was a significant accumulation of $T_H17$ cells in the colons of exGF mice inoculated with untreated human faeces (Fig. 1c and Supplementary Fig. 2). Notably, the capacity of human faeces to induce $T_H17$ cells was greatly diminished after treatment with chloroform (Fig. 1c). These results indicate that the chloroform-sensitive bacterial fraction in the human stool tested contained $T_H17$-cell-inducing bacteria, whereas the chloroform-resistant bacteria preferentially promoted $T_{reg}$ cell accumulation in the colon.

To investigate whether $T_{reg}$ cell induction by the chloroform-resistant fraction of human intestinal bacteria is transmissible, adult germ-free mice were co-housed with exGF mice inoculated with chloroform-treated human faeces for 4 weeks. Co-housed mice showed a significant increase in the frequency of colonic $T_{reg}$ cells (Fig. 1d). In addition, the progeny of exGF mice inoculated with chloroform-treated human faeces also showed increased numbers of $T_{reg}$ cells (Fig. 1d). Therefore, $T_{reg}$ cell induction by human intestinal bacteria is horizontally and vertically transmissible. Oral inoculation of germ-free mice with $2 \times 10^4$-fold diluted caecal samples from exGF mice inoculated with chloroform-treated human faeces fully induced the accumulation of $T_{reg}$ cells in the colon lamina propria, suggesting that abundant rather than minor members of the intestinal microbiota in exGF mice inoculated with chloroform-treated human faeces drive the observed induction of $T_{reg}$ cells (Fig. 1e). The $T_{reg}$-cell-inducing microbiota in mice inoculated with the $2 \times 10^4$-fold diluted sample ($+2 \times 10^4$ mice) was a stable community, because serial oral inoculation of caecal contents from these mice equally induced the accumulation of $T_{reg}$ cells in secondary ($+2\times10^4$-re mice) and tertiary recipients ($+2\times10^4$-re-re mice) (Fig. 1e). To minimize nonessential components of the microbiota for $T_{reg}$ cell induction, the caecal contents of $+2 \times 10^4$ mice were again diluted $2 \times 10^4$-fold and orally inoculated into another set of germ-free mice ($+(2 \times 10^4)^2$ mice). The $+(2 \times 10^4)^2$ mice had a marked accumulation of $T_{reg}$ cells in the colon (Fig. 1e). These results suggested that we succeeded in obtaining mice colonized with a relatively restricted and stable community of bacterial species enriched for $T_{reg}$ cell inducers.

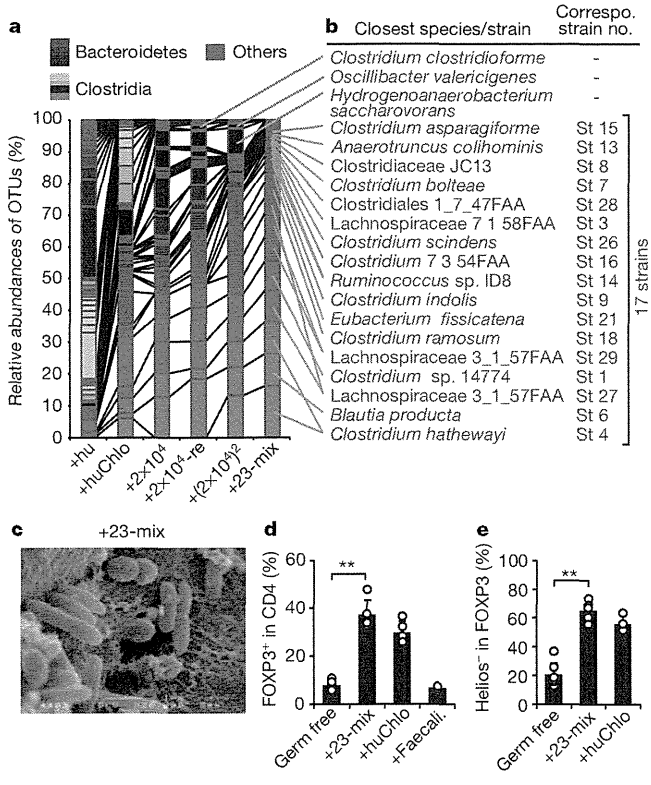


**Figure 2 | Assessment of microbiota composition and isolation of $T_{reg}$-cell-inducing strains. a, b,** Pyrosequencing of 16S rRNA genes was performed on caecal contents from the indicated mice. Relative abundance of OTUs (%) in the caecal bacterial community in each mouse (**a**), and the closest species/strain in the database and the corresponding isolated strain number for the indicated OTU (**b**) are shown. **c**, SEM showing the proximal colon of +23-mix mice. Original magnification, ~×20,000. **d, e**, The percentages of $FOXP3^+$ cells within the $CD4^+$ cell population (**d**) and $Helios^-$ cells in $CD4^+FOXP3^+$ cells (**e**) in the colon of the indicated mice. Circles represent individual animals. All experiments were performed more than twice with similar results. Error bars indicate s.d. $^{**}P < 0.01$.

The composition of the gut microbiota in mice treated with human samples was analysed by 16S ribosomal RNA (rRNA) gene amplicon sequencing using a 454 sequencer. Quality filter-passed sequences (3,000 reads for each sample) were classified into operational taxonomic units (OTUs) based on sequence similarity (>96% identity). The numbers of detected reads and closest known species for each OTU are shown in Supplementary Table 1, and the relative abundance of OTUs in each caecal sample is shown in Fig. 2a. As expected, the OTU profiles of mice treated with human faeces were quite different from those of conventional specific pathogen-free (SPF) mice (Supplementary Fig. 3). In mice inoculated with untreated human faeces, OTUs belonging to Bacteroidetes accounted for about 50% of the caecal microbial community (Fig. 2a). By contrast, most OTUs in exGF mice inoculated with chloroform-treated human faeces were related to Clostridia species. Most bacteria in $+2 \times 10^4$, $+2 \times 10^4$-re and $+(2 \times 10^4)^2$ mice had 16S rRNA gene sequence similarities with about 20 species of Clostridia, listed in Fig. 2b.

To isolate bacterial strains with $T_{reg}$-cell-inducing capabilities, we cultured caecal contents from $+2 \times 10^4$, $+2 \times 10^4$-re and $+(2 \times 10^4)^2$ mice *in vitro* and picked 442 colonies. BLAST searches of 16S rRNA gene sequences of the isolated colonies revealed that 31 strains in total were present, all of which were Clostridia (Supplementary Fig. 4). Of the 31 strains, we selected 23 that had less than 99% 16S rRNA gene sequence identity to any of the other 30 strains (Supplementary Fig. 4). We then individually cultured the 23 strains, mixed them in equal amounts, and orally inoculated the mixture into germ-free IQI mice (+23-mix mice). Numerous rod- and round-shaped bacteria were observed by scanning electron microscopy (SEM) on the epithelial cell surface in +23-mix mice (Fig. 2c), and the size and appearance of the caeca were quite different from those in germ-free mice, indicating successful colonization (Supplementary Fig. 5a). Pyrosequencing of 16S rRNA genes revealed that the caecal microbiota composition in +23-mix mice was quite similar to that in $+(2 \times 10^4)^2$ mice (Fig. 2a). In +23-mix mice, we observed efficient induction of $T_{reg}$ cells in the colonic lamina propria (Fig. 2d). The magnitude was comparable to that observed in exGF mice inoculated with chloroform-treated human faeces and much higher than that in mice colonized with *Faecalibacterium prausnitzii*, a human Clostridia strain well known for enhancing regulatory cell functions[9] (Fig. 2d). Most $T_{reg}$ cells in +23-mix mice expressed low levels of Helios (also known as IKZF2), indicating antigen-experienced cells (Fig. 2e, Supplementary Fig. 5b and ref. 10).

Only 17 strains listed in Fig. 2b and Supplementary Fig. 4 were detected in +23-mix mice by 16S rRNA gene sequencing, indicating that these 17 strains may be sufficient to induce $T_{reg}$ cells. Indeed, we found that the mixture of 17 strains (17-mix) induced $FOXP3^+$ $T_{reg}$ cells to a similar extent as the 23-mix (Fig. 3a). The increase in $T_{reg}$ cells induced by the 17-mix was reproducibly observed in exGF mice of different genetic backgrounds (IQI, BALB/c and C57BL/6) (Fig. 3a). Moreover, the mix was effective in other rodents: the frequency of colonic $T_{reg}$ cells in exGF rats inoculated with 17-mix was significantly higher than that in germ-free rats and comparable to that in SPF rats (Fig. 3a). The colonization with 17-mix induced a significant increase in the frequency of $IL\text{-}10^+$ and/or $ICOS^+$ cells within the $T_{reg}$ cell population, as revealed by analysis of exGF IL-10 reporter mice (*Il10*$^{Venus}$ mice, ref. 4) colonized with the 17-mix (Fig. 3b). Furthermore, $IL\text{-}10^+$ $T_{reg}$ cells in

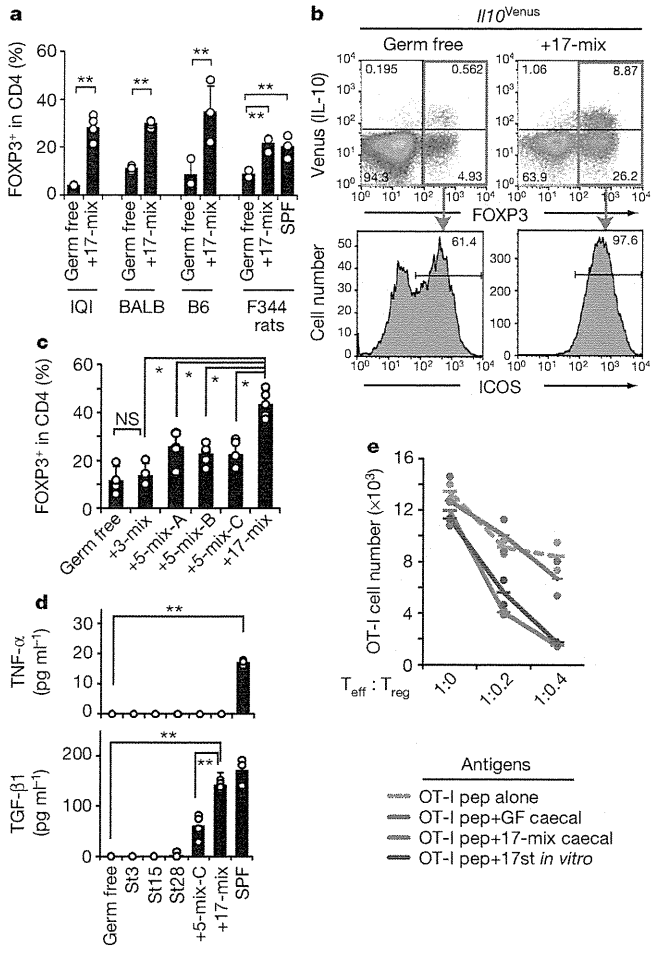


**Figure 3 | Characterization of 17 $T_{reg}$-cell-inducing strains.** **a**, The percentages of $FOXP3^+$ cells within the $CD4^+$ cell population in the colon lamina propria of the indicated mice and rats. **b**, The expression of Venus (IL-10) and FOXP3 by the gated colonic lamina propria $CD4^+$ cells, and ICOS expression by $CD4^+FOXP3^+$ cells in exGF *Il10*$^{Venus}$ mice colonized with or without 17-mix. **c**, Percentages of $FOXP3^+$ cells within the $CD4^+$ cell population in IQI exGF mice colonized with the indicated mix. **d**, The production of TNF-α and TGF-β1 in HCT8 cells stimulated with caecal extracts from the indicated mice. **e**, $CD8^+$ T cells from OT-I mice ($T_{eff}$) and the indicated ratio of colon lamina propria $CD4^+CD25^+$ cells from +17-mix mice ($T_{reg}$) were incubated with $CD11c^+$ cells pulsed with OT-I peptide alone or in combination with autoclaved caecal contents from +17-mix mice (+17-mix caecal), germ-free mice (+GF caecal), or autoclaved 17 strains cultured *in vitro* (+17 st *in vitro*). Depicted data represent average of duplicates (see also Supplementary Fig. 9c). Circles in **a**, **c**–**e** represent samples from individual animals. All experiments were performed more than twice with similar results. Error bars indicate s.d. $^{**}P < 0.01$; $^{*}P < 0.05$; NS, not significant.

+17-mix mice expressed high levels of CTLA4 (Supplementary Fig. 5c). Because IL-10 and CTLA4 are essential for the immunosuppressive activity of $T_{reg}$ cells[11,12], and ICOS is required for the $T_{reg}$-cell-mediated suppression of $T_H2$ responses[13], these results suggest that the mixture of 17 strains affects both the number and function of $T_{reg}$ cells in the colon. Next, we monocolonized germ-free mice with one of each of the 17 individual strains to determine their individual $T_{reg}$ cell induction capability. The monocolonized exGF mice exhibited low to intermediate levels of $T_{reg}$ cell induction with inter-individual variability (Supplementary Fig. 6a). As expected, none of the strains induced $T_H17$ cells in the monocolonized mice (Supplementary Fig. 6b). We also examined $T_{reg}$ cell induction by subsets of the 17-mix (randomly selected combinations of 3–5 strains: 3-mix, 5-mix-A, 5-mix-B, and 5-mix-C, see Supplementary Fig. 4). Although all tested combinations of 5-mix induced increases in the frequency of $T_{reg}$ cells, the magnitude was substantially lower than that observed in +17-mix mice (Fig. 3c). Therefore, it is likely that the 17 strains act synergistically to amplify the induction of $T_{reg}$ cells in a microbial-community-dependent fashion.

To investigate the mechanism for the $T_{reg}$ cell induction by the community of 17 strains, we incubated various human and mouse intestinal epithelial cell lines and primary cells with aqueous extracts from caecal contents from the +17-mix mice, and assessed the production of the active form of TGF-β1, a key cytokine for the differentiation and expansion of $T_{reg}$ cells. The caecal extracts from +17-mix mice routinely elicited TGF-β1, but not IL-6 and TNF-α, production, and the magnitude was significantly higher than that elicited by caecal extracts from single-strain or 5-mix-colonized mice (Fig. 3d and Supplementary Fig. 7). The induction of TGF-β1 was not inhibited by pre-treatment of the caecal extracts with a protease or nuclease (Supplementary Fig. 7c). Short-chain fatty acids (SCFAs) are protease- and nuclease-insensitive and have been associated with regulation of host immune homeostasis[14]. Quantitative analysis of SCFAs in caecal contents from +17-mix mice showed a significantly higher concentration of acetate, propionate, butyrate and isobutyrate than that in single-strain or 5-mix-colonized mice (Supplementary Fig. 8a). Furthermore, a mixture of sodium salts of these SCFAs elicited TGF-β1 production in epithelial cells to a level similar to that seen when the cells were stimulated with caecal extracts from +17-mix mice (Supplementary Fig. 8b). Therefore, the community of 17 strains cooperatively produces SCFAs that can elicit a TGF-β response, and this activity may contribute to the differentiation and expansion of $T_{reg}$ cells. We also investigated whether the 17 strains provide bacterial antigens to T cells. To do this, we addressed the antigen specificity of $T_{reg}$ cells accumulated in +17-mix mice using a cognate antigen-driven suppression assay. $CD4^+CD25^+$ lamina propria T cells from +17-mix mice substantially inhibited the OT-I ovalbumin (OVA) peptide-driven proliferation of OT-I CD8 T cells, and this suppression was markedly enhanced in the presence of autoclaved caecal content from +17-mix mice or autoclaved 17 strains cultured *in vitro*, but not in the presence of OT-II OVA peptide or caecal content from germ-free mice (Fig. 3e and Supplementary Fig. 9). These results are consistent with previous reports[15,16] and suggest that some fraction of colonic lamina propria $T_{reg}$ cells in +17-mix mice is specific to the 17 strains of Clostridia. Next, we assessed the kinetics of $T_{reg}$ cell accumulation and their expression of Ki67, a cell-cycle-associated nuclear protein, and gut-homing-associated molecules CD103 and β7 integrin. We observed a marked increase in the proportion of Ki67, CD103 and β7 expressing cells by 1 week after inoculation with the 17-mix (Supplementary Figs 10 and 11). Collectively, these observations indicate that the 17 strains provide SCFAs, bacterial antigens and probably other factors, which together contribute to differentiation, expansion and colonic homing of $T_{reg}$ cells.

To define the identity of the 17 bacterial strains fully, we sequenced their genomes (Supplementary Fig. 12). Phylogenetic comparison of the 17 strains using ribosomal multi locus sequencing typing (rMLST) revealed that the 17 strains belong to bacterial species falling within clusters XIVa, IV and XVIII of Clostridia as defined previously[17] (in a recent taxonomy, members of cluster XVIII Clostridia were reclassified in the class Erysipelotrichi) (Supplementary Fig. 13). The genome sequencing also revealed that the 17 strains all lack strong virulence-related genes such as collagenase and phospholipase C, often identified in pathogenic Clostridia species (Supplementary Table 2). We then examined the relative abundance of the 17 strains in healthy and ulcerative colitis human subjects using draft genome sequences of the 17 strains and publicly available human microbiome genomes generated through the MetaHIT project[18]. Ulcerative colitis subjects showed a tendency towards a reduction of the 17 strains, and 5 out of the 17 strains were significantly reduced in ulcerative colitis patients (Supplementary Fig. 14).

