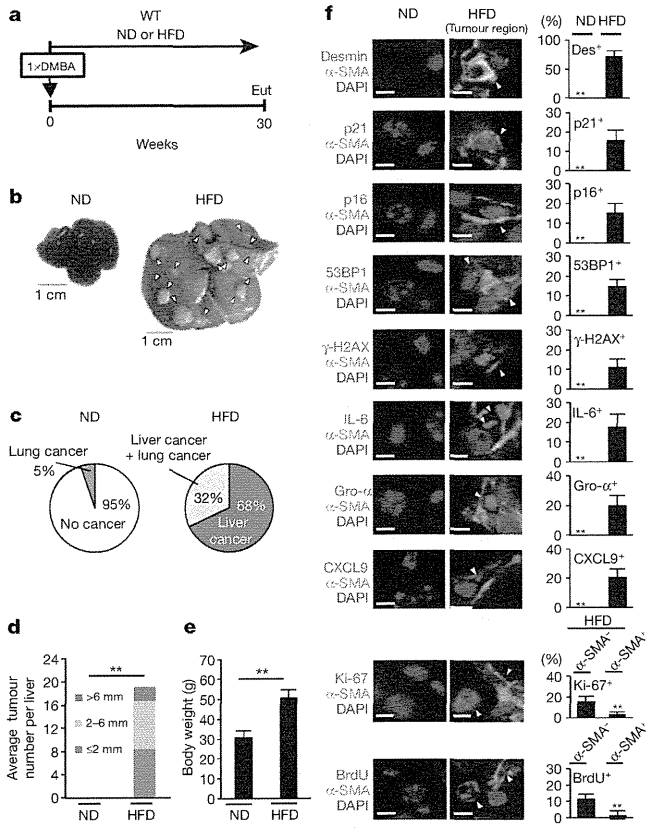

Figure 1 | Cellular senescence in HSCs. **a**, Timeline of the experimental procedure ($n = 19$ per group). Eut, euthanasia; ND, normal diet. **b**, Representative macroscopic photographs of livers. Arrowheads indicate HCCs. **c**, The ratios of cancer formation. **d**, The average liver tumour numbers 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.). $^{**}P < 0.01$.

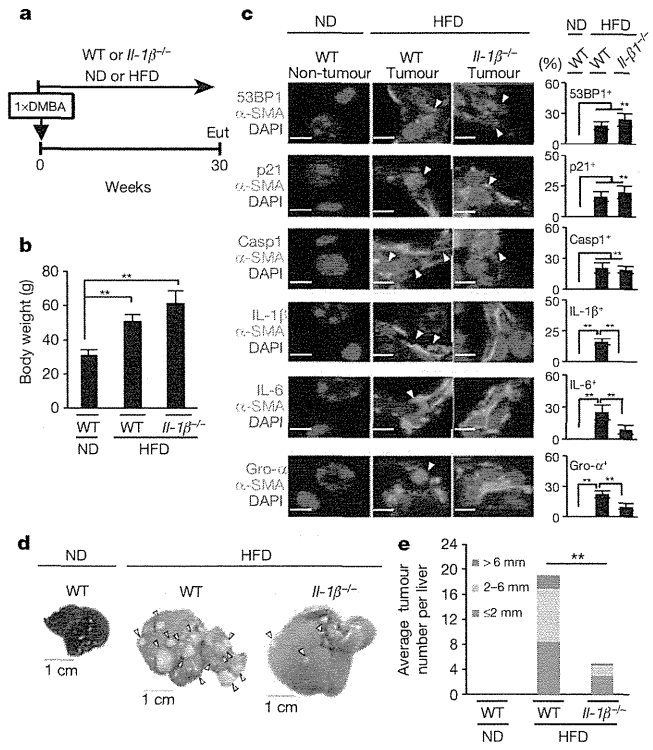

Figure 2 | IL-1β deficiency alleviates obesity-induced HCC development. **a**, Timeline of the experimental procedure (wild type (WT), $n = 19$; *Il-1β* $^{-/-}$, $n = 9$). **b**, The average body weights at the age of 30 weeks. **c**, Immunofluorescence analysis of liver sections. HSCs were visualized 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

