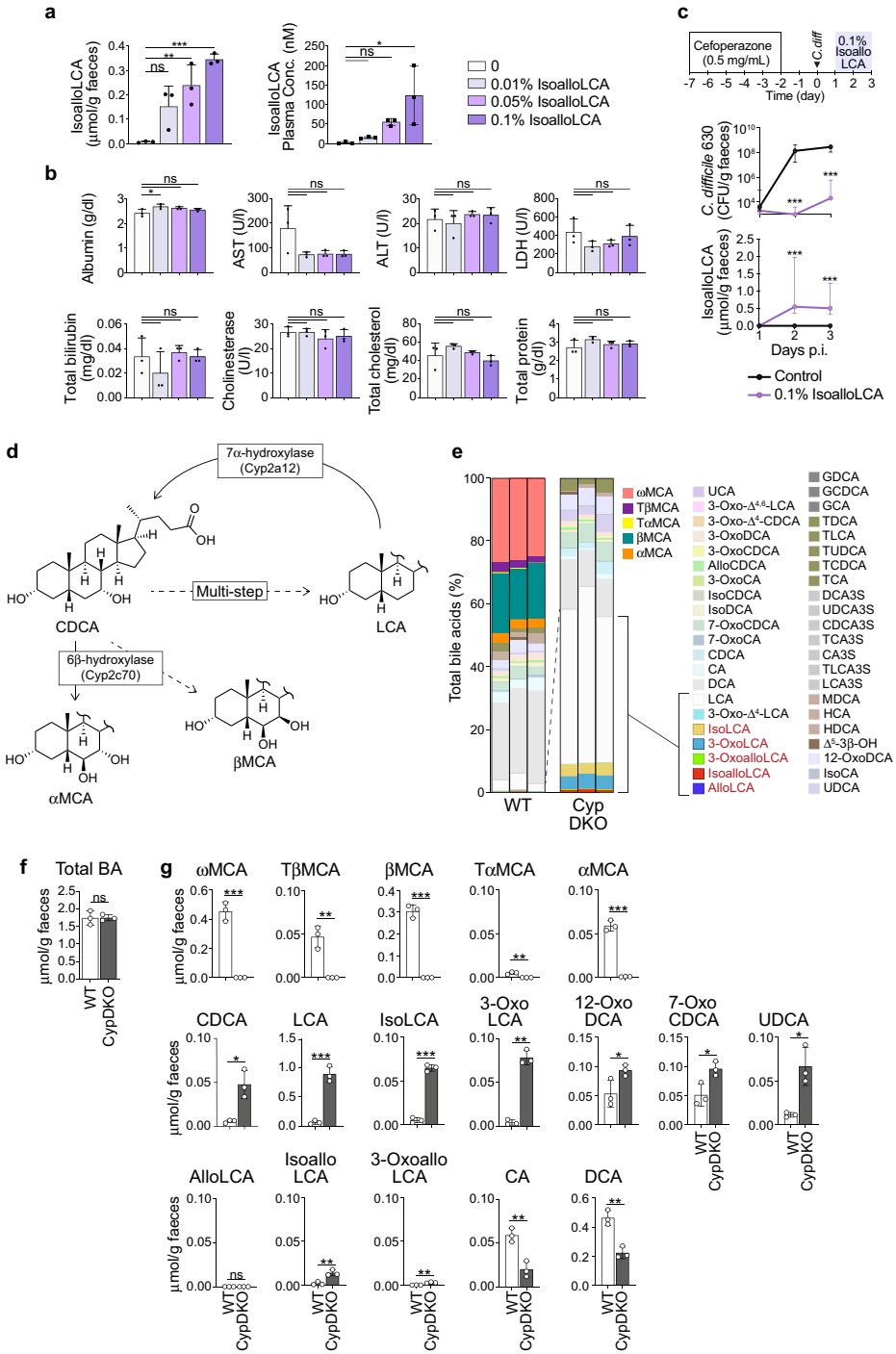
Extended Data Fig. 8 | Elucidated biosynthetic pathway of secondary bile acids. a, Elucidated biosynthetic pathway of secondary bile acids metabolized by CE91-derived bacterial strains. Colour-coded lines indicate in vitro identified metabolic capabilities of each strain. Responsible enzymes are indicated within boxes. Chemical structures are simplified by depicting only A and B steroid rings. **b**, The genome sequences of 68 isolates from a centenarian (CE91) were determined using PacBio Sequel and Illumina MiSeq sequencers. The gene prediction and annotation of the generated contigs were performed using the Rapid Annotations using Subsystem Technology (RAST) server. Gene clusters containing 5AR (magenta), 5BR (blue), 3β-HSDH-I (green), and 3β-HSDH-II (purple) homologues identified in 23 Bacteroidales strains are shown. Note that the gene clusters containing 5AR, 5BR, 3β-HSDH-I and 3β-HSDH-II homologues are present in close proximity to genes annotated with tricarboxylic acid (TCA) cycle-related functions (beige) and transporter/efflux system membrane fusion proteins (yellow). Gene homologues are defined as similarity of $<1 \times 10^{-12}$ E-value, >30% sequence identity, and >60% query coverage. Arrows represent coding sequences and annotated functions are colour-coded accordingly. Scale bar is 2kb. Note that 3α-HSDH homologue is not present in any of the 5AR clusters.