To evaluate the potential benefits of supplementation with the 17 strains, 17-mix or control PBS was orally administered into adult SPF

mice every 2 or 3 days (SPF + 17-mix or SPF + ctrl, respectively). We confirmed a significant increase in the frequency of colonic $T_{reg}$ cells in SPF + 17-mix mice compared with SPF + ctrl mice after 3 weeks of treatment (Fig. 4a). While being continuously treated with 17-mix or control, mice were subjected to the OVA-induced allergic diarrhoea model[19]. The occurrence and severity of diarrhoea and the OVA-specific serum IgE levels were significantly reduced in SPF + 17-mix mice relative to control mice (Fig. 4b–d). The protective effect of 17-mix was significantly attenuated by treatment of mice with a $T_{reg}$-cell-depleting anti-CD25 antibody (Supplementary Fig. 15). We also subjected mice to an experimental colitis model induced by trinitrobenzene sulphonic acid (TNBS)[20]. SPF + 17-mix mice showed less severe colon shortening and milder histological disease features, accompanied by lower mortality

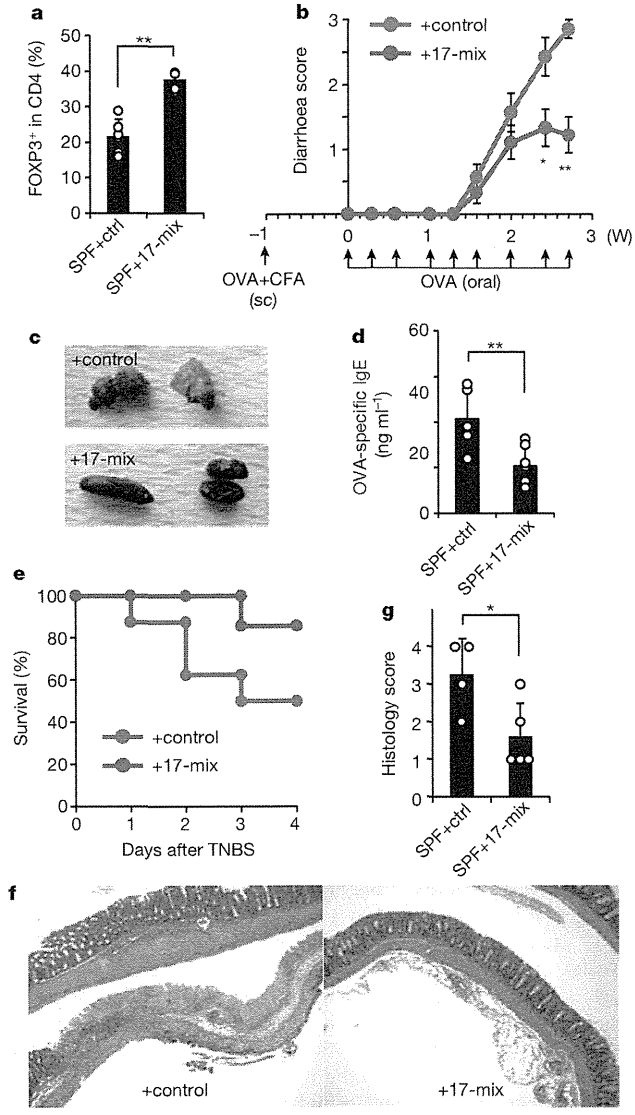


**Figure 4 | Treatment with 17-mix suppresses experimental colitis models.** **a**, The percentages of FOXP3$^+$ cells within the CD4$^+$ cell population in SPF + 17-mix or SPF + ctrl mice. **b–d**, SPF + 17-mix ($n = 9$) and SPF + ctrl ($n = 7$) mice were subjected to OVA-induced diarrhoea. The diarrhoea score (**b**; see Methods for definition), representative photographs of faeces (**c**), and OVA-specific IgE levels in the sera (**d**) are shown. sc, subcutaneous. **e–g**, SPF + 17-mix ($n = 8$) and SPF + ctrl ($n = 7$) were treated with TNBS. Animal survival (**e**), haematoxylin and eosin staining (original magnification, ×10) (**f**), and histology score of the distal colon (**g**) on day 4 after TNBS administration are shown. Data are representative of two independent experiments. Error bars indicate s.d. $^{**}P < 0.01$; $^{*}P < 0.05$.

than control mice (Fig. 4e–g and Supplementary Fig. 16a). In keeping with this clinical outcome, there was significantly increased expression of *Foxp3* and *Tgfb1* mRNA in SPF + 17-mix mice compared with control mice, as well as a tendency towards a reduction of inflammatory cytokine transcripts (Supplementary Fig. 16b). Identical suppression of colitis by 17-mix was also observed in an adoptive transfer model, in which germ-free SCID mice were orally inoculated with faeces from SPF mice together with or without 17-mix and then transferred with CD4$^+$CD45RB$^{hi}$ T cells (Supplementary Fig. 17).

The clinical track record of efficacy of single-strain probiotics has been modest. It has been postulated that a collection of functionally distinct bacterial species rationally selected from the human gut microbiota may be more effective than single strains in preventing/treating disease[21]. In the present study, we isolated 17 strains within Clostridia clusters XIVa, IV and XVIII from a human faecal sample; these strains affect $T_{reg}$ cell differentiation, accumulation and function in the mouse colon. It remains to be seen whether the 17 strains will have similar effects in the human intestine; however, a decreased prevalence of Clostridia clusters XIVa and IV in faecal samples from patients with inflammatory bowel disease and atopy[22–24] may suggest that supplementation with the 17-strain bacterial community might counterbalance dysbiosis, induce $T_{reg}$ cells and aid in the management of allergic and inflammatory conditions.

## METHODS SUMMARY

Experiments were performed with authorization from the Institutional Review Board for Human Research at RIKEN Yokohama Research Institute. Human stool from a healthy volunteer (Japanese, male, age 31 years) was obtained with informed consent. The sample was mixed with or without chloroform, and the aliquots were inoculated into germ-free IQI mice. Detailed procedures for lamina propria lymphocyte analysis, isolation of bacteria, extraction of bacterial DNA and sequencing are described in Methods. Statistical analysis was performed using the Student's *t*-test.

**Full Methods** and any associated references are available in the online version of the paper.

**Received 14 September 2012; accepted 22 May 2013.**
**Published online 10 July 2013.**

1. Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nature Rev. Immunol.* **9,** 313–323 (2009).
2. Honda, K. & Littman, D. R. The microbiome in infectious disease and inflammation. *Annu. Rev. Immunol.* **30,** 759–795 (2012).
3. O'Toole, P. W. & Cooney, J. C. Probiotic bacteria influence the composition and function of the intestinal microbiota. *Interdiscip. Perspect. Infect. Dis.* **2008,** 175285 (2008).
4. Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* **331,** 337–341 (2011).
5. Geuking, M. B. *et al.* Intestinal bacterial colonization induces mutualistic regulatory T cell responses. *Immunity* **34,** 794–806 (2011).
6. Russell, S. L. *et al.* Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. *EMBO Rep.* **13,** 440–447 (2012).
7. Round, J. L. & Mazmanian, S. K. Inducible Foxp3$^+$ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc. Natl Acad. Sci. USA* **107,** 12204–12209 (2010).
8. Chung, H. *et al.* Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* **149,** 1578–1593 (2012).
9. Sokol, H. *et al. Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl Acad. Sci. USA* **105,** 16731–16736 (2008).
10. Thornton, A. M. *et al.* Expression of Helios, an Ikaros transcription factor family member, differentiates thymic-derived from peripherally induced Foxp3$^+$ T regulatory cells. *J. Immunol.* **184,** 3433–3441 (2010).
11. Rubtsov, Y. P. *et al.* Regulatory T cell-derived interleukin-10 limits inflammation at environmental interfaces. *Immunity* **28,** 546–558 (2008).
12. Wing, K. *et al.* CTLA-4 control over Foxp3$^+$ regulatory T cell function. *Science* **322,** 271–275 (2008).
13. Zheng, Y. *et al.* Regulatory T-cell suppressor program co-opts transcription factor IRF4 to control $T_H2$ responses. *Nature* **458,** 351–356 (2009).
14. Maslowski, K. M. & Mackay, C. R. Diet, gut microbiota and immune responses. *Nature Immunol.* **12,** 5–9 (2011).
15. Lathrop, S. K. *et al.* Peripheral education of the immune system by colonic commensal microbiota. *Nature* **478,** 250–254 (2011).
16. Cebula, A. *et al.* Thymus-derived regulatory T cells contribute to tolerance to commensal microbiota. *Nature* **497,** 258–262 (2013).

17. Collins, M. D. *et al.* The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *Int. J. Syst. Bacteriol.* **44,** 812–826 (1994).
18. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
19. Kweon, M. N., Yamamoto, M., Kajiki, M., Takahashi, I. & Kiyono, H. Systemically derived large intestinal $CD4^+$ Th2 cells play a central role in STAT6-mediated allergic diarrhea. *J. Clin. Invest.* **106,** 199–206 (2000).
20. Strober, W., Fuss, I. J. & Blumberg, R. S. The immunology of mucosal models of inflammation. *Annu. Rev. Immunol.* **20,** 495–549 (2002).
21. Lawley, T. D. *et al.* Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing *Clostridium difficile* disease in mice. *PLoS Pathog.* **8,** e1002995 (2012).
22. Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl Acad. Sci. USA* **104,** 13780–13785 (2007).
23. Manichanh, C. *et al.* Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55,** 205–211 (2006).
24. Candela, M. *et al.* Unbalance of intestinal microbiota in atopic children. *BMC Microbiol.* **12,** 95 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by JSPS NEXT program, Grant in Aid for Scientific Research on Innovative Areas 'Genome Science' from the Ministry of Education, Culture, Sports, Science and Technology of Japan (No.221S0002), the global COE project of 'Genome Information Big Bang' and the Waksman Foundation of Japan Inc. We thank M. Suyama, K. Furuya, C. Yoshino, H. Inaba, E. Iioka, Y. Takayama, M. Kiuchi, Y. Hattori, N. Fukuda and A. Nakano for technical assistance, and P. D. Burrows for review of the manuscript.

**Author Contributions** K.Ho. planned experiments, analysed data and wrote the paper together with B.O. and M.H.; K.A. and T.Tano. performed immunological analyses and bacterial cultures together with Y.N., S.N. and H.M.; W.S., K.O., S.K. and M.H. performed bacterial sequence analyses; K.M. and S.U. provided essential materials; H.N., T.S. and S.S. supervised the $T_{reg}$ cell suppression assay; S.F., K.Ha., H.O., T.Tani., J.V.F. and P.W. were involved in data discussions.

**Author Information** All genome sequence data are deposited in DDBJ BioProject ID PRJDB521-543. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.H. (hattori@k.u-tokyo.ac.jp) or K.Ho. (kenya@rcai.riken.jp).

## METHODS

**Mice and rats.** C57BL/6, BALB/c, IQI/Jic and CB.17 SCID mice and F344 rats kept under SPF or germ-free conditions were purchased from Sankyo laboratories, Japan SLC, or CLEA Japan. IQI germ-free mice were used unless otherwise indicated. Germ-free and gnotobiotic mice were bred and maintained in vinyl isolators within the gnotobiotic facility of Sankyo laboratories. Germ-free *Il10*$^{Venus}$ mice were generated as previously described[4]. OT-I and OT-II T-cell receptor transgenic mice were purchased from Taconic Farms. All animal experiments were approved by the Animal Research Committee of RIKEN Yokohama Institute and the University of Tokyo.

**Chloroform treatment of human stool and generation of gnotobiotic mice.** Human stool from a healthy volunteer (Japanese, male, age 31 years) was obtained with informed consent. Human stool and mouse caecal contents were directly frozen at −80 °C, or suspended in 4 times volume (w/v) of phosphate-buffered saline (PBS) + 20% glycerol solution, snap-frozen in liquid nitrogen and stored at −80 °C until use. The frozen stocks were thawed, suspended in 10 times volume (w/v) of PBS, and passed through a 70 μm cell strainer to eliminate clumps and debris. Then suspensions were mixed with chloroform (final concentration 3%), and incubated in a shaking water bath for 60 min. After evaporation of chloroform by bubbling with $N_2$ gas for 30 min, the aliquots containing the chloroform-resistant fraction of intestinal bacteria were inoculated into germ-free mice by intra-gastric administration (250 μl; per mouse). To generate a series of gnotobiotic mice inoculated with diluted samples, caecal contents from exGF mice were treated with chloroform, diluted with PBS, and inoculated into germ-free IQI mice. The caecal suspensions diluted $2 \times 10^4$-fold correspond to $2.5 \times 10^4$ bacterial cells per mouse. Each group of exGF mice was individually caged in the gnotobiotic isolator for 3–4 weeks at Sankyo Lab service.

**Isolation of intestinal lamina propria lymphocytes and flow cytometry.** The colons were collected and opened longitudinally, washed with PBS to remove all luminal contents and shaken in Hanks' balanced salt solution (HBSS) containing 5 mM EDTA for 20 min at 37 °C. After removing epithelial cells, muscle layers and fat tissue using forceps, the lamina propria layers were cut into small pieces and incubated with RPMI1640 containing 4% fetal bovine serum, 0.5 mg ml$^{-1}$ collagenase D, 0.5 mg ml$^{-1}$ dispase and 40 μg ml$^{-1}$ DNase I (all Roche Diagnostics) for 1 h at 37 °C in a shaking water bath. The digested tissues were washed with HBSS containing 5 mM EDTA, resuspended in 5 ml of 40% Percoll (GE Healthcare) and overlaid on 2.5 ml of 80% Percoll in a 15-ml Falcon tube. Percoll gradient separation was performed by centrifugation at 800*g* for 20 min at 25 °C. The lamina propria lymphocytes were collected from the interface of the Percoll gradient and suspended in ice-cold PBS. For analysis of $T_{reg}$ cells, isolated lymphocytes were labelled with the LIVE/DEAD fixable dead cell stain kit (Invitrogen) to exclude dead cells from the analysis. The cells were washed with staining buffer containing PBS, 2% FBS, 2 mM EDTA and 0.09% $NaN_3$ and surface staining was performed with PECy7- or Pacific blue-labelled anti-CD4 antibody (RM4-5, BD Biosciences), PE-labelled anti-ICOS antibody (C938.4A, BioLegend), Alexa488-labelled anti-CD103 antibody (2E7, BioLegend), and PerCP/Cy5.5-labelled anti-integrin-β7 antibody (FIB27, BioLegend). Intracellular staining of FOXP3, CTLA4, Helios and Ki67 was performed using the Alexa647-labelled anti-FOXP3 antibody (FJK-16 s, eBioscience), PE-labelled anti-CTLA4 antibody (UC10-4F10-11, BD Biosciences), PE-labelled anti-Helios antibody (22F6, BioLegend), PECy7-labelled anti-Ki67 antibody (B56, BD Biosciences) and FOXP3 staining buffer set (eBioscience). For analysis of $T_H1$ and $T_H17$ cells, isolated lymphocytes were stimulated for 4 h with 50 ng ml$^{-1}$ phorbol 12-myristate 13-acetate (PMA, Sigma) and 1 μg ml$^{-1}$ ionomycin (Sigma) in the presence of GolgiStop (BD Biosciences). After incubation for 4 h, cells were washed in PBS, labelled with the LIVE/DEAD fixable dead cell stain kit and surface CD4 was stained with PECy7-labelled anti-CD4 antibody. Cells were washed, fixed in Cytofix/Cytoperm, permeabilized with Perm/Wash buffer (BD Biosciences), and stained with the APC-labelled anti-IL-17 antibody (eBio17B7, eBioscience) and FITC-labelled anti-IFN-γ antibody (XMG1.2, BD Biosciences). The antibody-stained cells were analysed with LSR Fortessa or FACSAriaIII (BD Biosciences), and data were analysed using FlowJo software (Treestar).