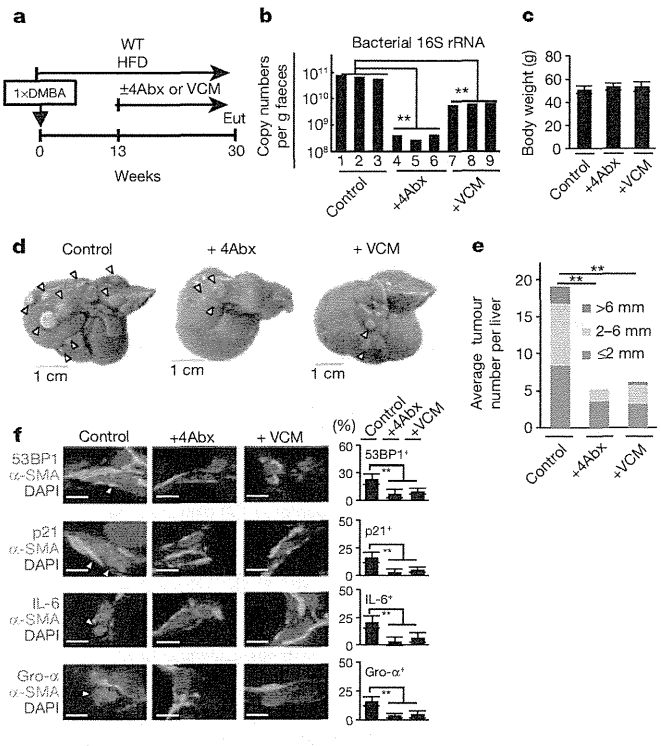

Figure 3 | Antibiotics treatments alleviate obesity-induced HCC development. **a**, Timeline of the experimental procedure (HFD, $n = 19$; HFD+4Abx, $n = 12$; HFD+VCM, $n = 6$). **b**, The copy number of intestinal bacteria in faeces of indicated mice. **c**, The average body weights at the age of 30 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. $^{**}P < 0.01$.

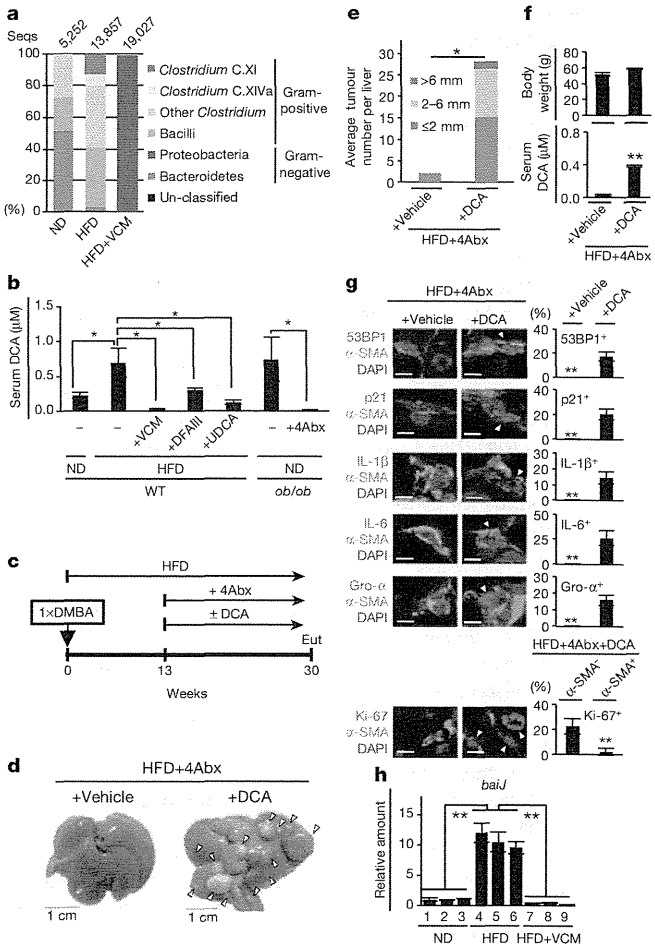

Figure 4 | Bacterial metabolite promotes obesity-induced HCC development. **a**, The relative abundance of OTUs (%) in the faecal bacterial community. Data are representative of five mice per group. **b**, Serum DCA concentration (ND, $n = 4$; HFD, $n = 6$; HFD+VCM, $n = 3$; HFD+DFAIII, $n = 3$; HFD+UDCA, $n = 3$; *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 Δ^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.