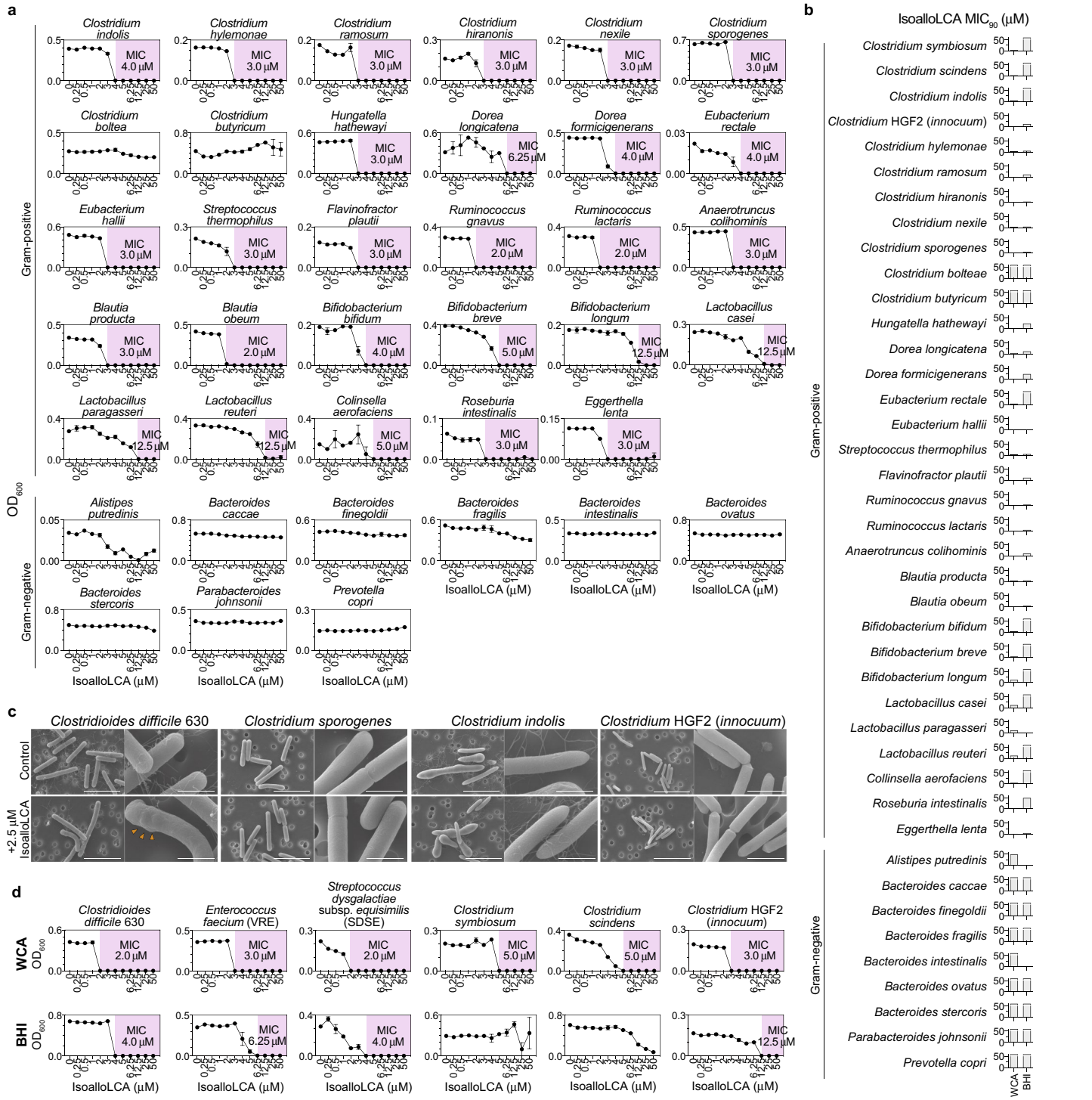
Extended Data Fig. 9 | Metabolism of LCA-related compounds by Bacteroidales strains and cooperative isoalloLCA production. a-**c**, Bacteroidales strains were cultured with 3-oxoalloLCA (**a**), 3-oxoLCA (**b**), or isoLCA (**c**) at a final concentration of 50 μM in pH 9-adjusted WCA medium. There was substantial substrate specificity and strain-to-strain variation in transformation efficiency. For instance, *P. merdae* St3, *P. distasonis* St4–5, and Odoribacteraceae St21 exhibited strong 3β-HSDH activity, reflected by both high isoalloLCA production from 3-oxoalloLCA (**a**) and isoLCA production from 3-oxoLCA (**b**), whereas other strains showed less efficient biotransformation or substrate specificity despite carriage of putative 3β-HSDH genes. The strength of 5BR activity also differed among the strains: *P. merdae* St3, *P. distasonis* St4–5, *B. dorei* St6–7, and *B. uniformis* St10-13 effectively transformed 3-oxoLCA to 3-oxo-Δ^4-LCA (and further to 3-oxoalloLCA by 5AR), while other strains displayed moderate to weak activity (**b**). *Porphyromonas somerae* St14 lacked a putative 3β-HSDH gene but was able to generate isoalloLCA from 3-oxoalloLCA nonetheless (**a**), suggesting that it carries a strain-specific gene with 3β-HSDH activity. **d**, Co-culture of *P. merdae* St3 or Odoribacteraceae St21 (5BR, 5AR, and 3β-HSDH encoders) with *E. lenta* St34 (top, green, a 3α-HSDH and 3β-HSDH encoder), or *P. distasonis* St4 (bottom, yellow, a 3α-HSDH, 3β-HSDH, and 5BR encoder) supplemented with 50 μM LCA in pH 9-adjusted