**Meta 16S rRNA gene sequencing.** The caecal contents from exGF mice were suspended in 10 ml of Tris-EDTA containing 10 mM Tris-HCl and 1 mM EDTA (pH 8), and incubated with lysozyme (Sigma, 15 mg ml$^{-1}$) at 37 °C for 1 h with gentle mixing. A purified achromopeptidase (Wako) was added (final concentration 2,000 unit ml$^{-1}$) and further incubated at 37 °C for another 30 min. Then, sodium dodecyl sulphate (final concentration 1%) was added to the cell suspension and mixed well. Subsequently, proteinase K (Merck) was added (final concentration 1 mg ml$^{-1}$) to the suspension and the mixture was incubated at 55 °C for 1 h. High-molecular-mass DNA was isolated and purified by phenol/chloroform extraction, ethanol and finally polyethyleneglycol precipitation[25]. PCR was performed using Ex Taq (TAKARA) and (1) the 454 primer A (5′-CCATCTCA TCCCTGCGTGTCTCCGACTCAG (454 adaptor sequence) + barcode (10 bases) + AGRGTTTGATYMTGGCTCAG-3′ (27Fmod)) and (2) the 454 primer B (5′-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG (454 adaptor sequence) + TGCTGCCTCCCGTAGGAGT-3′ (338R)) to the V1–V2 region of the 16S rRNA gene. Amplicons generated from each sample (~330 bp) were subsequently purified using AMPur XP (Beckman Coulter). The amount of DNA was quantified using Quant-iT Picogreen dsDNA assay kit (Invitrogen) and TBS-380mini fluorometer (Turner Biosystems). Then, the amplified DNA was used as template for 454 GS Junior (Roche) pyrosequencing using GS Junior Titanium emPCR Kit-Lib-L, GS Junior Titanium Sequencing Kit and GS Junior Titanium PicoTiterPlate Kit (all Roche) according to the manufacturer's instructions. Quality-filter-passed reads were obtained by removing reads that did not have both primer sequences, had the average quality value (QV) <25, and were possibly chimaeric[26]. Of the filter-passed reads, 3,000 reads trimming off both primer sequences for each sample were used and subjected to OTU analysis with the cutoff similarity of 96% identity. Representative sequences from each OTU were blasted to Ribosomal Database Project (RDP) of bacterial isolates, our genome database constructed from publically available genome sequences in NCBI and HMP databases, and 16S sequences of the 23 strains obtained in this study.

**Isolation of bacterial strains.** The frozen stocks of caecal contents from exGF mice were serially diluted with PBS and seeded onto non-selective agar plates (blood liver (BL) agar (Eiken Chemical) or Eggerth-Gagnon (EG) agar plates). EG agar plates contain the following components (quantities expressed per litre): Lab-Lemco Powder (2.8 g, Oxoid); proteose peptone no. 3 (10.0 g, Difco); yeast extract (5.0 g, Difco); $Na_2HPO_4$ (4.0 g); D(+)-glucose (1.5 g); soluble starch (0.5 g); L-cystine (0.2 g); L-cysteine-HCl-$H_2O$ (0.5 g); Tween 80 (0.5 g); Bacto agar (16.0 g, Difco); and defibrinated horse blood (50 ml). After culture under aerobic conditions or strictly anaerobic conditions (80% $N_2$, 10% $H_2$, 10% $CO_2$) at 37 °C for 2 or 4 days, individual colonies were picked up and cultured for an additional 2 or 4 days at 37 °C in ABCM broth (Eiken Chemical) or EG agar plate. The isolated strains were collected into EG stock medium (10% glycerol) and stored at −80 °C. To identify the isolated strains, 16S rRNA gene sequences were determined. The 16S rRNA gene was amplified by colony-PCR using KOD FX (TOYOBO) and GeneAmp PCR System9700 (Applied Biosystems) using 16S rRNA gene-specific primer pairs: 8F (5′-AGAGTTTGATCMTGGCTCAG-3′) and 519R (5′-ATTACCGCGGCKGCTG-3′) or 1513R (5′-ACGGCTACCTTGTTACGACTT-3′). The amplification program consisted of one cycle at 98 °C for 2 min, followed by 40 cycles at 98 °C for 10 s, 57 °C for 30 s and 68 °C for 1 min 30 s. Each amplified DNA was purified from the reaction mixture using Illustra GFX PCR DNA and gel band purification kit (GE Healthcare). Sequence analysis was performed using BigDye Terminator V3.1 cycle sequencing kit (Applied Biosystems) and Applied Biosystems 3730xl DNA analyser (Applied Biosystems). The resulting sequences were compared with sequences in RDP database and genome database using BLAST to determine close species/strains.

**Bacterial culture of isolated strains.** The isolated strains of Clostridia and Erysipelotrichi were cultured in EG broth without horse blood under a strictly anaerobic condition (80% $N_2$, 10% $H_2$, 10% $CO_2$) at 37 °C in an anaerobic chamber (Coy Laboratory Products). To prepare the bacterial mixture, bacterial strains were individually grown in EG broth to confluence and mixed at equal amounts of media volume.

**Scanning electron microscopy.** Scanning electron microscopy was performed by Filgen, Inc., Japan. The proximal colon was removed from +23-mix mice, cut open longitudinally, prefixed with 2% glutaraldehyde in 0.1 M phosphate buffer (pH 7.4) for 24 h at 4 °C, and then postfixed with 2% osmium tetroxide for 1 h at 4 °C. Fixed samples were dehydrated for 5 min each in sequential baths of 50%, 70%, 90% and 100% ethanol, inserted into a critical point dryer until dry and coated with osmium in an OPC-80N osmium plasma coater (Filgen). Scanning electron micrographs were taken by a JEOL JSM-6320F instrument.

**Measurement of organic acids.** Organic acid concentrations in caecal contents were determined by gas chromatography-mass spectrometry (GC-MS). Caecal contents (10 mg) were disrupted using 3-mm zirconia/silica beads (BioSpec Products) and homogenized in extraction solution containing 100 μl of internal standard (100 μM crotonic acid), 50 μl of HCl and 200 μl of ether. After vigorous shaking using a Shakemaster neo (Bio Medical Science) at 1,500 r.p.m. for 10 min, homogenates were centrifuged at 1,000*g* for 10 min and then the top ether layer was collected and transferred into new glass vials. Aliquots (80 μl) of the ether extracts were mixed with 16 μl of *N*-tert-butyldimethylsilyl-*N*-methyltrifluoroacetamide (MTBSTFA). The vials were sealed tightly by screwing and heated at 80 °C for 20 min in a water bath, and left at room temperature for 48 h for derivatization. The derivatized samples were run through a 6890N Network GC System (Agilent Technologies) equipped with HP-5MS column (0.25 mm × 30 m × 0.25 μm) and 5973 Network Mass Selective Detector (Agilent Technologies). Pure helium (99.9999%) was used as a carrier gas and delivered at a flow rate of 1.2 ml min$^{-1}$.

The head pressure was set at 10 p.s.i. with split 10:1. The inlet and transfer line temperatures were 250 °C and 260 °C, respectively. The following temperature program was used: 60 °C (3 min), 60–120 °C (5 °C $min^{-1}$), 120–300 °C (20 °C $min^{-1}$). One microlitre quantity of each sample was injected with a run time of 30 min. Organic acid concentrations were quantified by comparing their peak areas with the standards.

**Genome sequencing and gene prediction.** The genome sequences of 17 $T_{reg}$-cell-inducing strains were determined by the whole-genome shotgun strategy using a 454GS FLX Ti or Ion PGM sequencer. Each 1–5 μg of the genomic DNA was sheared to obtain DNA fragments. Template DNA was prepared according to the supplier's protocol. The generated sequence data were assembled using Newbler v2.8 software to obtain the draft genome sequences. All genome sequence data were deposited in DDBJ BioProject ID: PRJDB521-543. Protein-encoding genes were predicted using MetaGeneAnnotator software[27]. Putative toxins and virulence factors were searched using the BLASTP program and virulence factor databases, VFDB (http://www.mgc.ac.cn/VFs/main.htm) and MvirDB (http://mvirdb.llnl.gov), with the *e*-value cutoff of $1.0 \times 10^{-10}$, the identity >30% and the length coverage >60%.

**Phylogenetic tree.** Sequences concatenated with genes encoding 26 ribosomal proteins (large subunit L10, L11, L14, L16, L17, L19, L20, L23, L24, L29, L31, L32, L35, L7/L12, and small subunit S10, S12, S13, S15, S16, S17, S20, S20, S21, S3, S4, S7, S8) predicted from the genomes of each strain were used to construct a phylogenetic tree. The sequences of other bacterial species used for the tree construction were obtained from the ribosomal multi-locus sequencing typing (MLST) database[28]. The calculation was performed using the MEGA v5.0 package and the neighbour-joining method with a bootstrap of 1,000 replicates.

**Cognate antigen-driven $T_{reg}$ cell suppression assay.** Preparation of antigens in caecal contents was performed as previously reported[15]. Caecal contents from germ-free mice or +17-mix mice were collected and suspended in PBS (500 mg $ml^{-1}$); they were then filtered through a 70-μm mesh, and autoclaved at 121 °C for 15 min. To prepare antigens of bacterial components, the 17 strains of Clostridia were cultured *in vitro*, mixed, washed and suspended with 1 ml PBS, and autoclaved at 121 °C for 20 min. $CD11c^+$ cells were isolated by FACSAriaIII from spleens of SPF C57BL/6 mice and pulsed for 1 h with 0.5 μM SIINFEKL OT-I peptide alone or in combination with either of 5 μM ISQAVHAAHAEINEAGR OT-II peptide, autoclaved caecal contents from +17-mix mice or germ-free mice (diluted 1:200), or autoclaved 17 strains of bacteria cultured *in vitro* (diluted 1:200). The antigen-pulsed $CD11c^+$ cells were plated at $5 \times 10^4$ per well in 96-well round-bottomed plates. CD8 T cells ($T_{eff}$ cells) were sorted from spleens of SPF OT-I mice by FACSAriaIII and added to the $CD11c^+$ cell-seeded plates at $5 \times 10^4$ per well. Then, $CD4^+CD25^+$ T cells ($T_{reg}$ cells) sorted from colonic lamina propria of +17-mix mice or from spleens of SPF OT-II mice were added to the culture at the indicated ratio of $T_{reg}$ to $T_{eff}$ cells. After 3 days, all cells were harvested, stained with anti-CD4 and anti-CD8 antibodies, and analysed by FACSAriaIII to enumerate the number of CD8 OT-I T cells.

**Intestinal epithelial cell stimulation with caecal extracts and SCFAs.** To prepare caecal extracts, frozen caecal contents from germ-free, +17-mix or SPF mice were thawed and well suspended in 4 volumes of sterile water. After centrifugation (5,000 r.p.m. for 15 min), transparent supernatants were collected, filtered through 0.22 μm filter and used as caecal extracts. In some experiments, caecal extracts were treated with proteinase K (2 mg $ml^{-1}$, 55 °C for 1 h; Roche) or nuclease that degrades all forms of DNA and RNA (125 unit $ml^{-1}$, 37 °C for 4 h; Thermo), and subsequently heated at 95 °C for 5 min to inactivate the enzymes. Human intestinal epithelial cell lines (HCT8, HT29, Caco2, T84 and Colo205) and a mouse epithelial cell line (CMT93) were obtained from ATCC and maintained at 37 °C (5% CO2) in RPMI containing 10% heat-inactivated horse serum (Invitrogen). Cells were cultured at $1.5 \times 10^5$ cells in 150 μl medium in 48-well plates and stimulated with 4.5 μl caecal extract for 24 h. Human primary intestinal epithelial cells were obtained from Lonza and maintained at 33 °C (5% CO2) in SmGM-2 medium containing 10% FBS (Lonza) for 1–2 weeks ($6 \times 10^4$ cells in 48-well plates). The medium was changed to 150 μl SmGM-2 containing 1% FBS before stimulation. Caecal extracts (4.5 μl) were added to the culture and incubated for 24 h. Culture supernatants were collected and the level of the active form of TGF-β1 (Promega), TNF-α (R&D) and IL-6 (R&D) was measured by ELISA. To stimulate epithelial cell lines with SCFAs, sodium salts of acetate, butyrate, propionate and isobutyrate were dissolved in PBS. SCFAs were added to the culture individually (final 0.5 mM) or in combination (final 0.5 mM each), and incubated for 24 h.

**TNBS colitis.** C57BL/6 SPF adult mice were orally inoculated with 17-mix or control PBS every 2 or 3 days for 3 weeks. 2,4,6-Trinitrobenzene sulphonic acid (TNBS)-induced colitis was induced by the intracolonic administration of 2.5 mg of TNBS (Sigma) in 50% ethanol into anaesthetized mice via a thin round-tip needle. The tip of the needle was inserted 4 cm proximal to the anal verge, and mice were held in a vertical position for 30 s after the injection. All the mice were observed daily and were killed on day 4 after TNBS administration. Colons were fixed with 4% paraformaldehyde, sectioned, and stained with haematoxylin and eosin. The degree of inflammation in the distal part of colon was graded from 0 to 4 as follows: 0, normal; 1, ulcer with cell infiltration limited to the mucosa; 2, ulcer with limited cell infiltration in the submucosa; 3, focal ulcer involving all layers of the colon; 4, multiple lesions involving all layers of the colon, or necrotizing ulcer larger than 1 mm in length.

**Allergic diarrhoea.** BALB/c SPF adult mice were primed by subcutaneous injection with 1 mg of OVA (Fraction V; Sigma) in 100 μl of Complete Freund Adjuvant (CFA, DIFCO). One week after priming, mice were given 50 mg of OVA dissolved in 200 μl of PBS by intra-gastric administration three times per week. 17-mix or control PBS was orally administered to mice every 2 or 3 days for the entire period of the experiments. Diarrhoea was monitored visually 1 h after each oral OVA challenge. Diarrhoea was scored as follows: 0, normal faeces (solid); 1, moist faeces (semi-solid); 2, mild diarrhoea (loose); and 3, severe diarrhoea (watery). Serum was collected from the cheek vein 1 h after the last OVA challenge and OVA-specific IgE levels were measured by ELISA (Chondrex).

**Adoptive $CD4^+CD45RB^{hi}$ T-cell transfer model of colitis.** Germ-free CB.17 SCID mice were orally inoculated with SPF faeces together with or without 17-mix of Clostridia. One week later, exGF SCID mice received $4 \times 10^5$ $CD4^+CD45RB^{hi}$ T cells by intraperitoneal injection. Naive $CD4^+CD45RB^{hi}$ T cells were isolated from spleens of SPF BALB/c mice by FACS sorting. All the mice were observed daily and were killed on day 14 after T-cell transfer.


25. Morita, H. *et al.* An improved isolation method for metagenomic analysis of the microbial flora of the human intestine. *Microbes Environ.* **22,** 214–222 (2007).
26. Kim, S. W. *et al.* Robustness of gut microbiota of healthy adults in response to probiotic intervention revealed by high-throughput pyrosequencing. *DNA Res.* **20,** 241–253 (2013).
27. Noguchi, H., Taniguchi, T. & Itoh, T. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.* **15,** 387–396 (2008).
28. Jolley, K. A. *et al.* Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* **158,** 1005–1015 (2012).

LETTER

doi:10.1038/nature12347

# Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome

Shin Yoshimoto[1,2*], Tze Mun Loo[1,2,3], Koji Atarashi[4,5], Hiroaki Kanda[6], Seidai Sato[1,2], Seiichi Oyadomari[7], Yoichiro Iwakura[8], Kenshiro Oshima[9], Hidetoshi Morita[10], Masahira Hattori[9], Kenya Honda[2,4], Yuichi Ishikawa[6], Eiji Hara[1,2] & Naoko Ohtani[1,5*]

**Obesity has become more prevalent in most developed countries over the past few decades, and is increasingly recognized as a major risk factor for several common types of cancer[1]. As the worldwide obesity epidemic has shown no signs of abating[2], better understanding of the mechanisms underlying obesity-associated cancer is urgently needed. Although several events were proposed to be involved in obesity-associated cancer[1,3], the exact molecular mechanisms that integrate these events have remained largely unclear. Here we show that senescence-associated secretory phenotype (SASP)[4,5] has crucial roles in promoting obesity-associated hepatocellular carcinoma (HCC) development in mice. Dietary or genetic obesity induces alterations of gut microbiota, thereby increasing the levels of deoxycholic acid (DCA), a gut bacterial metabolite known to cause DNA damage[6]. The enterohepatic circulation of DCA provokes SASP phenotype in hepatic stellate cells (HSCs)[7], which in turn secretes various inflammatory and tumour-promoting factors in the liver, thus facilitating HCC development in mice after exposure to chemical carcinogen. Notably, blocking DCA production or reducing gut bacteria efficiently prevents HCC development in obese mice. Similar results were also observed in mice lacking an SASP inducer[8] or depleted of senescent HSCs, indicating that the DCA–SASP axis in HSCs has key roles in obesity-associated HCC development. Moreover, signs of SASP were also observed in the HSCs in the area of HCC arising in patients with non-alcoholic steatohepatitis[3], indicating that a similar pathway may contribute to at least certain aspects of obesity-associated HCC development in humans as well. These findings provide valuable new insights into the development of obesity-associated cancer and open up new possibilities for its control.**

Cellular senescence is a process occurring in normal cells in response to telomere erosion or oncogene activation, acting through checkpoint activation and stable cell-cycle arrest as a barrier to tumorigenesis[9,10]. Recent studies, however, reveal that senescent cells also develop a secretory profile composed mainly of inflammatory cytokines, chemokines and proteases, a typical signature termed the senescence-associated secretory phenotype (SASP)[4] or the senescence messaging secretome (SMS)[5], hereafter referred to as SASP. Some of the SASP factors have cell-autonomous activities that reinforce senescence cell-cycle arrest[5] and/or promote clearance of senescent cells[11,12], but other SASP factors have cell non-autonomous functions associated with inflammation and tumorigenesis promotion[4], indicating that SASP contributes positively and negatively to cancer development, depending on the biological context[4,5]. Because some of the SASP factors, such as IL-6 and PAI-1[4,5], are known to increase cancer risk in obesity[1,13], we propose that SASP may contribute to obesity-associated cancer.