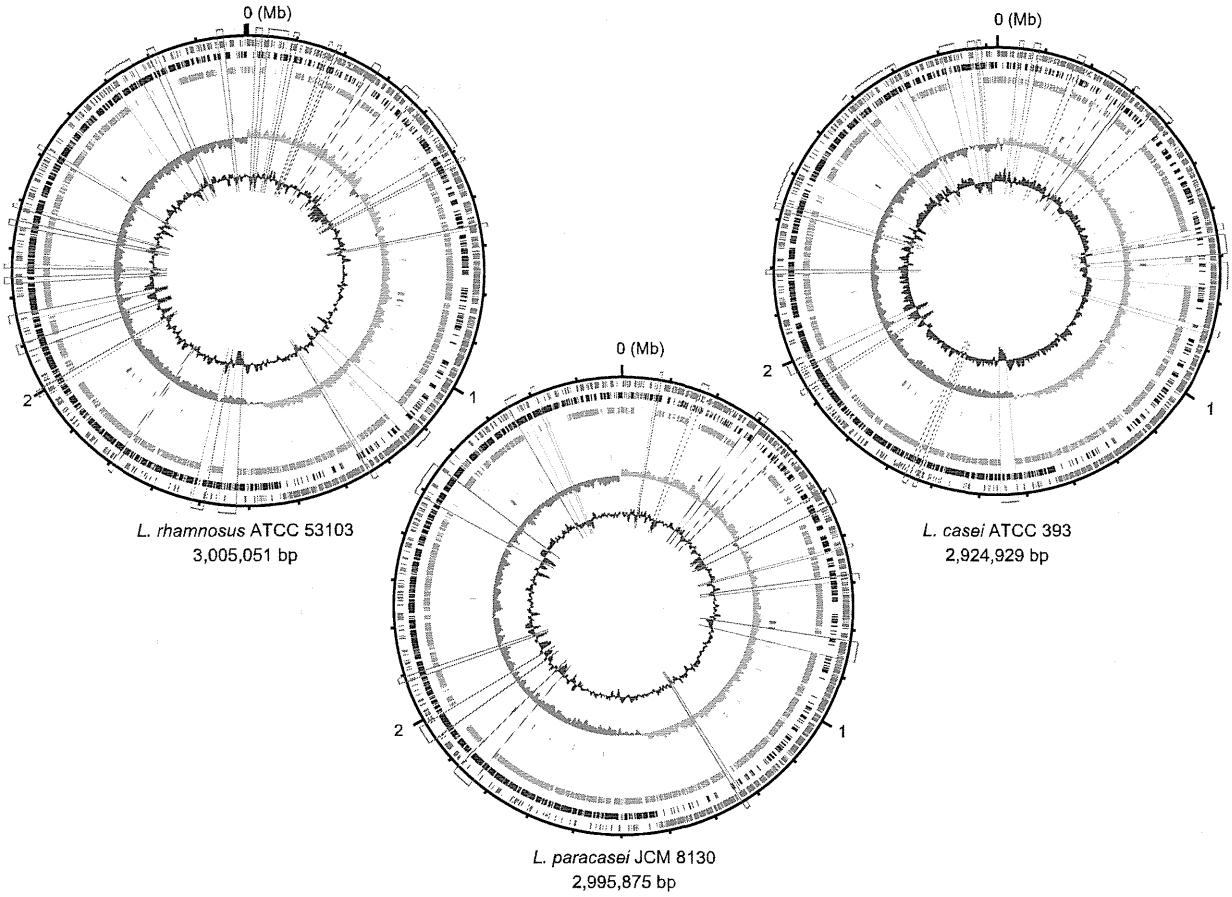

Figure 1. Circular representations of the chromosomes of *L. rhamnosus* ATCC 53103, *L. paracasei* JCM 8130, and *L. casei* ATCC 393. From the outside: circles 1 and 2 of the chromosome show the positions of protein-coding genes on the positive and negative strands, respectively. Circle 3 shows the positions 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