WCA medium. Notably, co-culture of *P. merdae* St3 and *E. lenta* St34 with LCA resulted in generation of alloLCA in addition to isoalloLCA. **e**, Different combinations of *P. merdae* St3, Odoribacteraceae St21, or *C. scindens* St59 (a *bai* operon encoder capable of converting CDCA to LCA) co-cultured with *E. lenta* St34 (top) or *P. distasonis* St4 (bottom) in the presence of 50 μM CDCA. In **a**–**e**, the presence of genes homologous to 5AR, 5BR, 3β-HSDH-I, 3β-HSDH-II, 3α-HSDH, or the *bai* operon in corresponding combinations of strains are indicated as grey boxes within the adjacent charts. Note that the *bai* operon in *C. scindens* contains 5BR and 3α-HSDH genes. Culture supernatants were collected after 48 h of anaerobic incubation at 37 °C for LC-MS/MS quantification. Data are mean + s.d. of duplicate samples and representative of two independent experiments.



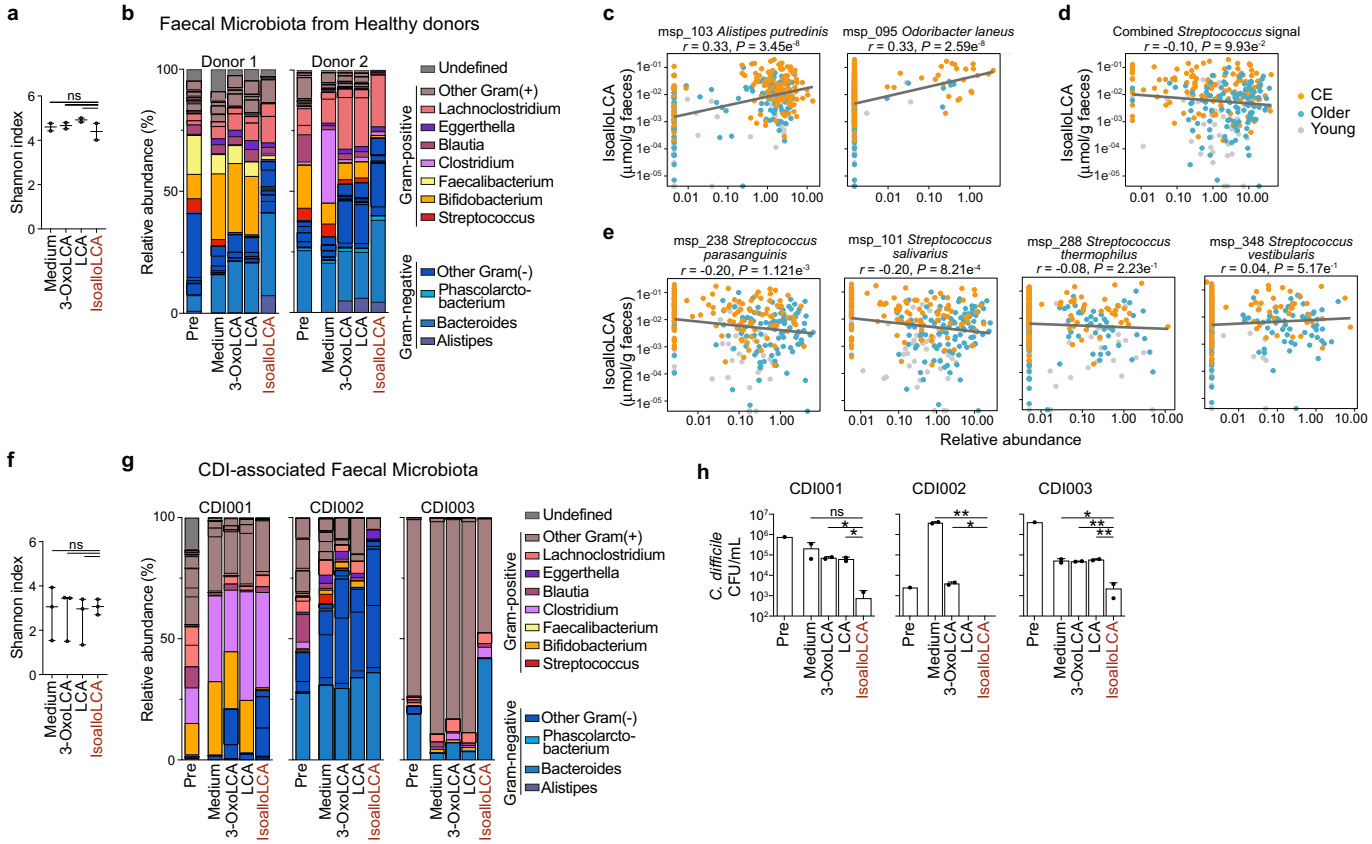
Extended Data Fig. 10 | IsoalloLCA inhibits Gram-positive pathogens. **a**, **b**, Gram-positive (upper) and -negative (lower) pathogens were incubated with varying concentrations of the indicated bile acids under anaerobic conditions at 37 °C in WCA medium until each strain reached stationary growth phase in the control medium (5-75 h). Bacterial growth was determined by OD_{600} measurement. Maximum growth densities (**a**) and growth curves (**b**) of pathogens in varying concentrations of bile acids are shown. Data are mean ± s.d. (error bars shown with fill area). **c**, MIC_{90} value (minimal inhibitory concentration required to prevent 90% growth) of