To explore this possibility, we first set up a system to examine the impact of dietary obesity on tumorigenesis, using wild-type C57BL/6 mice. However, we were unable to detect a statistically significant difference in cancer development between obese mice fed a high-fat diet (HFD) and lean mice fed a normal diet (data not shown), implying that a certain level of oncogenic stimuli might be required for obesity-associated cancer, especially in wild-type mice maintained in a specific pathogen free (SPF) environment. Because the Ras-pathway is frequently activated in human cancers, including hepatocellular carcinoma (HCC)[14], we decided to use a treatment with DMBA (7,12-dimethylbenz(*a*)anthracene, a chemical carcinogen that causes an oncogenic Ras mutation) at the neonatal stage, a protocol known to generate a variety of tumours throughout the body[15]. In this setting, we also took advantage of using p21-*p*-luc mice, in which the expression of the $p21^{Waf1/Cip1}$ gene (a senescence inducer, also known as *Cdkn1a*) can be monitored noninvasively using a bioluminescence imaging (BLI) technique[16]. The neonatal p21-*p*-luc mice were therefore treated with a single application of DMBA, followed by feeding either HFD or normal diet for 30 weeks (Fig. 1a). Interestingly, a marked increase of the bioluminescent signal was observed in the abdomen of the obese mice, and it originated mainly from liver cancer (Supplementary Fig. 1). Notably, all HFD-fed mice developed HCC, whereas only 5% of -mice fed normal diet developed malignant tumours in lung, but not liver (Fig. 1b–e and Supplementary Fig. 2). Importantly, moreover, similar HCC development was also observed when genetically obese (*ob/ob*, also known as $Lep^{ob/ob}$) mice were treated with DMBA at the neonatal stage (Supplementary Fig. 3a–d), indicating that obesity, but not the HFD, promotes HCC development.

Because the induction of $p21^{Waf1/Cip1}$ expression was observed in liver, particularly in the area of liver cancer (Supplementary Fig. 1c), we speculated that senescent cells might be present in the vicinity of cancerous hepatocytes. Indeed, $p21^{Waf1/Cip1}$ expression was observed only in activated hepatic stellate cells (HSCs), which express α-smooth muscle actin (α-SMA) and desmin[7] (Fig. 1f). Furthermore, a number of other senescence markers[9,10], such as $p16^{INK4a}$ expression, signs of DNA damage (53BP1 foci and γH2AX foci) and inhibited cell proliferation (the absence of bromodeoxyuridine incorporation and Ki-67 expression), were also observed in activated HSCs despite absence of oncogenic *ras* mutation (Fig. 1f and Supplementary Figs 3e and 4). Interestingly, moreover, increased expression of IL-6, Gro-α, CXCL9 (major components of SASP)[4,5], but not HGF (a differentiation marker)[7], was observed in activated HSCs, but not in other types of liver cells (Fig. 1f and Supplementary Figs 3e and 5), indicating that these activated HSCs are senescing and may promote obesity-associated HCC development via SASP. It should be noted that, unlike the study using carbon tetrachloride ($CCl_4$)[11], fibrosis was not apparent in HFD-fed mice (Supplementary Fig. 6), precluding the possibility that the appearance of senescent HSCs was a by-product of liver fibrosis.

[1]Division of Cancer Biology, Cancer Institute, Japanese Foundation for Cancer Research, Koto-ku, Tokyo 135-8550, Japan. [2]CREST, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan. [3]Department of Applied Biological Science, Tokyo University of Science, Noda, Chiba 278-8510, Japan. [4]Research Center for Allergy and Immunology, RIKEN, Yokohama, Kanagawa 230-0045, Japan. [5]PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan. [6]Division of Pathology, Cancer Institute, Japanese Foundation for Cancer Research, Koto-ku, Tokyo 135-8550, Japan. [7]Institute for Genome Research, University of Tokushima, Tokushima 770-8503, Japan. [8]Research Institute for Biological Science, Tokyo University of Science, Noda, Chiba 278-8510, Japan. [9]Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba 277-8561, Japan. [10]School of Veterinary Medicine, Azabu University, Sagamihara, Kanagawa 229-8501, Japan.
*These authors contributed equally to this work.

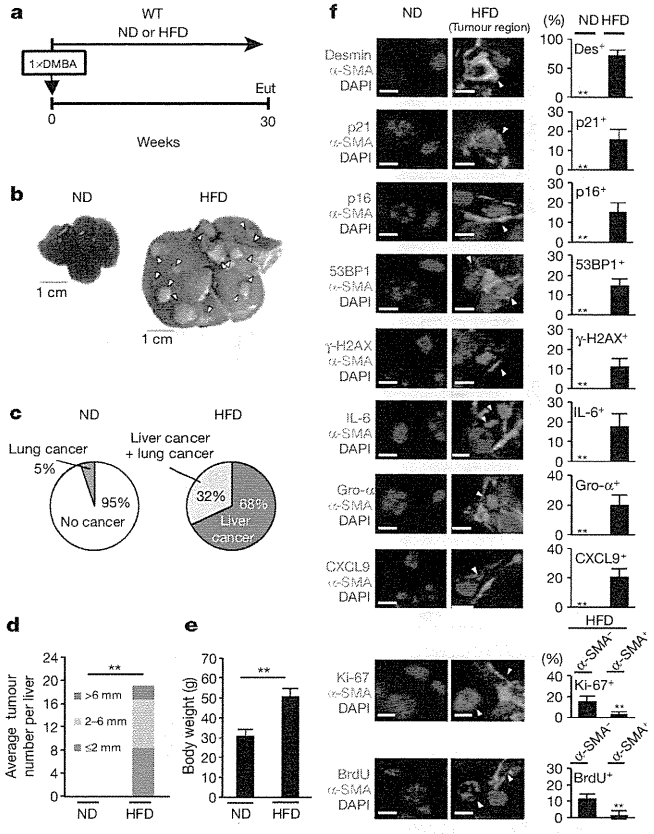


**Figure 1 | Cellular senescence in HSCs.** **a**, Timeline of the experimental procedure ($n = 19$ per group). Eut, euthanasia; ND, normal diet. **b**, Representative macroscopic photographs of livers. Arrowheads indicate HCCs. **c**, The ratios of cancer formation. **d**, The average liver tumour numbers and their relative size distribution. **e**, The average body weights at the age of 30 weeks. **f**, Immunofluorescence analysis of liver section. HSCs were visualized by α-SMA staining and DNA was stained by 4′,6-diamidino-2-phenylindole (DAPI). Scale bars, 2.5 μm. Arrowheads indicate α-SMA expressing cells that were positive for indicated markers. The histograms indicate the percentages of α-SMA-expressing cells that were positive for indicated markers. At least 100 cells were scored per group. For all graphs, error bars indicate mean ± standard deviation (s.d.). $^{**}P < 0.01$.

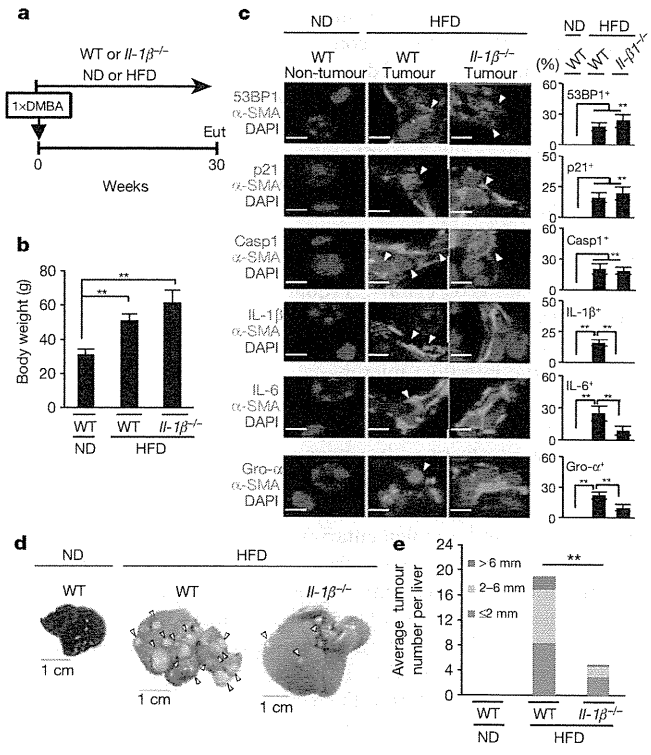


**Figure 2 | IL-1β deficiency alleviates obesity-induced HCC development.** **a**, Timeline of the experimental procedure (wild type (WT), $n = 19$; *Il-1β*$^{-/-}$, $n = 9$). **b**, The average body weights at the age of 30 weeks. **c**, Immunofluorescence analysis of liver sections. HSCs were visualized by α-SMA staining and DNA was stained by DAPI. Scale bars, 2.5 μm. The histograms indicate the percentages of α-SMA-expressing cells that were positive for indicated markers. At least 100 cells were scored per group. **d**, Representative macroscopic photographs of livers. Arrowheads indicate HCCs. **e**, The average liver tumour numbers and their relative size distribution. For all graphs, error bars indicate mean ± s.d. $^{**}P < 0.01$.

To ascertain the role of SASP in obesity-associated HCC development, we next sought evidence that the blockage of SASP can reduce obesity-associated HCC development. Although we were unable to detect the expression of IL-1α (an upstream regulator of SASP induction)[8] in HSCs, significant induction of IL-1β (a functional homologue of IL-1α) and its activator, caspase-1 (an essential component of the inflammasome), was observed in senescent HSCs (Fig. 2a–c). Moreover, the addition of recombinant IL-1β caused the dose-dependent induction of *Il-6* and *Gro-α* (also known as *Cxcl1*) gene expression in cultured primary murine HSCs (Supplementary Fig. 7a), indicating that inflammasome activation and subsequent IL-1β maturation can act as an upstream regulator of SASP induction in HSCs. Indeed, the levels of SASP factor expression in activated HSCs were substantially diminished in mice lacking the *Il-1β* gene (*Il-1β*$^{-/-}$ mice, also known as *Il1b*$^{-/-}$) (Fig. 2c), and the numbers and sizes of the liver tumours that developed in *Il-1β*$^{-/-}$ mice were strikingly reduced, as compared with wild-type mice (Fig. 2d, e), although the degree of steatohepatitis was not attenuated (Supplementary Fig. 8a, b). It should be noted, however, that other senescence markers, such as 53BP1 foci, p21$^{Waf1/Cip1}$ expression and inhibited cell proliferation, were still observed in the activated HSCs of *Il-1β*$^{-/-}$ mice (Fig. 2c and Supplementary Fig. 8c). These results are somewhat consistent with a recent observation that the expression of p21$^{Waf1/Cip1}$ can induce senescence cell-cycle arrest without SASP induction[17], suggesting that SASP, but not senescence cell-cycle arrest, promotes obesity-associated HCC development.

To further verify this idea, we next attempted to deplete senescent HSCs from obese wild-type mice treated with DMBA at the neonatal stage. As reported previously[18], an intravenous injection of liposomes carrying small interfering RNA (siRNA) against HSP47 substantially reduced the abundance of activated HSCs, coinciding with a significant reduction of HCC development (Supplementary Fig. 9a–f). Note that this was not accompanied by an attenuation of steatohepatitis (Supplementary Fig. 9g, h). These results, along with the data from *Il-1β*$^{-/-}$ mice (Fig. 2), strongly indicate that senescent HSCs have enhancing roles in HCC development via SASP, at least in the neonatal DMBA plus obesity-induced HCC model. It is also noteworthy that neither the deletion of the *Il-1β* gene nor the depletion of senescent HSCs caused appreciable weight loss (Fig. 2b and Supplementary Fig. 9c), implying that there may be an indirect link between obesity and HCC development, at least in this experimental setting. These observations then raised the question of how obesity provokes SASP in HSCs.

Emerging evidence has indicated that alterations of intestinal microbiota are associated with obesity[19]. Furthermore, the activation of toll-like receptor (TLR) 4 by lipopolysaccharide (LPS) from intestinal Gram-negative bacteria has been shown to promote HCC development, in an HCC model using DEN (diethyl nitrosamine) plus $CCl_4$ treatment[20]. We thus explored the possibility that intestinal bacteria have key roles in obesity-associated HCC development. Indeed, a

treatment with a well-established oral antibiotic cocktail (4Abx), which reduces the number of commensal intestinal bacteria[20], caused a marked reduction of HCC development, accompanied by a marked decrease in senescent HSCs in the neonatal DMBA plus obesity-induced HCC model (Fig. 3 and Supplementary Fig. 3). As reported[20], 4Abx treatment resulted in not only a > 99.5% reduction of the presence of bacterial 16S ribosomal RNA gene in faeces, but also an enlargement of caecum commonly observed in germ-free mice (Fig. 3b and data not shown). Unexpectedly, however, a slight increase, rather than decrease, in HCC development was observed in mice lacking the *Tlr4* gene (*Tlr4*$^{-/-}$) (Supplementary Fig. 10), indicating that LPS from Gram-negative bacteria is unlikely to promote HCC development in this setting. Indeed, meta 16S rRNA gene sequencing analysis of the intestinal microbiota revealed that the percentage of Gram-positive bacterial strains indigenous to the human and rodent intestinal tracts[6] was dramatically increased with a HFD (Fig. 4a). Moreover, a treatment with vancomycin (VCM), an antibiotic that preferentially targets Gram-positive bacteria, alone was sufficient to block HCC development and the appearance of senescent HSCs (Figs 3d–f, 4a and Supplementary Fig. 3). These results lead us to propose that the obesity-associated increase of Gram-positive bacteria may promote HCC development, presumably through the enterohepatic circulation of gut bacterial metabolites or toxins.

To substantiate this idea, the serum metabolites of HFD- and normal-diet-fed mice were analysed by liquid chromatography mass spectrometry (LC-MS). Interestingly, the level of deoxycholic acid (DCA), a secondary bile acid produced solely by the 7α-dehydroxylation of

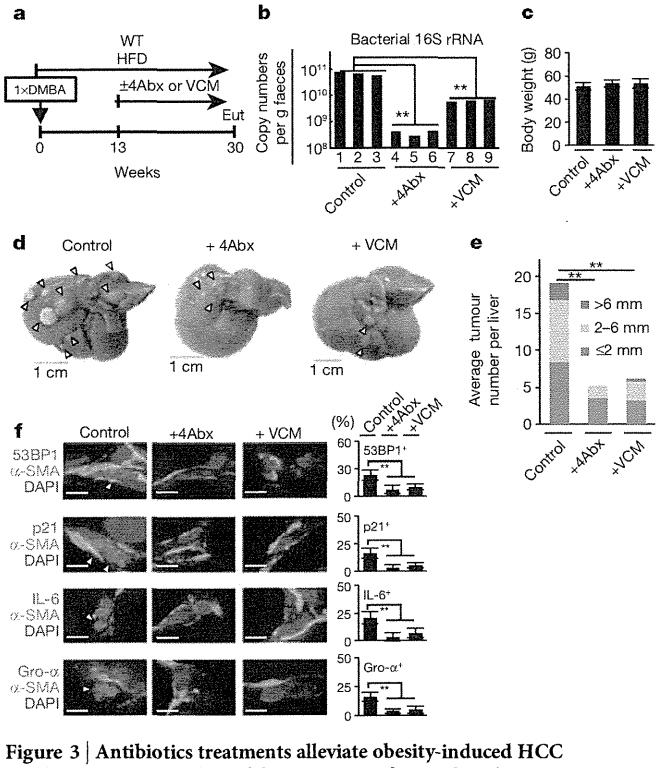


**Figure 3 | Antibiotics treatments alleviate obesity-induced HCC development.** **a**, Timeline of the experimental procedure (HFD, $n = 19$; HFD+4Abx, $n = 12$; HFD+VCM, $n = 6$). **b**, The copy number of intestinal bacteria in faeces of indicated mice. **c**, The average body weights at the age of 30 weeks. **d**, Representative macroscopic photograph of livers. Arrowheads indicate HCCs. **e**, The average tumour numbers and their relative size distribution. **f**, Immunofluorescence analysis of liver sections. HSCs were visualized by α-SMA staining and DNA was stained by DAPI. Scale bars, 2.5 μm. The histograms indicate the percentages of α-SMA expressing cells that were positive for indicated markers. At least 200 cells were scored per group. For all graphs, error bars indicate mean ± s.d. $^{**}P < 0.01$.