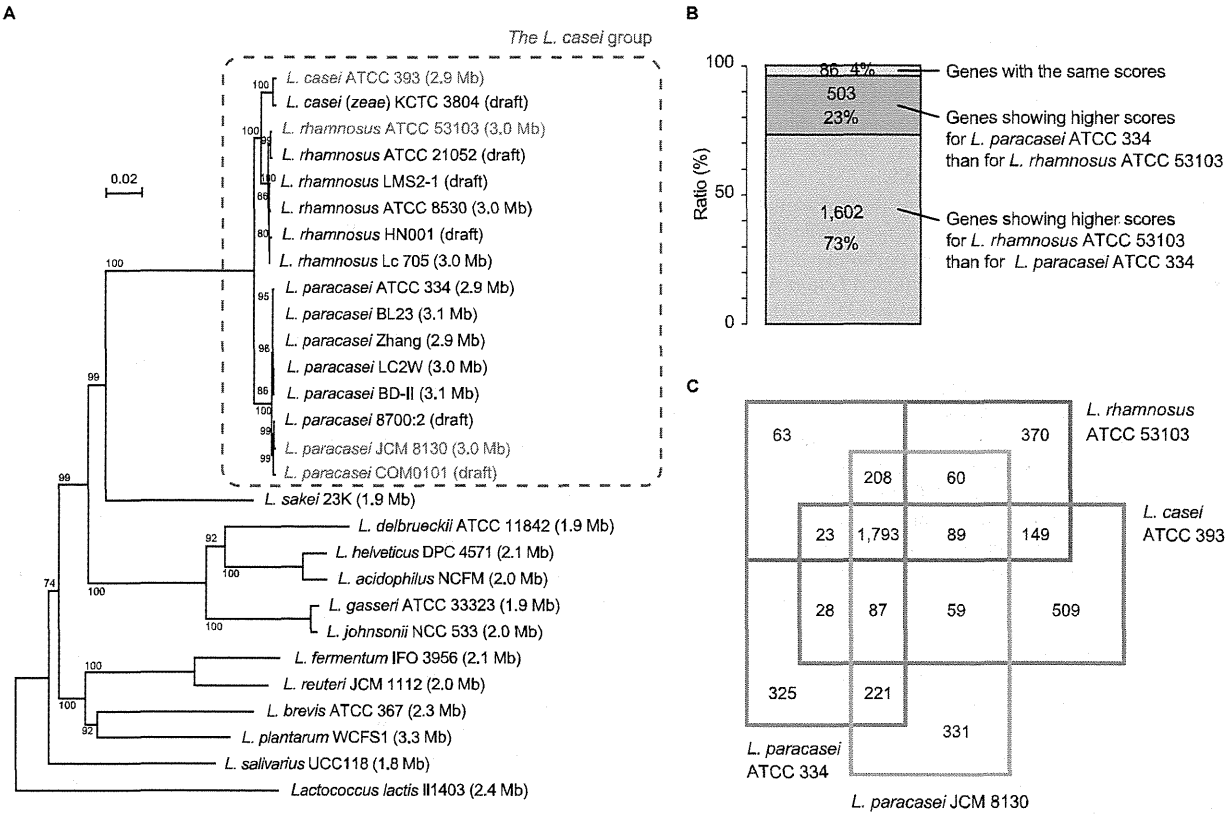

Figure 2. Genome-based phylogenetic analysis of the *L. casei* group. (**A**) Phylogenetic relationships between the genomes of sequenced lactobacilli inferred from 34 concatenated ribosomal protein amino acid sequences. The scale bar represents an evolutionary distance. Sequences were aligned with ClustalW 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 (Lsa11 and Ldbu1) of CRISPR loci were identified in the *L. casei* genomes [25]. *L. casei* strains BD-II, BL23, LC2W, and Zhang also have an Lsa11-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