isoalloLCA for each pathogenic strain was determined by serially diluting the compound and incubating each dilution with the pathogen. Shaded area indicates isoalloLCA concentrations with growth inhibitory effects, and MIC_{90} is reported inside the shaded box. Data are mean ± s.d. of duplicate samples and are representative of two independent experiments. **d**, In vitro growth inhibition of *C. difficile* 630 and vancomycin-resistant *E. faecium* (VRE) by co-culturing with CE91-derived isolates in the presence or absence of 12.5 µM 3-oxo-Δ^4-LCA in WCA medium. Average CFU of overnight cultures are shown ($n = 6$). Data are mean ± s.d. and are representative of two independent experiments. ***$P < 0.001$; Mann–Whitney test (two-tailed) with Welch's correction. ns, not significant.



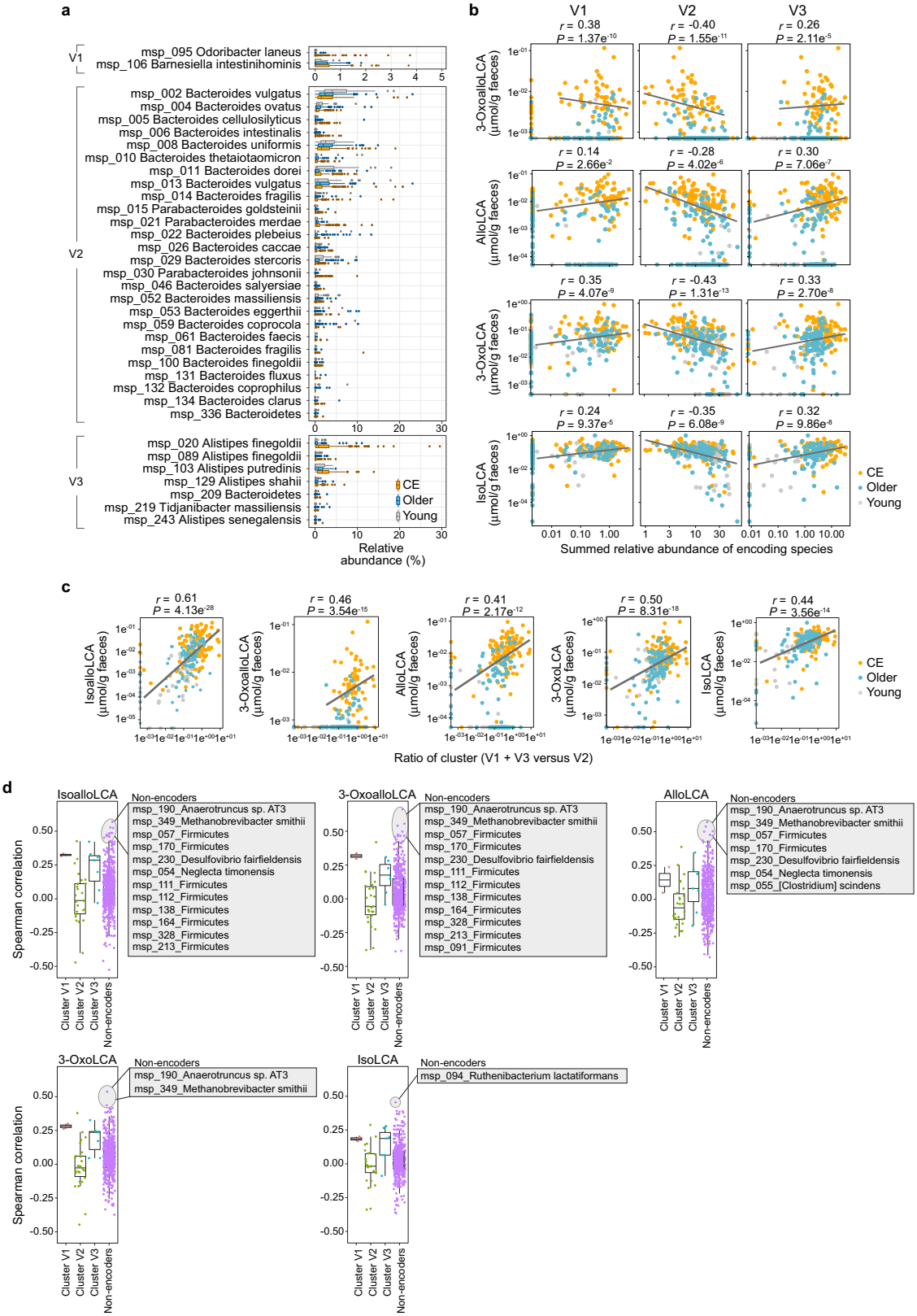
Extended Data Fig. 11 | In vivo suppression of faecal *C. difficile* shedding in *Cyp2a12*$^{-/-}$*Cyp2c70*$^{-/-}$ mice. a, **b**, Faecal and plasma concentrations of isoalloLCA (**a**) and plasma analytes (**b**) from SPF C57BL/6N mice after 14 days on an 0.01%, 0.05%, or 0.1% (w/w) isoalloLCA-containing diet. Plasma analytes from each animal: albumin; aspartate aminotransferase (AST); alanine transaminase (ALT); lactate dehydrogenase (LDH); total bilirubin; cholinesterase; total cholesterol; and total protein. Data are mean ± s.d.; $n = 3$ each; ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; one-way ANOVA with Tukey's test. ns, not significant. **c**, SPF C57BL/6N mice were pretreated with cefoperazone through the drinking water during