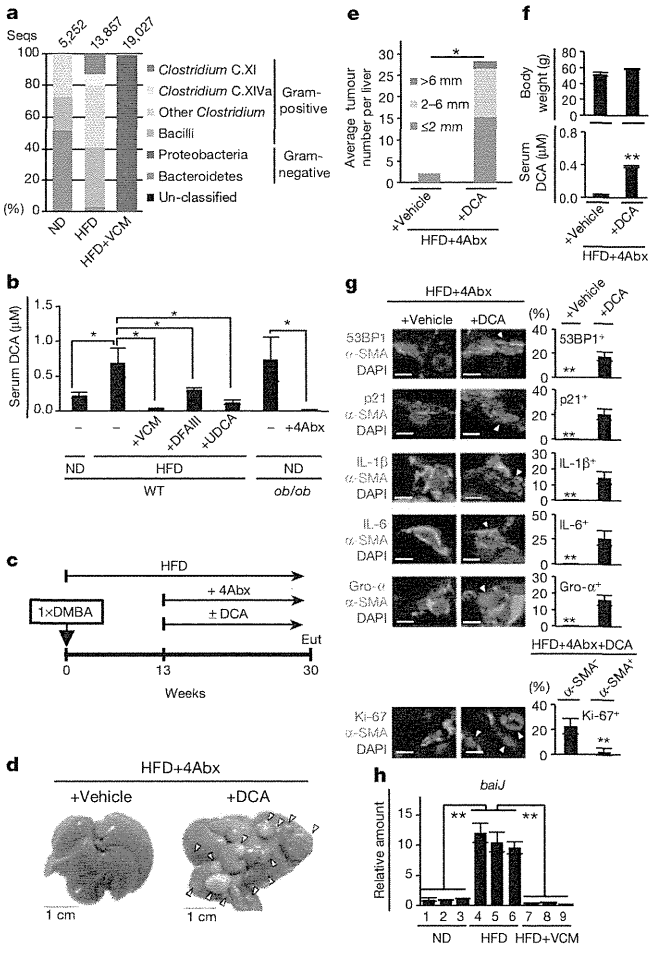


**Figure 4 | Bacterial metabolite promotes obesity-induced HCC development.** **a**, The relative abundance of OTUs (%) in the faecal bacterial community. Data are representative of five mice per group. **b**, Serum DCA concentration (ND, $n = 4$; HFD, $n = 6$; HFD+VCM, $n = 3$; HFD+DFAIII, $n = 3$; HFD+UDCA, $n = 3$; *ob/ob*, $n = 3$; *ob/ob*+4Abx, $n = 3$). Error bars indicate mean ± s.e.m. **c**, Timeline of the experimental procedure ($n = 3$ per group). **d**, Representative macroscopic photographs of livers. Arrowheads indicate HCCs. **e**, The average tumour numbers and their relative size distribution. **f**, The average body weight and serum DCA concentration. **g**, Immunofluorescence analysis of liver sections. Scale bars, 2.5 μm. The histograms indicate the percentages of α-SMA-expressing cells that were positive for indicated markers. At least 100 cells were scored per group. **h**, The quantitative real time PCR (qPCR) analysis of *baiJ* gene in the faeces (180 mg) of indicated mice used in **a**. For all graphs except **b**, error bars indicate mean ± s.d. $^{*}P < 0.05$, $^{**}P < 0.01$.

primary bile acids carried out by gut bacteria such as strains belonging to *Clostridium* cluster XI and XIVa[6] (VCM-sensitive Gram-positive bacteria), was substantially increased by the HFD feeding, and was reduced by antibiotic treatments (Figs 3a and 4b). Note that DCA is known to cause DNA damage through reactive oxygen species production[21] and DNA damage is a critical inducer of SASP[4,22]. Moreover, in addition to colon carcinogenesis[23], DCA has been shown to enhance liver carcinogenesis[24]. These notions prompted us to examine if DCA has key roles in obesity-associated HCC development. To this end, we attempted to lower the levels of DCA, by either decreasing the 7α-dehydroxylation activity with difructose anhydride III (DFA III)[25] or stimulating bile acid secretion with ursodeoxycholic acid (UDCA)[26]. Notably, lowering the DCA concentration substantially reduced HCC development, accompanied by a marked decrease in senescent HSCs in obese mice treated with DMBA at the neonatal stage (Fig. 4b and

Supplementary Figs 11 and 12). In a reciprocal set of experiments, we also assessed whether DCA-feeding enhances HCC development in mice treated with DMBA at the neonatal stage (Fig. 4c). Intriguingly, although DCA feeding alone was insufficient to enhance HCC development in lean mice fed a normal diet at 30 weeks (data not shown), a significant enhancement of HCC development (Fig. 4c–f), accompanied by the appearance of senescence cell-cycle arrest and SASP in HSCs (Fig. 4g), was observed when HFD-fed mice treated with 4Abx were fed DCA for 17 weeks.

Notably, operational taxonomic unit (OTU)-based bacterial diversity analysis (Fig. 4a), in conjunction with a quantitative PCR analysis (Supplementary Fig. 13), revealed that the population of cluster XI of the genus *Clostridium* was strikingly increased in HFD-fed mice. Interestingly, phylogenetic analysis of the bacterial OTUs revealed that the population of *Clostridium* cluster XI is composed of a single bacterial taxon (OTU-1105) close to the DCA-producing strain *Clostridium sordellii*, and represents more than 12% of the faecal bacteria in HFD-fed mice (Supplementary Fig. 14). Concordantly, moreover, the abundance of the *baiJ* gene, a gene involved in bile acid 7α-dehydroxylation[27], was remarkably increased in faeces of mice fed HFD and was reduced by VCM treatment (Fig. 4h). On the other hand, a bacterial taxon (OTU-154) close to other DCA producing strains belonging to *Clostridium* cluster XIVa (*Clostridium hylemonae* and *Clostridium scindens*) represents only 0.5% of the total faecal bacteria in HFD-fed mice (Supplementary Fig. 14). Thus, although other bacteria may also be involved here, the simplest explanation for our data is that OTU-1105 belonging to *Clostridium* cluster XI contribute to an increase in the DCA level at least to some extent in HFD-fed mice.

Finally, to further support and extend our murine data to human biology, we tested whether IL-1β treatment can induce SASP in cultured primary human HSCs. As in murine HSCs, the addition of recombinant IL-1β caused the induction of *Il-6* and *Il-8* (a functional homologue of murine *Gro-α*) gene expression in cultured primary human HSCs (Supplementary Fig. 7b). Importantly, moreover, signs of cellular senescence and SASP were also observed in the HSCs without serious fibrosis in the area of HCC arising in patients with non-alcoholic steatohepatitis (NASH)[3] (8 out of 26) (Supplementary Fig. 15). This is somewhat consistent with previous observations that replicative senescence of cultured human HSCs is accompanied by a pronounced inflammatory but less fibrogenic phenotype[28] and a certain percentage of NASH- associated HCC arose from the non-cirrhotic liver[29]. Unlike rodents, the human liver cannot 7α-hydroxylate DCA, forming cholic acid[6]. Hence, DCA can accumulate to very high levels (>50%) in the bile acid pool of humans[6]. These data, together with the previous observation that high fat consumption resulted in higher faecal DCA concentrations in healthy male volunteers (ages 20-60)[30], suggest that DCA-induced senescent HSCs may contribute to at least certain aspects of obesity-associated HCC development via SASP in humans as well.

It should be noted that although many of the perturbations, for example, the *Il-1β* knockout, antibiotics treatment and lower DCA levels, significantly prevent HCC development, residual HCCs were still observed with these perturbations (Figs 2e and 3e and Supplementary Figs 11c and 12c). These results, in conjunction with the observation that DCA-feeding alone was insufficient to enhance HCC development in lean mice fed a normal diet until at least 30 weeks (data not shown), imply that an additional factor associated with obesity may exist to promote obesity-associated HCC development. Nevertheless, combining published data[14,19,21,24,30] with our findings, it is clear that the increased levels of DCA produced by gut bacteria play key roles in the promotion of obesity-associated HCC development via provoking SASP in HSCs, at least in the neonatal DMBA plus obesity-induced HCC model (Supplementary Fig. 16). A greater understanding of the molecular mechanisms linking gut microbial metabolite to SASP will therefore provide valuable new insights into how to bypass this undesirable side effect of cellular senescence.

## METHODS SUMMARY

**Chemically-induced carcinogenesis.** DMBA treatments[15] consisted of a single application of 50 μl of a solution 0.5% DMBA (7,12-dimethylbenz [*a*]anthracene, Sigma) in acetone to the dorsal surface on postnatal day 4–5. Mother mice with pups were then fed either normal diet or HFD until weaning. At the age of 4 weeks old, pups were weaned and continuously fed either normal diet or HFD until euthanized.

**Bacterial 16S rRNA amplicon sequencing and analysis.** Bacterial genomic DNA was isolated from mice faeces, amplified for V1–V4 hypervariable regions of the 16S rRNA gene, and used for pyrosequencing analysis.

**Full Methods** and any associated references are available in the online version of the paper.

**Received 10 December 2012; accepted 4 June 2013.**
**Published online 26 June 2013.**

1. Khandekar, M. J., Cohen, P. & Spiegelman, B. M. Molecular mechanisms of cancer development in obesity. *Nature Rev. Cancer* **11,** 886–895 (2011).
2. Calle, E. E. & Kaaks, R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nature Rev. Cancer* **4,** 579–591 (2004).
3. Sun, B. & Karin, M. Obesity, inflammation, and liver cancer. *J. Hepatol.* **56,** 704–713 (2012).
4. Coppé, J. P. *et al.* Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor. *PLoS Biol.* **6,** 2853–2868 (2008).
5. Kuilman, T. & Peeper, D. S. Senescence-messaging secretome: SMS-ing cellular stress. *Nature Rev. Cancer* **9,** 81–94 (2009).
6. Ridlon, J. M., Kang, D. J. & Hylemon, P. B. Bile salt biotransformations by human intestinal bacteria. *J. Lipid Res.* **47,** 241–259 (2006).
7. Friedman, S. L. Hepatic stellate cells: protean, multifunctional, and enigmatic cells of the liver. *Physiol. Rev.* **88,** 125–172 (2008).
8. Orjalo, A. V., Bhaumik, D., Gengler, B. K., Scott, G. K. & Campisi, J. Cell surface-bound IL-1α is an upstream regulator of the senescence-associated IL-6/IL-8 cytokine network. *Proc. Natl Acad. Sci. USA* **106,** 17031–17036 (2009).
9. Collado, M. & Serrano, M. Senescence in tumours: evidence from mice and humans. *Nature Rev. Cancer* **10,** 51–57 (2010).
10. Kuilman, T., Michaloglou, C., Mooi, W. J. & Peeper, D. S. The essence of senescence. *Genes Dev.* **24,** 2463–2479 (2010).
11. Krizhanovsky, V. *et al.* Senescence of activated stellate cells limits liver fibrosis. *Cell* **134,** 657–667 (2008).
12. Kang, T. W. *et al.* Senescence surveillance of pre-malignant hepatocytes limits liver cancer development. *Nature* **479,** 547–551 (2011).
13. Park, E. J. *et al.* Dietary and genetic obesity promote liver inflammation and tumorigenesis by enhancing IL-6 and TNF expression. *Cell* **140,** 197–208 (2010).
14. Newell, P. *et al.* Ras pathway activation in hepatocellular carcinoma and anti-tumoral effect of combined sorafenib and rapamycin *in vivo*. *J. Hepatol.* **51,** 725–733 (2009).
15. Serrano, M. *et al.* Role of the *INK4a* locus in tumor suppression and cell mortality. *Cell* **85,** 27–37 (1996).
16. Ohtani, N. *et al.* Visualizing the dynamics of $p21^{Waf1/Cip1}$ cyclin-dependent kinase inhibitor expression in living animals. *Proc. Natl Acad. Sci. USA* **104,** 15034–15039 (2007).
17. Coppé, J.-P. *et al.* Tumor suppressor and aging biomarker $p16^{INK4a}$ induces cellular senescence without the associated inflammatory secretory phenotype. *J. Biol. Chem.* **286,** 36396–36403 (2011).
18. Sato, Y. *et al.* Resolution of liver cirrhosis using vitamin A-coupled liposomes to deliver siRNA against a collagen-specific chaperone. *Nature Biotechnol.* **26,** 431–442 (2008).
19. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444,** 1022–1023 (2006).
20. Dapito, D. H. *et al.* Promotion of hepatocellular carcinoma by the intestinal microbiota and TLR4. *Cancer Cell* **21,** 504–516 (2012).
21. Payne, C. M. *et al.* Deoxycholate induces mitochondrial oxidative stress and activates NF-κB through multiple mechanisms in HCT-116 colon epithelial cells. *Carcinogenesis* **28,** 215–222 (2007).
22. Takahashi, A. *et al.* DNA damage signaling triggers degradation of histone methyltransferases through $APC/C^{Cdh1}$ in senescent cells. *Mol. Cell* **45,** 123–131 (2012).
23. McGarr, S. E., Ridlon, J. M. & Hylemon, P. B. Diet, anaerobic bacterial metabolism, and colon cancer: a review of the literature. *J. Clin. Gastroenterol.* **39,** 98–109 (2005).
24. Kitazawa, S. *et al.* Enhanced preneoplastic liver lesion development under 'selection pressure' conditions after administration of deoxycholic or lithocholic acid in the initiation phase in rats. *Carcinogenesis* **11,** 1323–1328 (1990).
25. Minamida, K., Ohashi, M., Hara, H., Asano, K. & Tomita, F. Effects of ingestion of difructose anhydride III (DFA III) and the DFA III-assimilating bacterium *Ruminococcus productus* on rat intestine. *Biosci. Biotechnol. Biochem.* **70,** 332–339 (2006).
26. Beuers, U. Drug insight: mechanisms and sites of action of ursodeoxycholic acid in cholestasis. *Nat. Clin. Pract. Gastroenterol. Hepatol.* **3,** 318–328 (2006).
27. Ridlon, J. M. & Hylemon, P. B. Identification and characterization of two bile acid coenzyme A transferases from *Clostridium scindens*, a bile acid 7α-dehydroxylating intestinal bacterium. *J. Lipid Res.* **53,** 66–76 (2012).

28. Schnabl, B., Purbeck, G. A., Choi, Y. H., Hagedorn, C. H. & Brenner, D. A. Replicative senescence of activated human hepatic stellate cells is accompanied by a pronounced inflammatory but less fibrogenic phenotype. *Hepatology* **37,** 653–664 (2003).
29. Takuma, Y. & Nouso, K. Nonalcoholic steatohepatitis-associated hepatocellular carcinoma: our case series and literature review. *World J. Gastroenterol.* **16,** 1436–1441 (2010).
30. Rafter, J. J. *et al.* Cellular toxicity of fecal water depends on diet. *Am. J. Clin. Nutr.* **45,** 559–563 (1987).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The authors thank M. Oshima for suggestions in antibiotics treatment and members of the Hara laboratory for discussion during the preparation of this manuscript. This work was supported by grants from Japan Science and Technology Agency (JST), Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT), Ministry of Health, Labour and Welfare of Japan (MHLW), Uehara Memorial Foundation and Takeda Science Foundation. S.Y. was partly supported by a postdoctoral fellowship from the Japan Society for Promotion of Science (JSPS). T.M.L. was partly supported by an international scholarship from the Ajinomoto Scholarship Foundation.

**Author Contributions** E.H. and N.O. designed the experiments, analysed the data and wrote the manuscript. N.O., S.Y. and T.M.L. performed obesity-induced liver cancer experiments. K.A., K.O., H.M., M.H. and K.H. performed bacterial genome data analysis. H.K., S.S. and Y.I. performed histopathological analysis of mouse and human liver cancer specimens. S.O. performed metabolite analysis. Y.I. provided *Il-1β*$^{-/-}$ mice. E.H. oversaw the projects.

**Author Information** Bacterial 16S rRNA amplicon sequence data have been deposited in DDBJ (http://www.ddbj.nig.ac.jp/index-e.html) with the accession number DRA000952. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.H. (eiji.hara@jfcr.or.jp).

## METHODS

**Mice and diet.** The *p21-p-luc* mice (CD1)[16] were backcrossed with C57BL/6 mice for eight generations. The leptin-deficient (*ob/ob*) mice (C57BL/6) were purchased from Charles River Laboratories Japan, Inc. *Tlr4*$^{-/-}$ mice (C57BL/6) were purchased from Oriental Bioservices. *Il-1β*$^{-/-}$ mice (C57BL/6) were provided by Y. Iwakura[31]. Male mice were used for all the experiments in this study. The mice were maintained under specific pathogen-free (SPF) conditions, on a 12-h light-dark cycle, and fed normal diet (CE-2 from CLEA Japan Inc., composed of 12 kcal% fat, 29 kcal% protein, 59 kcal% carbohydrates) or high-fat diet (HFD, D12492 from Research Diets Inc., composed of 60 kcal% fat, 20 kcal% protein, 20 kcal% carbohydrates) *ad libitum*. Mice with more than 45 g weight at the age of 30 weeks old were used as obese mice for all the experiments. We measured the amount of food our mice eat and found that a 50 g HFD mouse eats 3.44 g food a day. This equates to 1.2 g of fat per day or 24 g per kg. According to the Reagan-Shaw equation[32] (human equivalent dose (mg kg$^{-1}$) = mouse dose (mg kg$^{-1}$) × mouse $K_m$ factor ÷ human $K_m$ factor; where the mouse and human $K_m$ factors are 3 and 37, respectively), this is equivalent to a 70 kg human eating 136 g of fat a day. The sample size used in this study was determined based on the expense of data collection, and the need to have sufficient statistical power. Randomization and blinding were not used in this study. All animal experiments were cared for by protocols approved by the Committee for the Use and Care of Experimental Animals of the Japanese Foundation for Cancer Research (JFCR).