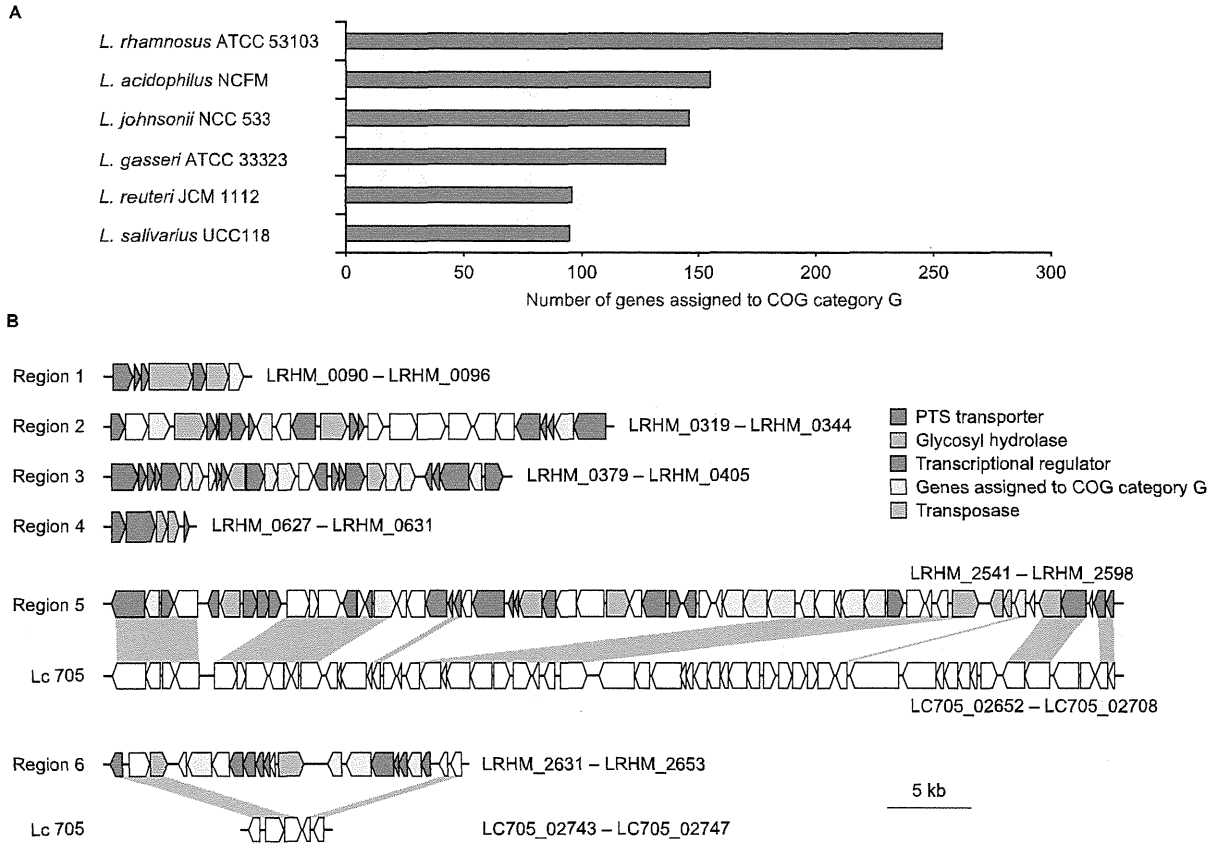

Figure 3. Abundance of genes related to carbohydrate transport and metabolism in *L. rhamnosus* ATCC 53103. (**A**) Comparative analysis by functional categories of the gene repertoires of sequenced intestinal lactobacilli. The number of genes assigned to COG category G in each genome is shown. (**B**) 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

Table 1. Cont.

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_2085–LRHM_2097	12.3	conserved hypothetical protein				LBCZ_2402–LBCZ_2414	12.5	carbohydrate utilization gene cluster
LRHM_2115–LRHM_2119	8.3	CRISPR-associated protein				LBCZ_2437–LBCZ_2492	66.7	putative cell surface protein, carbohydrate utilization gene cluster
LRHM_2193–LRHM_2198	11.8	cell surface protein, glycosyltransferase				LBCZ_2499–LBCZ_2517	21.1	transposase, conserved hypothetical protein
LRHM_2223–LRHM_2230	7.3	multidrug ABC transporter, hypothetical protein				LBCZ_2616–LBCZ_2643	31.3	carbohydrate utilization gene cluster, transposase
LRHM_2351–LRHM_2356	8.1	multidrug ABC transporter				LBCZ_2678–LBCZ_2694	15.0	transposase
LRHM_2545–LRHM_2597	57.7	carbohydrate utilization gene cluster (region-5)				LBCZ_2698–LBCZ_2704	7.6	PTS transporter
LRHM_2635–LRHM_2651	15.4	carbohydrate utilization gene cluster (region-6)						
LRHM_2779–LRHM_2793	12.5	prophage region III						

doi:10.1371/journal.pone.0075073.t001

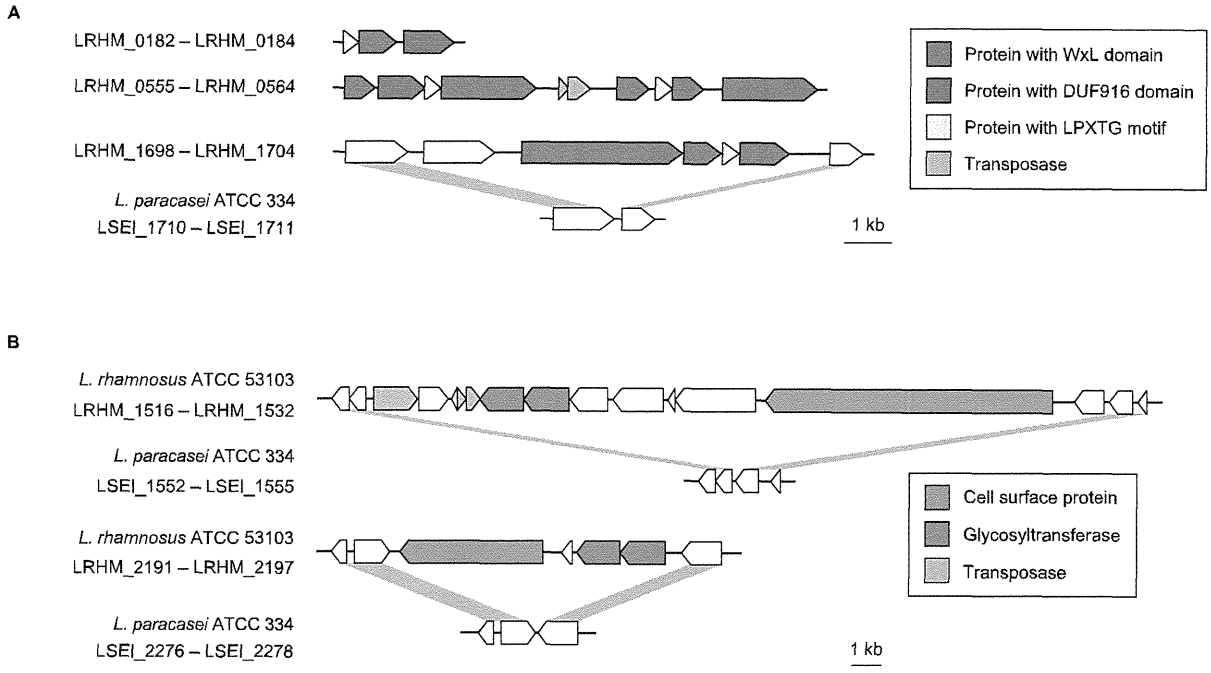

Figure 4. Gene clusters encoding cell surface proteins in *L. rhamnosus* ATCC 53103. (**A**) WxL clusters. (**B**) Putative glycosylated cell-surface protein clusters. Genes and their orientations are depicted with arrows. Gray bars indicate orthologous regions between *L. rhamnosus* ATCC 53103 and *L. paracasei* ATCC 334.
doi:10.1371/journal.pone.0075073.g004