the period from 7 to 2 days before inoculation and then infected with *C. difficile* 630 by oral gavage on day 0. The mice were subsequently placed on an 0.1% isoalloLCA-containing diet from day 1 to day 3 after infection (p.i.) (n = 17-18 animals each). Faecal *C. difficile* CFUs and isoalloLCA levels throughout the course of infection were determined. Data are mean ± s.d. ***$P < 0.001$; Mann–Whitney test (two-tailed) with Welch's correction. **d**, Simplified pathway of bile acid metabolism in rodents. In rodents' livers, CDCA is immediately hydroxylated at the 6β-position to generate muricholic acids (α- and β-muricholic acids (MCAs)) by CYP2C70, and as such we employed a *Cyp2c70*$^{-/-}$ mouse model. As resulting excessive CDCA causes hepatotoxicity in *Cyp2c70*$^{-/-}$ mice, we cross-bred *Cyp2c70*$^{-/-}$ mice with *Cyp2a12*$^{-/-}$ mice lacking 7α-rehydroxylation capacity to minimize liver injury. **e**-**g**, Faecal bile acids of SPF wild-type C57BL/6N mice (WT) and *Cyp2a12*$^{-/-}$*Cyp2c70*$^{-/-}$ double-knockout (CypDKO) mice ($n = 3$ each) were quantified by LC-MS/MS (µmol/g wet weight faeces). The percent of each bile acid among total faecal bile acids from each mouse was calculated (**e**). Note that *Cyp2a12*$^{-/-}$*Cyp2c70*$^{-/-}$ double-knockout mice showed elevated levels of LCA and related bile acids. Amount of total faecal bile acids (**f**) and each bile acid (**g**) from each mouse. Data are mean ± s.d. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; one-way ANOVA with Tukey's test. ns, not significant. Each dot represents an individual.