**Chemically induced carcinogenesis.** DMBA treatments[15] consisted of a single application of 50 μl of a solution 0.5% DMBA (7,12-dimethylbenz(*a*)anthracene, Sigma) in acetone to the dorsal surface on postnatal day 4–5. After this application, mother mice with pups were fed normal diet or HFD. At the age of 4 weeks old, pups were weaned and continuously fed either normal diet or HFD until euthanized. Evaluation of tumour number and size was determined by counting the number of visible tumours and measuring the size of the tumour.

**Bioluminescence imaging.** Bioluminescence imaging was performed as previously described[16,33]. In brief, for the detection of luciferase expression, mice were anesthetized, injected intraperitoneally with D-luciferin sodium salt (75 mg kg$^{-1}$) 5 min before beginning photon recording. Mice were placed in the light-tight chamber and a grey-scale image of the mice was first recorded with dimmed light followed by acquisition of luminescence image using a cooled charged-coupled device (CCD) camera (PIXIS 1024B; Princeton Instruments). The signal-to-noise ratio was increased by 2 × 2 binning and 5 min exposure. For colocalization of the luminescent photon emission on the animal body, grey scale and pseudo-colour images were merged by using IMAGE-PRO PLUS (Media Cybernetics).

**Antibiotics treatment.** Antibiotics treatment was performed as previously described[20] using a combination of four antibiotics (4Abx) of ampicillin (1 g l$^{-1}$), neomycin (1 g l$^{-1}$), metronidazole (1 g l$^{-1}$) and vancomycin (500 mg l$^{-1}$), or vancomycin (500 mg l$^{-1}$) alone (VCM) in drinking water at the age of 13 weeks old until killed.

**Histology and immunofluorescence analysis.** Haematoxylin and eosin staining and immunofluorescence analysis were performed as previously described[16]. The primary antibodies used for mouse samples were as follows: α-SMA (Sigma A5228), desmin (abcam ab15200), p21 (abcam ab2961), p16 (Santa Cruz sc1207), 53BP1 (Santa Cruz sc22760), γ-H2AX (CST 9718), IL-6 (abcam ab6672), Gro-α (abcam ab17882), Ki-67 (Thermo RM9106), bromodeoxyuridine (abcam ab6326), caspase-1 (Millipore 06-503), IL-1β (R&D systems AF-401-NA), HSP47 (Santa Cruz sc8352), CXCL9 (abcam ab137792), F4/80 (Invitrogen BM8) and CD45(Millipore 05-1416). The primary antibodies used for human samples were as follows: α-SMA (Dako M0851), γ-H2AX (CST 9718), p16 (Santa Cruz sc56330), p21 (CST #2947), IL-6 (abcam ab6672), IL-8 (abcam ab18672), 53BP1 (Santa Cruz sc22760) and caspase-1 (Millipore 06-503).

**Quantitative PCR.** Total RNA was extracted from mouse tissues using TRIzol reagent (Life technologies) and reverse transcription and quantitative PCR were performed as previously described[22]. Primers used were as follows: human *GAPDH*, 5′-CAACTACATGGTTTACATGTTC-3′ (forward) and 5′-GCCAGTGGACT CCACGAC-3′ (reverse), mouse *Gapdh*, 5′-CAACTACATGGTCTACATGTTC-3′ (forward) and 5′-CACCAGTAGACTCCACGAC-3′ (reverse), human *IL-6*, 5′-CTCGACGGCATCTCAGCCCTGA-3′ (forward) and 5′-CTGCCAGTGCCTC TTTGCTGCTTT-3′ (reverse), mouse *Il-6*, 5′-TGATTGTATGAACAACGATG ATGC-3′ (forward) and 5′-GGACTCTGGCTTTGTCTTTCTTGT-3′ (reverse), human *IL-8*, 5′-AAGGAAAACTGGGTGCAGAG-3′ (forward) and 5′-ATTGC ATCTGGCAACCCTAC-3′ (reverse), mouse *Gro-α*, 5′-GCTGGGATTCACC TCAAGAA-3′ (forward) and 5′-AGGTGCCATCAGAGCAGTCT-3′ (reverse), bacterial *baiJ* 5′-TCAGGACGTGGAGGCGATCCA-3′ (forward) and 5′-TACRTGATACTGGTAGCTCCA-3′ (reverse), *Clostridium* cluster XI 16S rRNA gene 5′-TGACGGTACYYNRKGAGGAAGCC-3′ (forward) and 5′-ACT ACGGTTRAGCCGTAGCCTTT-3′ (reverse).

***In vivo* RNAi experiment.** 250 μl of siRNA solution (3 mg ml$^{-1}$) against HSP47 or control siRNA were mixed with 250 μl of complexation buffer and 500 μl of Invivofectamine (Life Technologies), incubated for 30 min at 50 °C, and dialysed at room temperature for 2 h in 1 l of PBS (pH 7.4). Dialysed siRNA-Invivofectamine complex was collected and 3 μg per g (weight) was injected through mice's tail vein twice a week for 15 weeks until killed. The sequences of HSP47 targeting oligo are as follows. 5′-GCACUGCUUGUGAACGCCAU GUUCU-3′ (sense), 5′-AGAACAUGGCGUUCACAAGCAGUGC-3′ (antisense). As a negative control, Ambion *In vivo* Negative Control #1 siRNA(4457289) was used.

**Treatment with DCA, UDCA and DFAIII.** Deoxycholic acid (DCA) was dissolved in absolute ethanol and diluted in 66% propylene glycol to reduce the concentration of alcohol to 5%. HFD-fed mice treated with DMBA at neonatal stage were fed a combination of four antibiotics (4Abx) with 40 μg per g (weight) of DCA or vehicle (control) three times per week using a plastic feeding tube at the age of 13 weeks old until killed. Ursodeoxycholic acid (UDCA) tablets (Tanabe-Mitsubishi Pharma) were powdered and dissolved in 66% propylene glycol. HFD-fed mice treated with DMBA at neonatal stage were fed 60 μg per g (weight) of UDCA or vehicle (control) using a plastic feeding tube every day at the age of 15 weeks old until killed. Difructose anhydride III (DFAIII)was dissolved in saline. HFD-fed mice treated with DMBA at neonatal stage were fed 0.1 mg per g (weight) of DFAIII or vehicle (control) using a plastic feeding tube every day at the age of 17 weeks old until killed.

**Bacterial 16S rRNA amplicon sequencing and analysis.** Bacterial genomic DNA was isolated from faeces using a QIAamp DNA Stool mini kit (QIAGEN), and 100 ng of DNA was used for PCR for V1–V4 hyper variable regions of the 16S rRNA gene. Twenty five cycles of amplification was performed with universal 16S rRNA primers 27F 5′-AGAGTTTGATCCTGGCTCAG-3′ and 519R 5′-GWATTACCGCGGCKGCTG-3′ with 10-bp barcode tags using KOD Fx plus DNA polymerase (TOYOBO). All amplicons were sequenced on a 454 Genome Sequencer FLX Titanium platform (Roche Diagnostics and Beckman Coulter Genomics). Quality filter-passed sequence reads were obtained by removing reads that had no both primer sequences, had less than 500 bp in length, had the average quality value (QV) < 25, or were possible chimaeric. Of the filter-passed reads, more than 2,500 sequence reads trimming off both primer sequences for each sample were used and subjected to OTU analysis with the cutoff similarity of 97% identity using QIIME software. Representative sequences from each OTU were blasted to the database in Ribosomal Database Project (RDP) and aligned. The obtained OTU sequences were grouped at class level[34,35] and were subjected to phylogenetic analysis using MEGA software as described previously[35].

**Determination of the copy number of faecal bacteria.** The copy number of faecal bacteria was calculated from the standard curve of known bacterial copy number by quantitative real-time PCR of 16S rRNA gene using 341f, 5′-CCTACGGGAGGC AGCAG-3′ and 534r 5′-ATTACCGCGGCTGCTGG-3′ primers as described previously[36].

**Measurement of serum ALT and AST.** The levels of serum alanine aminotransferase (ALT) and aspartate aminotransferase (AST) were measured using kits from WAKO Pure Chemical Industries, Ltd.

**Measurement of serum deoxycholic acid.** The metabolomic analysis of mice serum were performed by liquid chromatograph mass spectrometry(LC-MS) in Human Metabolome Technologies Inc. Japan as previously described[37]. The amount of serum DCA was measured by gas chromatograph mass spectrometry (GC-MS) in the Bile Acid Institute of Junshin Clinic, Japan as described[38].

**Human subjects.** Informed consent was obtained from all patients according to the protocol approved by the ethics committee of the Japanese Foundation for Cancer Research (JFCR).

**Statistical analysis.** Data were analysed by unpaired *t*-test with Welch correction (two-side) or Mann–Whitney test (two-side). *P*- values less than 0.05 were considered significant.

**Cell culture.** Murine primary HSCs were isolated as previously described[18,39], and were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum in 3% $O_2$ and 5% $CO_2$ condition. Human primary HSCs were purchased from Health Science Research Resources Bank and were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum in 3% $O_2$ and 5% $CO_2$ condition.

***H-ras* sequencing.** Total RNA was prepared from HCCs and HSCs isolated from tumour regions using TRIzol reagent (Invitrogen). RNA was converted to cDNA by using oligo (dT) primer and a 330-bp PCR fragment containing exon 2 of *H-ras* gene was amplified with 5′-TGGGGCAGGAGCTCCTGGAT-3′ and 5′-GAA GGACTTGGTGTTGTTGA-3′ primers. PCR fragments were sub-cloned using Target Clone Plus system (TOYOBO) and were sequenced by using Dye-Terminator and Big-Dye cycle sequencing system (Applied Biosystems) as described previously[33].

31. Horai, R. *et al.* Production of mice deficient in genes for interleukin (IL)-1α, IL-1β, IL-1α/β, and IL-1 receptor antagonist shows that IL-1β is crucial in turpentine-induced fever development and glucocorticoid secretion. *J. Exp. Med.* **187,** 1463–1475 (1998).
32. Reagan-Shaw, S., Nihal, M. & Ahmad, N. Dose translation from animal to human studies revisited. *FASEB J.* **22,** 659–661 (2008).
33. Yamakoshi, K. *et al.* Real-time *in vivo* imaging of $p16^{Ink4a}$ reveals cross talk with p53. *J. Cell Biol.* **186,** 393–407 (2009).
34. Collins, M. D. *et al.* The phylogeny of the genus *Clostridium*: proposal of five new general and eleven new species combinations. *Int. J. Syst. Bacteriol.* **44,** 812–826 (1994).
35. Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* **331,** 337–341 (2011).
36. Song, Y., Liu, C. & Finegold, S. M. Real-time PCR quantitation of clostridia in feces of autistic children. *Appl. Environ. Microbiol.* **70,** 6459–6465 (2004).
37. Ooga, T. *et al.* Metabolomic anatomy of an animal model revealing homeostatic imbalances in dyslipidaemia. *Mol. Biosyst.* **7,** 1217–1223 (2011).
38. Muto, A. *et al.* Detection of $\Delta^4$-3-oxo-steroid 5β-reductase deficiency by LC–ESI-MS/MS measurement of urinary bile acids. *J. Chromatogr. B* **900,** 24–31 (2012).
39. Sekiya, Y. *et al.* Down-regulation of cyclin E1 expression by microRNA-195 accounts for interferon-β-induced inhibition of hepatic stellate cell proliferation. *J. Cell. Physiol.* **226,** 2535–2542 (2011).

# CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature13004

## Corrigendum: Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome

Shin Yoshimoto, Tze Mun Loo, Koji Atarashi, Hiroaki Kanda, Seidai Sato, Seiichi Oyadomari, Yoichiro Iwakura, Kenshiro Oshima, Hidetoshi Morita, Masahira Hattori, Kenya Honda, Yuichi Ishikawa, Eiji Hara & Naoko Ohtani

*Nature* **499,** 97–101 (2013); doi:10.1038/nature12347

In this Letter, the forename of author Masahira Hattori was spelled incorrectly as 'Masahisa'. It has been corrected in the HTML and PDF versions of the manuscript.

OPEN ACCESS Freely available online

PLOS ONE

# Genomic Adaptation of the *Lactobacillus casei* Group

**Hidehiro Toh[1☯], Kenshiro Oshima[2☯], Akiyo Nakano[3], Muneaki Takahata[3], Masaru Murakami[3], Takashi Takaki[4], Hidetoshi Nishiyama[4], Shizunobu Igimi[5], Masahira Hattori[2], Hidetoshi Morita[3]***

**1** Medical Institute of Bioregulation, Kyushu University, Higashi-ku, Fukuoka, Japan, **2** Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan, **3** School of Veterinary Medicine, Azabu University, Sagamihara, Kanagawa, Japan, **4** JEOL Ltd., Akishima, Tokyo, Japan, **5** Division of Biomedical Food Research, National Institute of Health Sciences, Kamiyoga, Setagaya, Tokyo, Japan

## Abstract

*Lactobacillus casei*, *L. paracasei*, and *L. rhamnosus* form a closely related taxonomic group (*Lactobacillus casei* group) within the facultatively heterofermentative lactobacilli. Here, we report the complete genome sequences of *L. paracasei* JCM 8130 and *L. casei* ATCC 393, and the draft genome sequence of *L. paracasei* COM0101, all of which were isolated from daily products. Furthermore, we re-annotated the genome of *L. rhamnosus* ATCC 53103 (also known as *L. rhamnosus* GG), which we have previously reported. We confirmed that ATCC 393 is distinct from other strains previously described as *L. paracasei*. The core genome of 10 completely sequenced strains of the *L. casei* group comprised 1,682 protein-coding genes. Although extensive genome-wide synteny was found among the *L. casei* group, the genomes of ATCC 53103, JCM 8130, and ATCC 393 contained genomic islands compared with *L. paracasei* ATCC 334. Several genomic islands, including carbohydrate utilization gene clusters, were found at the same loci in the chromosomes of the *L. casei* group. The *spaCBA* pilus gene cluster, which was first identified in GG, was also found in other strains of the *L. casei* group, but several *L. paracasei* strains including COM0101 contained truncated *spaC* gene. ATCC 53103 encoded a higher number of proteins involved in carbohydrate utilization compared with intestinal lactobacilli, and extracellular adhesion proteins, several of which are absent in other strains of the *L. casei* group. In addition to previously fully sequenced *L. rhamnosus* and *L. paracasei* strains, the complete genome sequences of *L. casei* will provide valuable insights into the evolution of the *L. casei* group.

**Citation:** Toh H, Oshima K, Nakano A, Takahata M, Murakami M, et al. (2013) Genomic Adaptation of the *Lactobacillus casei* Group. PLoS ONE 8(10): e75073. doi:10.1371/journal.pone.0075073

**Editor:** Joseph Schacherer, University of Strasbourg, France

**Received** February 16, 2013; **Accepted** August 10, 2013; **Published** October 8, 2013



**Funding:** This research was supported by the Promotion and Mutual Aid Corporation for Private Schools of Japan, Grant-in-Aid for Matching Fund Subsidy for Private Universities, and a research project grant awarded by the Azabu University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: morita@azabu-u.ac.jp

☯ These authors contributed equally to this work.

## Introduction

The genus *Lactobacillus* is the largest group of the family *Lactobacteriaceae* and contains more than 130 species. The species *Lactobacillus casei*, *L. paracasei*, and *L. rhamnosus* are phylogenetically and phenotypically closely related and are regarded together as the *Lactobacillus casei* group within the facultatively heterofermentative lactobacilli [1]. The classification and nomenclature of this group are controversial [2–7]. Some strains of *L. casei*, *L. paracasei*, and *L. rhamnosus* have for long been used as probiotics in a wide range of different products marketed in many countries. *L. casei* and *L. paracasei* have also been isolated from a variety of environmental habitats, including raw and fermented dairy (especially cheese) and plant materials (e.g., wine, pickle, silage, and kimchi). They are used as acid-producing starter cultures in milk fermentation as adjunct cultures for intensification and for acceleration of flavor development in bacterial-ripened cheeses. They are commonly the dominant species of nonstarter lactic acid bacteria in ripening cheese.

In the *L. casei* group, the genomes of five *L. paracasei* strains (ATCC 334, BD-II, BL23, LC2W, and Zhang) and three *L. rhamnosus* strains (ATCC 53103, Lc 705, and ATCC 8530) have been fully sequenced to date [8–14]. We have also previously reported the complete genome sequence of *L. rhamnosus* ATCC 53103 [15]. *L. rhamnosus* GG, the original strain of *L. rhamnosus* ATCC 53103, was isolated from a healthy human intestinal flora, and is one of the most widely used and well-documented probiotics, which confer a health benefit on the host when administered in adequate amounts [16]. It has been reported that *L. rhamnosus* GG can shorten the duration of infectious diarrhea, reduce antibiotic-associated symptoms, and alleviate food allergy and atopic dermatitis in children [16].