Extracellular Components

Another group has also determined the complete genome sequence of *L. rhamnosus* GG, and revealed the presence of the SpaCBA pili on the cell surface of *L. rhamnosus* GG [9]. SpaA is a backbone-forming major pilin, SpaB is a minor pilin, and SpaC located at the pilus tip is essential for the mucus adherence of *L. rhamnosus* GG [9,30]. The *spaCBA* genes are encoded in the largest GI (LRHM_0376 to LRHM_0466) in *L. rhamnosus* ATCC 53103 (Fig. S5). The *L. paracasei* Zhang, *L. paracasei* BL23, and *L. paracasei* ATCC 334 genomes also encode the *spaCBA* genes (Fig. S5). In contrast, *L. casei* ATCC 393 completely lacks the *spaCBA* genes. The *spaCBA* genes were also encoded in *L. paracasei* COM0101, but the *spaC* gene was truncated by a nonsense mutation [25] (Fig. S5), which probably encodes a non-functional protein. Douillard *et al.*, (2013) clearly showed that the *L. paracasei* strain isolated from Yakult produced no pilus structures by an immunoelectron microscopy using immunogold staining [31]. It has been reported that the adhesion capacity of *L. rhamnosus* GG to Caco-2 cells and intestinal mucus was approximately 10 times that of strain Shirota, which was obtained from Yakult [32]. This may be because *L. rhamnosus* GG encodes the intact SpaCBA and *L. paracasei* COM0101 encodes truncated SpaC. Furthermore, *L. paracasei* JCM 8130, *L. paracasei* BD-II, and *L. paracasei* LC2W also contained truncated *spaC* gene (Fig. S5), and *L. rhamnosus* Lc 705 and ATCC 8530 completely lacked the *spaCBA* genes. The *spaCBA* genes have been found only in the *L. casei* group to date. Because different lineages in *L. casei* strains contained the *spaCBA* genes, it has been suggested that the *spaCBA* genes were not recently acquired [25]. It could thus be speculated that the ancestral strain of the *L. casei* group had encoded the intact *spaCBA* genes and then *spaCBA* may have been lost or disrupted in certain strains of the *L. casei* group.

L. rhamnosus ATCC 53103 had three gene clusters encoding proteins with a C-terminal WxL domain (Fig. 4A). The WxL domain is conserved in the surface proteins in low-GC gram-positive bacteria [33] and attaches to the peptidoglycan on the cell surface [34]. The WxL protein cluster was not found in other sequenced intestinal lactobacilli. The proteins with the WxL domain were present together with the proteins containing the DUF916 domain (PF06030) of unknown function and the small proteins with the LPXTG-like sorting motif, and their gene organizations were similar to that in *L. plantarum* WCFS1 [35]. Of the three WxL protein clusters, one (LRHM_1699 to LRHM_1702) was not conserved in the sequenced *L. paracasei* strains (Fig. 4A, Table 2). There were 14 genes encoding proteins that had both a signal sequence for secretion and an LPXTG-type motif for covalent anchoring to the peptidoglycan matrix (Table 2), and these proteins can be cleaved by sortase. The protein LRHM_1529 was composed of 3,275 amino acid residues, representing the largest protein in this genome, and it contained imperfect repeats consisting of serine, alanine, and aspartic acid. This serine-rich motif has been found in the extracellular proteins in the genomes of other gram-positive bacteria such as *L. plantarum*, *L. johnsonii*, and *Streptococcus pneumoniae* [29,36,37]. The protein LRHM_1529 was encoded in the region (LRHM_1518 to LRHM_1530), which contained two glycosyltransferase genes (Fig. 4B). It has been suggested that glycosyltransferase, encoded by the adjacent genes, caused O-linked glycosylations on the serines in the putative cell surface protein, thus producing mucin-like structures [36]. Similarly, the protein LRHM_2193 had an LPXTG-type motif, and it contained imperfect repeats consisting of serine and alanine and two adjacent glycosyltransferase genes (Fig. 4B). Thus, LRHM_1529 and LRHM_2193 could encode glycosylated cell-surface adhesives. The protein LRHM_1797 (2,357 amino acids) plays an important modulating role in

Table 2. Putative cell surface adherence proteins of *L. rhamnosus* ATCC 53103.