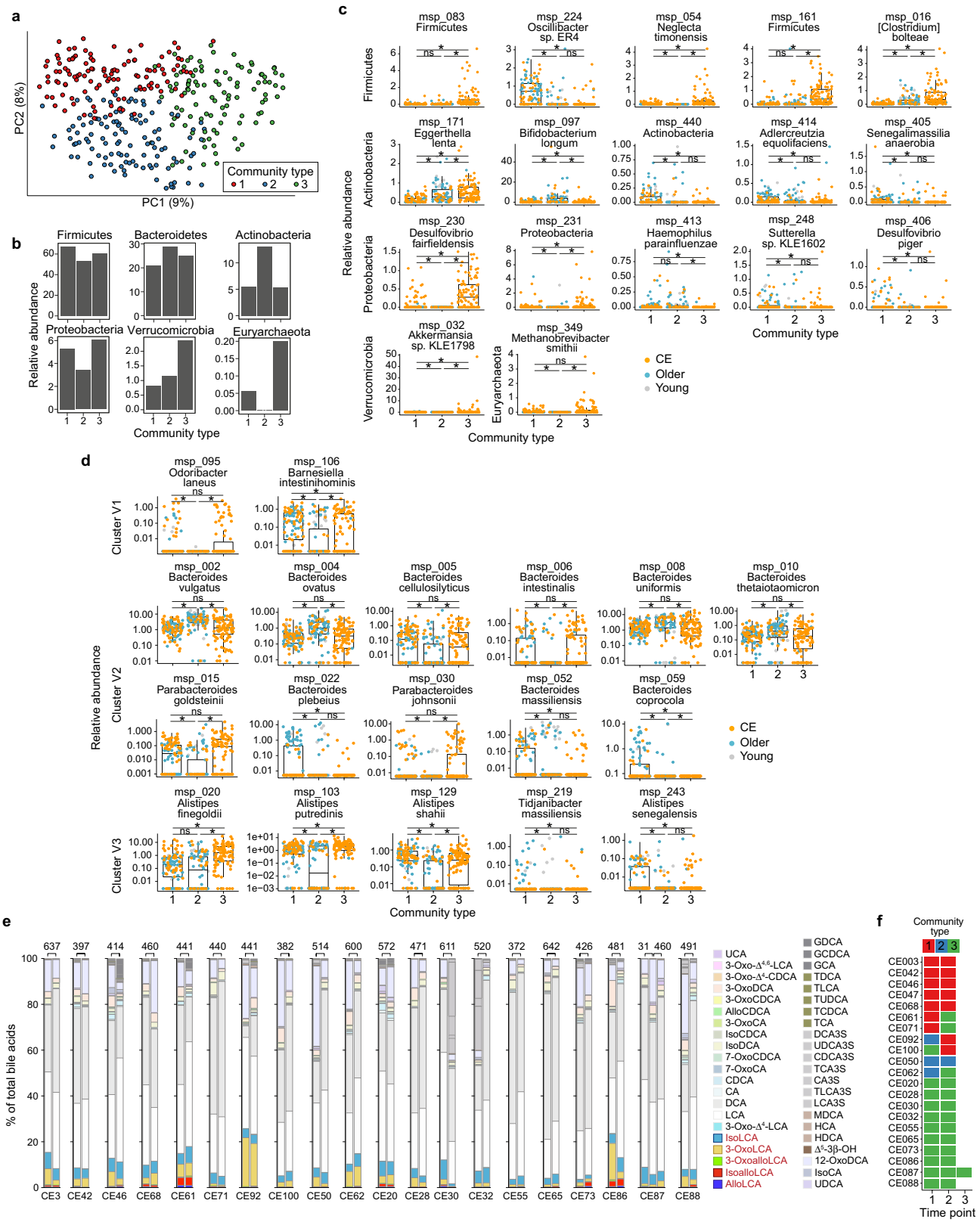
Extended Data Fig. 12 | MIC$_{90}$ of isoalloLCA against gut commensals in WCA and BHI media. a, Growth inhibitory effect of isoalloLCA on gut commensals in WCA medium. Gram-positive and Gram-negative commensal isolates were selected from our culture collection and incubated anaerobically at 37 °C with the indicated concentrations of isoalloLCA. Bacterial growth was determined by OD_{600} measurement. Each strain was cultured until it reached stationary growth phase in the control medium (7-48 h). Shaded area indicates isoalloLCA concentrations with growth inhibitory effects. MIC$_{90}$ is shown inside the shaded box. **b**, MIC$_{90}$ of isoalloLCA against Gram-positive and Gram-negative commensals in WCA and BHI media. **c**, Representative scanning electron microscopy images of *C. difficile* 630, *C. sporogenes*, *C. indolis* and *Clostridium* HGF2 (*innocuum*) grown in control or 2.5 μM isoalloLCA-containing WCA medium for 5 h. The right panels show the corresponding high-resolution images. Scale bars are 5.0 μm and 1.0 μm. Arrows indicate morphological alterations after isoalloLCA treatment. **d**, The differences in growth inhibitory effect of isoalloLCA between WCA and BHI medium on *C. difficile* 630, *E. faecium* (VRE), *S. dysgalactiae* subsp. *equisimilis* (SDSE), *C. symbiosum*, *C. scindens* and *Clostridium* HGF2 (*innocuum*). Data in **a**, **d** are mean ± s.d. from duplicate samples and are representative of two independent experiments.