In this paper, we present the complete genome sequences of *L. casei* ATCC 393 and *L. paracasei* JCM 8130 (also known as ATCC 25302), which were isolated from a cheese and milk product, respectively, and the draft genome sequence of *L. paracasei* COM0101 isolated from a commercial fermented milk product. Furthermore, we re-annotated the genome of *L. rhamnosus* ATCC 53103. We then compared sequenced genomes of the *L. casei* group to gain a broader view of the genetic variability within the group. Comparison of the genome sequences of strains isolated from the human gut and dairy products can provide valuable insights into the lifestyle adaptation of the *L. casei* group.

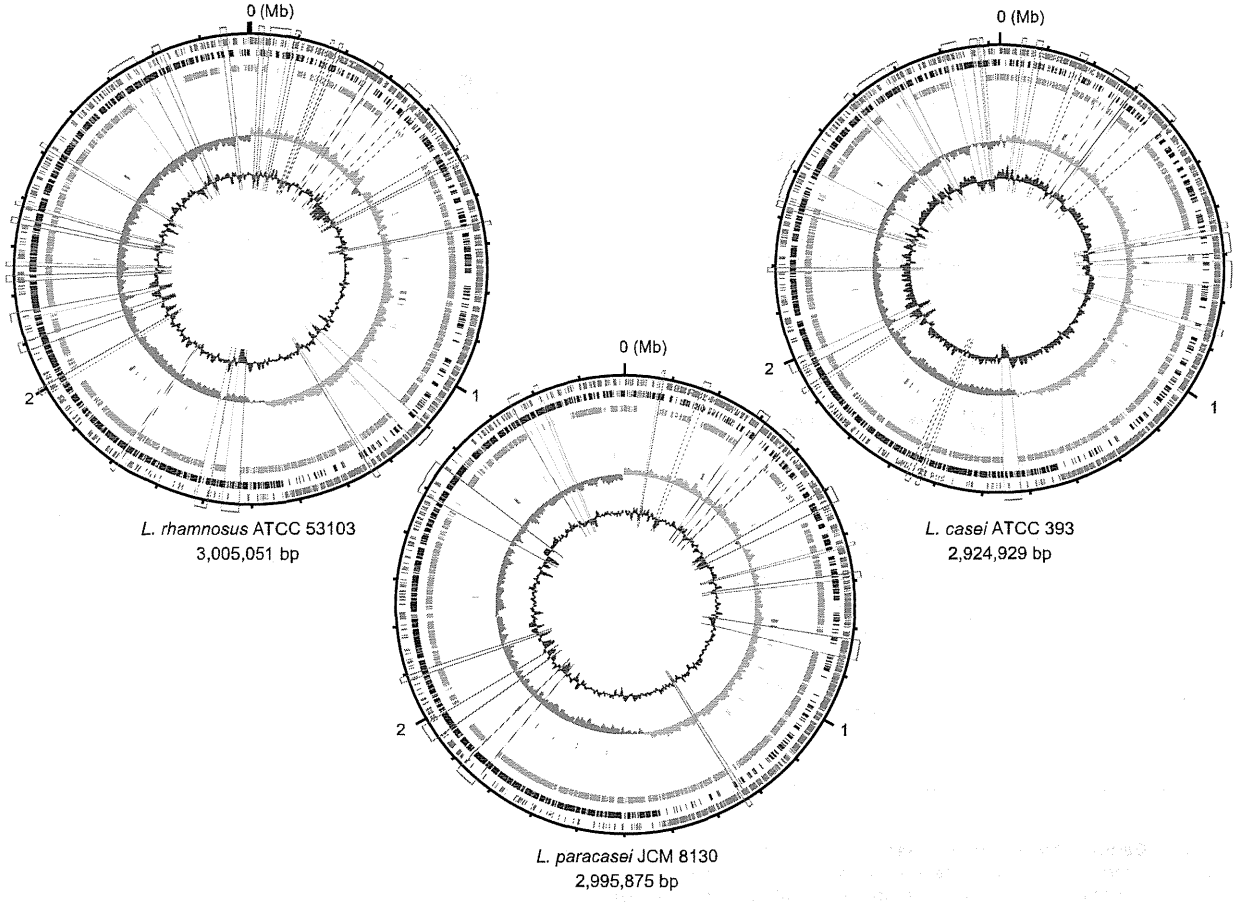


**Figure 1. Circular representations of the chromosomes of *L. rhamnosus* ATCC 53103, *L. paracasei* JCM 8130, and *L. casei* ATCC 393.** From the outside: circles 1 and 2 of the chromosome show the positions of protein-coding genes on the positive and negative strands, respectively. Circle 3 shows the positions of protein-coding genes that are shared among the 10 completely sequenced genomes of the *L. casei* group. Circle 4 shows the positions of tRNA genes (orange) and rRNA genes (blue). Circle 5 shows a plot of GC skew [(G − C)/(G+C); orange indicates values >0; blue indicates values <0]. Circle 6 shows a plot of G+C content (outward: higher values than the average). The genomic islands in each strain are boxed: regions including carbohydrate utilization gene cluster (pink), prophage-like regions (green), and the others (blue).
doi:10.1371/journal.pone.0075073.g001

## Materials and Methods

### Genome Sequencing

*L. paracasei* JCM 8130 and *L. casei* ATCC 393 were obtained from the Japan Collection of Microorganisms (JCM) and the American Type Culture Collection (ATCC), respectively. In this study, ten strains of putative *L. paracasei* isolated from the fermented milk product Yakult (Yakult Ltd., Japan) exhibited the same pattern by random amplification of polymorphic DNA fingerprinting [17]. We thus selected one *L. paracasei* strain designated as COM0101 for sequencing. *L. paracasei* JCM 8130, *L. casei* ATCC 393, and *L. paracasei* COM0101 were cultured in MRS (deMan, Rogosa and Sharpe) broth (Difco) at 37°C for 24 h, and the genomic DNAs were isolated and purified as previously described [18].

The genome sequences of *L. paracasei* JCM 8130, *L. casei* ATCC 393, and *L. paracasei* COM0101 were determined by the whole-genome shotgun strategy using Sanger sequencing (3730xl DNA sequencers) and 454 pyrosequencing (GS-FLX sequencers). We generated 19,200 (3.9-fold, 3730xl) and 284,003 (25.7-fold, GS-FLX) sequences from the *L. paracasei* JCM 8130 genome, 28,416 (5.9-fold, 3730xl) sequences from the *L. casei* ATCC 393 genome, and 131,707 (15.4-fold, GS-FLX) sequences from the *L. paracasei* COM0101 genome. The 454 pyrosequencing reads were assembled using the Newbler assembler software. A hybrid assembly of 454 and Sanger reads was performed using the Phred-Phrap-Consed program. Gap closing and re-sequencing of low-quality regions were conducted by Sanger sequencing to obtain the high-quality finished sequence. The overall accuracy of the finished sequence was estimated to have an error rate of <1 per 10,000 bases (Phrap score of ≥40). The deep sequencing dataset of *L. paracasei* JCM 8130 and *L. paracasei* COM0101 are deposited in the DDBJ/GenBank/EMBL Sequence Read Archive under the accession numbers DRA000955 and DRA000956, respectively.

### Informatics

An initial set of predicted protein-coding genes was identified using Glimmer 3.0 [19]. Genes consisting of <120 base pairs (bp) and those containing overlaps were eliminated. All predicted proteins were searched against a non-redundant protein database (nr, NCBI) using BLASTP with a bit-score cutoff of 60. The start codon of each protein-coding gene was manually refined from BLASTP alignments. The tRNA genes were predicted by the

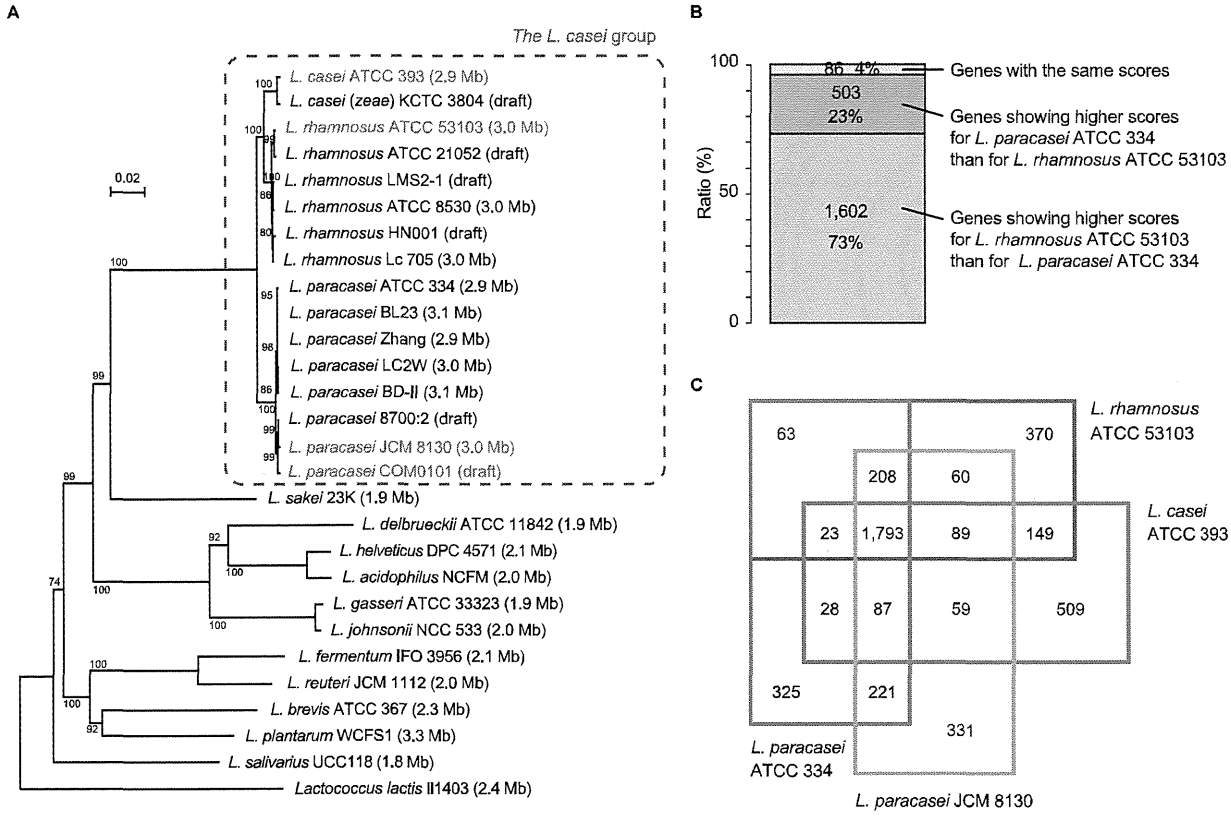


**Figure 2. Genome-based phylogenetic analysis of the *L. casei* group.** (**A**) Phylogenetic relationships between the genomes of sequenced lactobacilli inferred from 34 concatenated ribosomal protein amino acid sequences. The scale bar represents an evolutionary distance. Sequences were aligned with ClustalW with a bootstrap trial of 1,000 and bootstrap values (%) are indicated at the nodes. An unrooted tree was generated using NJplot. The chromosome size is shown in parentheses. (**B**) Three-way comparisons between *L. casei* ATCC 393 with *L. rhamnosus* ATCC 53103 and *L. paracasei* ATCC 334. The 2,191 genes shared by the three strains were classified into three categories on the basis of the BLAST score ratio analysis [23]. (**C**) Venn diagram comparing the gene inventories of four strains of the *L. casei* group. Data resulted from reciprocal BLASTP analysis. The numbers of shared and unique genes are shown.
doi:10.1371/journal.pone.0075073.g002

tRNAscan-SE [20], and the rRNA genes were detected by BLASTN search using known *Lactobacillus* rRNA sequences as queries. Protein domains were identified using HMMER with the Pfam database. Orthology across whole genomes has been determined using BLASTP reciprocal best hits in all-against-all comparisons of amino acid sequences. Two sequences were identified as highly conserved orthologs if their BLAST score ratio is more than 0.8. When two genome sequences were compared using BLASTN, non-matching regions were predicted as genomic islands. The presence of an N-terminal signal peptide sequence was predicted using the SignalP [21]. Clustered regularly interspaced short palindromic repeats (CRISPR) were predicted using the CRISPRFinder [22]. Draft genome sequences of *L. rhamnosus* ATCC 21052 (accession no. AFZY01000000), *L. rhamnosus* HN001 (ABWJ00000000), *L. rhamnosus* LMS2-1 (ACIZ00000000), *L. paracasei* 8700:2 (ABQV00000000), and *L. casei* (*zeae*) KCTC 3804 (BACQ01000000) were obtained from GenBank.

The complete genome sequences of *L. paracasei* JCM 8130, *L. casei* ATCC 393, and *L. rhamnosus* ATCC 53103 are deposited in the DDBJ/GenBank/EMBL database under the accession numbers AP012541–AP012543, AP012544–AP012546, and AP011548, respectively. The draft genome sequence of COM0101 has been deposited in public database under the accession numbers BAGT01000001–BAGT01000184.

## Results and Discussion

### Comparative Genome Analysis within the *L. casei* Group

We first re-annotated the genome of *L. rhamnosus* ATCC 53103, which we previously reported in the short paper [15]. Next, we determined and annotated the complete genome sequences of *L. paracasei* JCM 8130 and *L. casei* ATCC 393. The genome of *L. paracasei* JCM 8130 consists of a circular chromosome of 2,995,875 bp and two plasmids, and that of *L. casei* ATCC 393 consists of a circular chromosome of 2,924,929 bp and two plasmids (Fig. 1). The chromosomes of *L. paracasei* JCM 8130 and *L. casei* ATCC 393 contained 2,848 and 2,737 predicted protein-coding genes, respectively. The larger plasmid (27 kilobases [kb]) of ATCC 393 shared 14 genes, such as beta-galactosidase and cystathionine beta-synthase, with a 65-kb plasmid (accession no. FM179324) of *L. rhamnosus* Lc 705 (Fig. S1), thus indicating that both plasmids may be derived from the same origin. Furthermore, we generated a draft genome sequence of *L. paracasei* COM0101 that consists of 184 contigs (>500 bp) with a total length of 3,003,364 bp. The COM0101 genome contained 2,767 predicted protein-coding genes. One of the highly redundant contigs

contained a gene for plasmid replication protein that showed 100% amino acid identity with that of *L. paracasei* strains, indicating that the COM0101 genome probably has at least one plasmid. Their chromosome sizes (2.9–3.0 megabases [Mb]) were among the largest group in the *Lactobacillus* genomes, with an average size of 1.8–2.0 Mb (Fig. 2A). General features of these genomes are summarized in Table S1.

We constructed a phylogenetic tree for concatenated sequences of ribosomal proteins from sequenced *Lactobacillus* (Fig. 2A). *L. casei* ATCC 393 and the *L. casei–paracasei* phylum were found to form a distinct clade within the *L. casei* group, and *L. casei* ATCC 393 was shown to be closer to *L. casei* (*zeae*) KCTC 3804. A three-way comparison between the genomes of *L. casei* ATCC 393, *L. rhamnosus* ATCC 53103, and *L. paracasei* ATCC 334 using the BLAST score ratio analysis [23] revealed a greater number of proteins in *L. casei* ATCC 393 showing a high score for *L. rhamnosus* ATCC 53103 than those showing a high score for *L. paracasei* ATCC 334 (Fig. 2B). Moreover, *L. casei* ATCC 393 shared more genes with *L. rhamnosus* ATCC 53103 than with *L. paracasei* ATCC 334 (Fig. 2C). We thus found that *L. casei* ATCC 393 is more closely related to *L. rhamnosus* strains than to *L. paracasei* strains based on the phylogeny, overall protein similarities, and number of shared genes. This result supports the previous reports that *L. casei* ATCC 393 is distinct from other strains previously described as *L. paracasei* [2,3,5,6]. Furthermore, we also constructed a multi-locus sequence typing (MLST)-based phylogenetic tree [24] for *L. paracasei* strains (Fig. S2A), showing that COM0101 shares the same MLST lineage with BL23, LC2W, and BD-II. Moreover, COM0101 shared more genes with BL23 than with ATCC 334 and JCM 8130 (Fig. S2B). Thus, COM0101 is phylogenetically closely related to BL23, LC2W, and BD-II in *L. paracasei* strains.