Locus	Size (aa)	Contained domain	SignalP	*L. paracasei* ATCC 334	*L. paracasei* BL23	*L. paracasei* Zhang	*L. paracasei* JCM 8130	*L. casei* ATCC 393	*L. rhamnosus* Lc 705	*L. rhamnosus* ATCC 8530
LRHM_0051	1,492	fibronectin-binding	+	−	−	−	−	−	+	+
LRHM_0182	106	LPXTG	+	−	−	−	−	−	+	+
LRHM_0183	268	WxL	+	+	+	+	+	+	+	+
LRHM_0184	359	DUF916	+	+	+	+	+	+	+	+
LRHM_0426	334	LPXTG (SpaA)	+	−	+	+	+	−	−	−
LRHM_0427	241	LPXTG (SpaB)	+	−	+	+	+	−	−	−
LRHM_0428	895	LPXTG (SpaC)	+	+	+	+	−	−	−	−
LRHM_0555	220	WxL1	+	+	+	+	+	+	+	+
LRHM_0556	340	DUF916	+	+	+	+	+	+	+	+
LRHM_0557	118	LPXTG	+	+	+	+	+	+	+	+
LRHM_0558	688	WxL2	+	−	+	+	−	+	+	+
LRHM_0561	238	WxL1	+	+	+	+	+	−	+	+
LRHM_0562	124	LPXTG	+	+	+	+	+	−	+	+
LRHM_0563	229	WxL1	+	+	+	+	+	−	+	+
LRHM_0564	679	WxL2	+	+	+	+	+	−	+	+
LRHM_1138	401	LPXTG	−	+	+	+	+	+	+	+
LRHM_1331	213	LysM	−	+	+	+	+	+	+	+
LRHM_1393	567	fibronectin-binding	−	+	+	+	+	+	+	+
LRHM_1528	913	Ig-like fold	+	−	−	−	−	+	+	+
LRHM_1529	3,275	LPXTG	+	−	−	−	−	+	+	+
LRHM_1699	351	DUF916	+	−	−	−	−	+	+	+
LRHM_1700	114	LPXTG	+	−	−	−	−	−	+	+
LRHM_1701	262	WxL	+	−	−	−	−	+	+	+
LRHM_1702	1,131	WxL	−	−	−	−	−	+	+	+
LRHM_1797	2,357	LPXTG	−	−	−	−	−	+	+	+
LRHM_2006	1,561	LPXTG	+	−	−	−	−	−	+	−
LRHM_2185	1,973	LPXTG	+	+	+	+	+	+	+	+
LRHM_2193	1,653	LPXTG	−	−	−	−	−	+	+	+
LRHM_2248	388	LPXTG, mucin-binding domain	−	+	+	+	+	+	+	+
LRHM_2279	517	LPXTG (SpaD)	+	+	+	+	+	−	+	+
LRHM_2281	983	LPXTG (SpaF)	+	+	+	+	+	−	+	+
LRHM_2626	1,494	LPXTG	+	−	−	−	−	−	+	+
LRHM_2815	2,603	LPXTG	+	+	+	+	+	−	+	+

*'+' indicates that the orthologous gene is present, and '−' indicates that the orthologous gene is absent.

doi:10.1371/journal.pone.0075073.t002

adhesion to intestinal epithelial cells and biofilm formation [38]. These genes (LRHM_1529, LRHM_1797, and LRHM_2193) were absent in the sequenced *L. paracasei* strains. The presence of a variety of the cell surface adherence proteins could contribute to the probiotic properties of *L. rhamnosus* ATCC 53103.

Conclusions

We determined the complete genome sequences of *L. paracasei* JCM 8130 and *L. casei* ATCC 393, and the draft genome sequence of *L. paracasei* COM0101. Furthermore, we re-annotated the genome of *L. rhamnosus* ATCC 53103. We confirmed that *L. casei* ATCC 393 is distinct from the *L. paracasei* strains previously. Comparative genome analysis revealed 1,682 core genes and genome-wide synteny in the *L. casei* group. Chromosomes of the *L. casei* group contained GIs, many of which are also found at the same loci, suggesting that the chromosomes of the *L. casei* group contain several hypervariable regions at the same loci, which may contribute to the adaptation to each ecological niche. The *spaCBA* pilus gene cluster, which was first identified in *L. rhamnosus* GG, was also found in other strains of the *L. casei* group, but several *L. paracasei* strains including COM0101 contained truncated *spaC* gene. *L. rhamnosus* ATCC 53103 encodes SpaCBA pili, proteins with WxL domain, two glycosylated cell-surface adhesives, and several large proteins with the LPXTG motif. The complete genome sequences of *L. rhamnosus*, *L. paracasei*, and *L. casei* will provide a framework that will help understand the genomic differences between strains within the *L. casei* group.