Extended Data Fig. 13 | The effects of isoalloLCA on the complex gut microbiota. a, **b**, **f**–**h**, Human faecal samples from healthy young donors (**a**, **b**) and patients with *C. difficile* infection (CDI) (**f**–**h**) were incubated for 48 h in modified WCA medium supplemented with 3-oxoLCA, LCA, or isoalloLCA (50 µM). Shannon index of diversity (**a** and **c**) and a compositional shift in the microbiome at the genus level (**b** and **g**) of faecal cultures after secondary bile acid treatment. Each dot in **a** and **f** represents a single donor's faecal culture. ns, not significant by Kruskal–Wallis with Dunn's test. Although α-diversity was not significantly affected, isoalloLCA induced broad changes in microbial community structure, with a significant alteration in the ratio of Gram-negative to Gram-positive bacteria. *C. difficile* CFUs in each CDI microbiota pre- or post- 48 h incubation with medium or secondary bile acids were determined (**h**). Data are mean ± s.d. of duplicate samples. $^{**}P < 0.01$; $^{*}P < 0.05$; Unpaired *t*-test. ns, not significant. **c**–**e**, Correlation between the levels of faecal isoalloLCA and gut MSPs in centenarian (CE, $n = 119$, orange), older ($n = 107$, blue), and young ($n = 39$, grey) participants. Spearman's coefficient (r) and significance (P) for msp_103 *A. putredinis* and msp_095 *O. laneus* (both 5AR and 3β-HSDH encoders) (**c**), combined *Streptococcus* signal (**d**), and *Streptococcus* species (**e**). Each dot represents an individual. Several *Streptococcus* spp. showed a negative association with isoalloLCA concentration in our cohort.



Extended Data Fig. 14 | Associations between species encoding cluster V1, V2 or V3, or non-encoders with secondary bile acids. a, Relative abundance of gut MSPs categorized into clusters V1, V2, and V3 from the assembled gut metagenomes of centenarian (CE, $n = 119$, orange), older ($n = 107$, blue), and young ($n = 39$, grey) groups. **b**, **c**, Spearman's coefficient (r) and significance (P) between faecal bile acid level and summed relative abundance of cluster V1-, V2-, or V3-encoding species (**b**) and ratio of species encoding clusters V1 and V3 versus V2 from assembled gut metagenomes (**c**). Each dot represents an individual. **d**, Spearman correlation coefficient (r) for each bile acid against species encoding cluster V1, V2, or V3, or non-encoders. Non-encoders that correlate strongly with bile acid concentration are indicated within a grey circle, and the species are listed.



Extended Data Fig. 15 | See next page for caption.

Extended Data Fig. 15 | Stratification into microbial community types with respective microbiome characteristics. a, Principal coordinate analysis of three microbial community types (1, 2, and 3) stratified using Dirichlet Multinomial Mixtures from the gut microbiomes of centenarian [CE, $n = 176$ (153 individuals)], older ($n = 110$), and young ($n = 44$) individuals. **b**, **c**, Relative abundances across phyla (**b**) and top differentially abundant species from each phylum (**c**) from the assembled gut metagenomes in each community type. Community type 1 is characterized by a high relative abundance of Firmicutes (for example, *Oscillibacter* spp.) and Proteobacteria (for example, *Desulfovibrio* spp.), whereas community type 2 exhibits relative enrichment of Actinobacteria and depletion of Proteobacteria. Community type 3 is structurally similar to type 1, but exhibits a higher abundance of Verrucomicrobia (for example, *Akkermansia* spp.) and Euryarchaeota (for example, *Methanobrevibacter* spp.). **d**, Species harbouring gene clusters V1, V2, and V3 from the assembled gut metagenomes in each community type. Each dot in **c** and **d** represents an individual from the centenarian (orange), older (blue), or young (grey) groups. Horizontal lines indicate the median; box boundaries indicate interquartile range (IQR); whiskers represent values within 1.5 x IQR of the first and third quartiles. Asterisks indicate significantly different abundance in the specified comparison at FDR $P < 0.05$ based on a Wilcoxon rank-sum test. ns, not significant. **e**, Longitudinal change in faecal bile acid composition from the same individual over the course of 1-2 years. Upper numbers indicate collection interval (days). Faecal bile acids were quantified by LC-MS/MS (μmol/g wet weight faeces), and the percent of each bile acid among total faecal bile acids in each sample was calculated. **f**, Stability of microbial community type (1, 2, and 3) in each individual over time.

nature research

Corresponding author(s): Kenya Honda

Last updated by author(s): Jul 7, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	LabSolutions v5.97 software was used for LC-MS/MS and GCMSsolution v4.45 software for GC-MS/MS data collection. PLATEmanager v5/S software was used for collecting OD600 measurement.