We compared the genomes of *L. rhamnosus* ATCC 53103, *L. paracasei* JCM 8130, *L. casei* ATCC 393, and *L. paracasei* ATCC 334 (Fig. 2C). Thus, 1,793 genes were common to the four strains, and a total of 4,315 ortholog clusters were assigned to the pan-genome of the four strains. Of the 1,793 core genes, 1,682 (94%) were also conserved among the other six completely sequenced strains (BD-II, BL23, LC2W, Zhang, Lc 705, and ATCC 8530) of the *L. casei* group. Broadbent *et al.* (2012) showed that 1,715 protein-coding genes were common to 17 sequenced *L. casei* strains [25]. These results suggest that approximately 1,700 genes constitute the core genome of the *L. casei* group, likely inherited from their common ancestor. All dispensable protein-coding genes, which were found in one or more but not all the 10 completely sequenced strains of the *L. casei* group, were functionally classified based on the clusters of orthologous groups from the NCBI COGs database, and the gene repertoires were compared (Fig. S3). There was a considerable difference in the number of genes assigned to COG category G (carbohydrate transport and metabolism) and category L (replication, recombination, and repair) among the strains. *L. rhamnosus* strains had a lower number of genes assigned to COG category L because the *L. rhamnosus* genomes contained a lower number of transposase genes compared with the other strains, suggesting that insertion element-mediated genome diversification is less frequent in *L. rhamnosus* strains. In contrast, *L. paracasei* JCM 8130 and *L. casei* ATCC 393 contained a higher number of transposase genes. Most of the genes assigned to COG category G were encoded in hypervariable regions in the genomes of the *L. casei* group (described later). We next classified all protein-coding genes of *L. rhamnosus* ATCC 53103 and sequenced intestinal lactobacilli on the basis of the COGs database (Fig. 3A). *L. rhamnosus* ATCC 53103 contained a higher number of genes assigned to COG category G compared with intestinal lactobacilli. The abundance of genes related to carbohydrate transport and metabolism in *L. rhamnosus* ATCC 53103 may contribute to the wide variety of qualities in this strain compared with other probiotics.

Bacteriocins are small antimicrobial peptides produced widely by lactic acid bacteria. The *L. rhamnosus* ATCC 53103 genome encoded the bacteriocin gene cluster (LRHM_2289 to LRHM_2312), which contained genes encoding the two-component sensor and regulator, four bacteriocin immunity proteins, ATP-binding cassette (ABC) transporter with the proteolytic domain, and small peptides. The cluster was conserved in the genomes of the *L. casei* group, but in the corresponding region of *L. casei* ATCC 393, a gene for bacteriocin ABC transporter was interrupted by transposase (LBCZ_2129 to LBCZ_2133) and genes for immunity proteins were absent, suggesting that *L. casei* ATCC 393 may not be able to produce bacteriocin.

CRISPRs, along with their associated *cas* genes, are known to constitute a defense system against the propagation of phages and plasmids; these were observed in the genomes of a number of lactic acid bacteria [26]. *L. rhamnosus* ATCC 53103 contained a CRISPR region (2,260,261–2,261,880) and four CRISPR-associated genes (LRGG_2116 to LRGG_2119). The 36-bp-long sequence was present 25 times and separated by 30-bp unique spacer sequences. It has been reported that two distinct types (LsaI1 and LdbuI) of CRISPR loci were identified in the *L. casei* genomes [25]. *L. casei* strains BD-II, BL23, LC2W, and Zhang also have an LsaI1-type CRISPR region at the same locus on the chromosome, suggesting that the ancestral strain of the *L. casei* group had encoded a CRISPR region. However, the 36-bp repeat sequence of the four *L. casei* strains differs by two bases from that of *L. rhamnosus* ATCC 53103, and the number of the repeat sequences was different (17–22) among these strains. COM0101 has the orthologs of the four CRISPR-associated genes, indicating that COM0101 also may have a CRISPR region. In contrast, *L. paracasei* JCM 8130, *L. casei* ATCC 393, *L. rhamnosus* Lc 705, and *L. rhamnosus* ATCC 8530 had no CRISPR, suggesting that these strains may have lost a CRISPR region during adaptation to their environment where phage detection is not essential.

## Genomic Islands

Whole-genome alignment showed a high level of synteny among the strains of the *L. casei* group (Fig. S4). A previous report showed that there was a high degree of synteny among the genomes of 17 *L. casei* strains [25]. These results indicate that strains of the *L. casei* group have a stable genome structure. However, each genome contained specific genes, many of which were grouped in clusters as genomic islands (GIs). It has been reported that the comparison of the genomes of *L. paracasei* ATCC 334 and BL23 revealed 12 and 19 GIs (>5 kb) in ATCC 334 and BL23, respectively [27]. Similarly, we identified 26 GIs (>5 kb) in *L. rhamnosus* ATCC 53103 that were not conserved in *L. paracasei* ATCC 334 (a cheese isolate) (Table 1, Fig. 1). The 26 genomic islands of *L. rhamnosus* ATCC 53103 included six carbohydrate utilization gene clusters (regions −1 to −6), four of which were completely or partially present in *L. paracasei* BL23, whose ecological origin is unclear. This result supports the previous findings that cheese isolates, including *L. paracasei* ATCC 334, have undergone significant gene decay, including loss of many genes involved in carbohydrate utilization [25,27]. Thus, *L. paracasei* ATCC 334 contains a lower number of genes related to carbohydrate transport and metabolism compared with the other sequenced *L. paracasei* strains (Fig. S3). In probiotic lactobacilli, horizontal gene transfer played an important role in shaping the common ancestor [28]. Such acquisition of new genes can expand a bacterium's potential for adaptation to a new niche. The

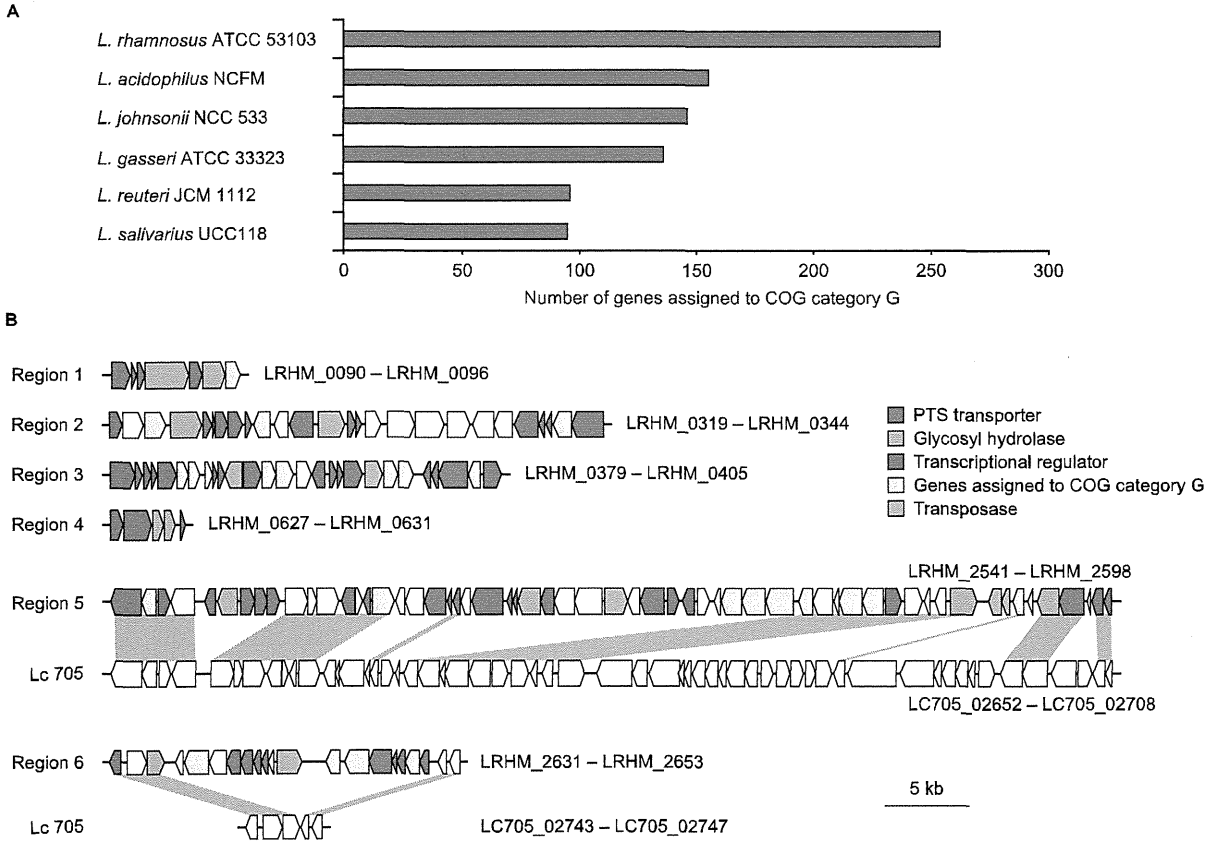


**Figure 3. Abundance of genes related to carbohydrate transport and metabolism in *L. rhamnosus* ATCC 53103.** (**A**) Comparative analysis by functional categories of the gene repertoires of sequenced intestinal lactobacilli. The number of genes assigned to COG category G in each genome is shown. (**B**) Carbohydrate utilization gene clusters of *L. rhamnosus* ATCC 53103. Genes and their orientations are depicted with arrows. Regions -5 and -6 are compared with the corresponding genomic locations in *L. rhamnosus* Lc 705. Gray bars indicate orthologous regions.
doi:10.1371/journal.pone.0075073.g003

common ancestor of *L. rhamnosus* ATCC 53103 and *L. paracasei* ATCC 334 seems to have acquired carbohydrate utilization gene clusters via lateral gene transfer. These carbohydrate utilization gene clusters may have provided adaptive features to some strains including ATCC 53103 for their survival and proliferation in the human intestine. In contrast, these carbohydrate utilization gene clusters may have been lost in the lineage to ATCC 334 during its adaptation to the cheese environment.

Similarly, compared with *L. paracasei* ATCC 334, 15 and 24 GIs were found in *L. paracasei* JCM 8130 and *L. casei* ATCC 393, respectively (Table 1, Fig. 1). Of these GIs, 6 (JCM 8130) and 10 (ATCC 393) were found at the same loci with those of *L. rhamnosus* ATCC 53103. A comparative genome hybridization in 22 *L. casei* strains isolated from various habitats has revealed 25 hypervariable regions [27], of which 11 were found at the same loci of the GIs in *L. rhamnosus* ATCC 53103. Thus, these results suggest that the chromosomes of the *L. casei* group contain several hypervariable regions at the same loci.

The six carbohydrate utilization gene clusters of *L. rhamnosus* ATCC 53103 contained the genes for phosphoenolpyruvate-carbohydrate phosphotransferase (PTS)-type transporter systems, glycosyl hydrolases, transcriptional regulators, and other carbohydrate-related proteins (Fig. 3B). *L. rhamnosus* ATCC 53103 encoded 28 complete PTS-type transporter systems, 11 of which were encoded adjacent to genes for glycosyl hydrolase and transcriptional regulator, thereby allowing localized transcriptional control. The organization (carbohydrate transporter, glycosyl hydrolase, and transcriptional regulator) is reminiscent of the many clusters found in *Bifidobacterium longum* [29].

Six of the 26 GIs of *L. rhamnosus* ATCC 53103 overlapped with all the hypervariable regions among the sequenced *L. rhamnosus* strains (ATCC 53103, Lc 705, ATCC 8530, ATCC 2105, HN001, and LMS2-1). Three of the six hypervariable regions were prophage-like regions (LRHM_1038 to LRHM_1090, LRHM_1455 to LRHM_1475, and LRHM_2779 to LRHM_2794 in ATCC 53103). The other three regions corresponded to regions containing carbohydrate utilization gene clusters (regions -3, -5, and -6), indicating that *L. rhamnosus* strains show flexibility in sugar utilization. Two of the five PTS-type transporter systems in region-5 and two in region-6 were missing in Lc 705, ATCC 8530, and LMS2-1 strains (Fig. 3B). Comparative genomic hybridization analyses have showed that the region corresponding to regions -5 and -6 contains an overrepresentation of genes involved in carbohydrate utilization and transcriptional regulation in 22 *L. casei* strains [27]. Taken together, the region corresponding to regions -5 and -6 in the genomes of the *L. casei* group may be required to fine-tune its ability to utilize carbohydrates.

**Table 1.** Genomic islands in *L. rhamnosus* ATCC 53103, *L. paracasei* JCM 8130, and *L. casei* ATCC 393.

| *L. rhamnosus* ATCC 53103 | | | *L. paracasei* JCM 8130 | | | *L. casei* ATCC 393 | | |
|---|---|---|---|---|---|---|---|---|
| **Locus** | **Size (kb)** | **Product description** | **Locus** | **Size (kb)** | **Product description** | **Locus** | **Size (kb)** | **Product description** |
| LRHM_0019–LRHM_0031 | 12.5 | ammonium transporter protein, hypothetical protein | LBPC_0071–LBPC_0078 | 6.1 | conserved hypothetical protein | LBCZ_0041–LBCZ_0051 | 8.3 | hypothetical protein, transposase |
| LRHM_0044–LRHM_0073 | 39.7 | fibronectin-binding protein, beta-glucuronidase, 2-dehydro-3-deoxygluconokinase, mannonate dehydratase, fructuronate reductase | LBPC_0157–LBPC_0170 | 15.5 | conserved hypothetical protein | LBCZ_0065–LBCZ_0076 | 12.6 | transposase, conserved hypothetical protein |
| LRHM_0086–LRHM_0096 | 10.5 | carbohydrate utilization gene cluster (region-1) | LBPC_0276–LBPC_0297 | 23.8 | carbohydrate utilization gene cluster | LBCZ_0159–LBCZ_0174 | 15.7 | myo-inositol catabolism protein |
| LRHM_0149–LRHM_0156 | 6.1 | carbohydrate transporter, two-component system | LBPC_0331–LBPC_0359 | 30.6 | PTS transporter, amino acid ABC transporter | LBCZ_0223–LBCZ_0252 | 33.8 | carbohydrate utilization gene cluster |
| LRHM_0172–LRHM_0177 | 6.6 | taurine ABC transporter | LBPC_0470–LBPC_0499 | 33.4 | hypothetical protein | LBCZ_0277–LBCZ_0286 | 8.9 | conserved hypothetical protein |
| LRHM_0256–LRHM_0268 | 14.8 | myo-inositol catabolism protein | LBPC_0579–LBPC_0584 | 6.7 | PTS transporter, 6-phospho-beta-galactosidase | LBCZ_0338–LBCZ_0388 | 41.1 | prophage region I |
| LRHM_0319–LRHM_0350 | 34.6 | carbohydrate utilization gene cluster (region-2) | LBPC_0636–LBPC_0648 | 12.1 | prophage region I | LBCZ_0605–LBCZ_0617 | 11.9 | hypothetical protein |
| LRHM_0376–LRHM_0466 | 97.8 | carbohydrate utilization gene cluster (region-3), amino acid ABC transporter, beta-N-acetylglucosaminidase, N-acylamino acid racemase, cell surface protein, transposase | LBPC_0763–LBPC_0817 | 41.2 | prophage region II | LBCZ_0620–LBCZ_0675 | 39.9 | prophage region II |
| LRHM_0493–LRHM_0500 | 8.0 | hypothetical protein | LBPC_1168–LBPC_1176 | 9.2 | conserved hypothetical protein | LBCZ_0685–LBCZ_0742 | 54.2 | prophage region III |
| LRHM_0624–LRHM_0631 | 7.3 | carbohydrate utilization gene cluster (region-4) | LBPC_1739–LBPC_1789 | 42.7 | prophage region III | LBCZ_0821–LBCZ_0832 | 13.4 | prophage region IV |
| LRHM_1038–LRHM_1090 | 39.7 | prophage region I | LBPC_1864–LBPC_1906 | 36.7 | prophage region IV | LBCZ_1343–LBCZ_1372 | 36.1 | prophage region V |
| LRHM_1192–LRHM_1199 | 10.4 | amino acid transporter, hypothetical protein | LBPC_1988–LBPC_1998 | 10.9 | glycosyltransferase, transposase | LBCZ_1552–LBCZ_1559 | 7.8 | truncated formate C-acetyltransferase, transcriptional regulator |
| LRHM_1455–LRHM_1484 | 36.5 | prophage region II | LBPC_2364–LBPC_2427 | 70.9 | glycosyltransferase, cell surface protein, conserved hypothetical protein | LBCZ_1571–LBCZ_1577 | 9.9 | putative cell surface protein |
| LRHM_1518–LRHM_1530 | 24.9 | cell surface protein, glycosyltransferase | LBPC_2603–LBPC_2630 | 27.3 | carbohydrate utilization gene cluster | LBCZ_1817–LBCZ_1825 | 9.7 | glycosyltransferase |
| LRHM_1699–LRHM_1703 | 7.7 | cell surface protein | LBPC_2661–LBPC_2670 | 8.4 | putative cell surface protein, transposase | LBCZ_1857–LBCZ_1870 | 23.6 | conserved hypothetical protein |
| LRHM_1877–LRHM_1891 | 13.2 | conserved hypothetical protein, transposase | | | | LBCZ_2040–LBCZ_2046 | 9.2 | conserved hypothetical protein |
| LRHM_1959–LRHM_1977 | 19.3 | glycosyltransferase | | | | LBCZ_2167–LBCZ_2179 | 12.2 | conserved hypothetical protein, ABC transporter |
| LRHM_2012–LRHM_2019 | 16.4 | conserved hypothetical protein | | | | LBCZ_2185–LBCZ_2247 | 81.7 | putative cell surface protein, conjugative transposon protein |