Data analysis	To construct amplicon sequence variants (ASVs), Cutadapt v1.15 and DADA2 R package v1.18.0 were used. Taxonomy was assigned to each ASVs by search against the National Center fro Biotechnology Information (NCBI) using the GLSEARCH program. For metagenome analysis, sequence were pre-processed using Broad Picard Pipeline v2.9.4, Trim_Galore! v.0.4.4, and KneadData v0.7.2, then assembled by MegaHIT v1.1.4, Prodigal v2.6.3, and CD-HIT v4.7. USERACH v7.0.1090 was used for the 16S search of metagenome data. Reads were mapped to the gene catalogue with BWA v0.7.17, counted (count matrix) and normalized to transcripts per kilobase million (TPM matrix) using in-house scripts (https://gitlab.com/xavier-lab-computation/public/centenarianmicrobiome), and binned into metagenomic species using MSPminer. Assembled genes were annotated with NCBIRefseq (version May 2018) and MSPs that had no match to any species were annotated using PhyloPhlan v0.99. Microbiome diversities and Bray-Curtis dissimilarity were analysed using Vegan package in R v3.6.1. Bacterial whole-genome sequencing was prepared using TruSeq DNA PCR-Free kit, FASTX-toolkit v0.0.13, SMRTbell template prep kit 2.0, and Canu v1.8. Sequences were assembled using Unicycler v0.4.8 and annotated using the Rapid Annotations based on Subsystem Technology (RAST) Prokaryotic Genome Annotation Server v2.0 and Prokka: rapid prokaryotic genome annotation software tool v1.14.5. LC-MS/MS data was obtained using Analyst software v1.71 and analysed by SCIEX OS-MQ software ver1.7.0.36606 and Labsolutions Insight v3.5. GraphPadPrism v8.0, Excel, and RStudio v1.2.1335 were used to generate graphs.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Genome sequences of the 68 strains isolated from a centenarian and 16S rRNA amplicon sequence datasets have been deposited in the DDBJ under BioProject PRJDB11902 and PRJDB11894, respectively. Metagenomic data have also been deposited in the NCBI under Bioproject PRJNA675598. LC-MS/MS data were deposited in Metabolomics Workbench under project ID PR001168 with study ID ST001851 for human faeces data and study ID ST001852 for in vitro data. All deposited data will be released before publication. Source data are provided with this paper. European Nucleotide Archive (accession number PRJEB25514) was used for the Sardinian metagenome data.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size for the human or mouse studies. For the human cohorts, no sample size calculation was conducted as we were not testing for end clinical outcomes nor testing any intervention. Sample sizes therefore represent the maximum number of samples we could obtain during the recruitment period. The number of animals studied per treatment group was based on our previous knowledge of the reproducibility, balancing statistical robustness and animal welfare.
Data exclusions	We excluded subjects undergoing antibiotic treatment (3 centenarians, 1 elderly) as this is known to heavily influence the composition of microbiome as well as the metabolomics data. Two human subjects who had high serum C-reactive protein levels and died shortly after the data collection were excluded entirely from this study. These criteria were pre-established in the laboratory.
Replication	Each hypothesis was tested with multiple types of experiment. Reproducibility of C. difficile infection level in animal studies was verified by performing 2-3 independent experiments, which yielded comparable results (Main Fig. 3c, d and Extended Data Fig. 10c).
Randomization	For the human cohorts no random allocation was used as our study was observational and did not test any intervention. For animal studies, mice were randomized into separate cages upon arrival from the vendor. Sex-matched littermates were used and experiments were intended to test a single variable. For experiments using Cyp2a12/Cyp2c70 double knockout mice, animals were allocated across treatment groups to balance age, litter, and body weight.
Blinding	All experiments were not strictly blinded because the measurements were quantitative. Analysis of metabolite levels were conducted in an objective manner by more than two researchers.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	8-15 weeks old SPF and germ-free C57BL/6N WT female mice, 8-20 weeks old Cyp2a12/Cyp2c70 double knockout male and female mice were used in this study. All animals were maintained on the 12-hour light-dark cycle and received gamma-irradiated pellet food (50 kGy radiated CL2, CLEA Japan). Temperature of 20-24°C and humidity 40-60% were used for the housing conditions.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	All animal experiments were approved by the Keio University Institutional Animal Care and Use Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Participants in this study were centenarians (n = 160, +100 yo), elderly subjects (n = 112, 85-89 yo), young subjects (n = 47, 21-55 yo), and lineal descendants and siblings of centenarians (n = 22). Detailed population characteristics are presented in Extended Data Figure 1 and Table S1.
Recruitment	Centenarians and lineal relatives of centenarians were recruited as part of The Japan Semi-supercentenarian Study. The elderly cohorts were recruited as a part of Kawasaki Aging and Wellbeing project. Young cohorts were recruited under the IRB code 20150075 at Keio University School of Medicine. Faecal sample collection of patients with CDI was conducted under the IRB of Keio University School of Medicine (code 20150075). There was no bias towards selection of any particular group.
Ethics oversight	Institution Review Boards of Keio University School of Medicine

Note that full information on the approval of the study protocol must also be provided in the manuscript.