Supporting Information

Figure S1 Linear representations of the plasmids of *L. casei* 393 and of *L. rhamnosus* Lc 705. Genes and their orientations are depicted with arrows. Several lines connect orthologs with the following colors: red, genes sharing over 95% amino acid identity; orange, genes sharing 70–95% amino acid identity; blue, transposase genes; and green, partially conserved genes.
(EPS)

Figure S2 Genetic relationships among *L. paracasei* strains as defined by multilocus sequence typing. (A) Concatenated sequences of five MLST loci (*ftsZ*, *metRS*, *mutL*, *pgm*, and *polA*) were analyzed as described previously [24]. **(B)** Venn diagram comparing the gene inventories of four *L. paracasei* strains. Data resulted from reciprocal BLASTP analysis. The numbers of shared and unique genes are shown.
(EPS)

Figure S3 COG classification of dispensable protein-coding genes of the *L. casei* group.
(EPS)

Figure S4 Synteny between the chromosomes in the *L. casei* group. Each plot point represents reciprocal best matches by BLASTP comparisons between orthologs.
(EPS)

Figure S5 The *spaCBA* pili cluster arrangement. Genes and their orientations are depicted with arrows.
(EPS)

Table S1 General genomic features of strains sequenced in this study.
(PDF)

Acknowledgments

We thank K. Furuya, C. Shindo, H. Inaba, K. Motomura, and Y. Hattori (The University of Tokyo), and A. Tamura and N. Itoh (Kitasato University) for technical assistance, and Dr. H. Zhang for supplying *L. paracasei* Zhang.

Author Contributions

Conceived and designed the experiments: HM. Performed the experiments: AN MT TT HN SI. Analyzed the data: HT KO MM MH HM. Contributed reagents/materials/analysis tools: AN. Wrote the paper: HT HM.

References

1. Felis GE, Dellaglio F (2007) Taxonomy of Lactobacilli and Bifidobacteria. Curr Issues Intest Microbiol 8: 44–61.
2. Dicks LM, Du Plessis EM, Dellaglio F, Lauer E (1996) Reclassification of *Lactobacillus casei* subsp. *casei* ATCC 393 and *Lactobacillus rhamnosus* ATCC 15820 as *Lactobacillus zeae* nom. rev., designation of ATCC 334 as the neotype of *L. casei* subsp. *casei*, and rejection of the name *Lactobacillus paracasei*. Int J Syst Bacteriol 46: 337–340.
3. Felis GE, Dellaglio F, Mizzi L, Torriani S (2001) Comparative sequence analysis of a *recA* gene fragment brings new evidence for a change in the taxonomy of the *Lactobacillus casei* group. Int J Syst Evol Microbiol 51: 2113–2117.
4. Dellaglio F, Felis GE, Torriani S (2002) The status of the species *Lactobacillus casei* (Orla-Jensen 1916) Hansen and Lessel 1971 and *Lactobacillus paracasei* Collins *et al.* 1989. Request for an opinion. Int J Syst Evol Microbiol 52: 285–287.
5. Acedo-Félix E, Pérez-Martínez G (2003) Significant differences between *Lactobacillus casei* subsp. *casei* ATCC 393T and a commonly used plasmid-cured derivative revealed by a polyphasic study. Int J Syst Evol Microbiol 53: 67–75.
6. Diancourt L, Passet V, Chervaux C, Garault P, Smokvina T, et al. (2007) Multilocus sequence typing of *Lactobacillus casei* reveals a clonal population structure with low levels of homologous recombination. Appl Environ Microbiol 73: 6601–6611.
7. Judicial Commission of the International Committee on Systematics of Bacteria (2008) The type strain of *Lactobacillus casei* is ATCC 393, ATCC 334 cannot serve as the type because it represents a different taxon, the name *Lactobacillus paracasei* and its subspecies names are not rejected and the revival of the name '*Lactobacillus zeae*' contravenes Rules 51b (1) and (2) of the International Code of Nomenclature of Bacteria. Opinion 82. Int J Syst Evol Microbiol 58: 1764–1765.
8. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, et al. (2006) Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci U S A 103: 15611–15616.
9. Kankainen M, Paulin L, Tynkkynen S, von Ossowski I, Reunanen J, et al. (2009) Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human– mucus binding protein. Proc Natl Acad Sci U S A 106: 17193–17198.
10. Mazé A, Boël G, Zúñiga M, Bourand A, Loux V, et al. (2010) Complete genome sequence of the probiotic *Lactobacillus casei* strain BL23. J Bacteriol 192: 2647–2648.
11. Zhang W, Yu D, Sun Z, Wu R, Chen X, et al. (2010) Complete genome sequence of *Lactobacillus casei* Zhang, a new probiotic strain isolated from traditional homemade koumiss in Inner Mongolia, China. J Bacteriol 192: 5268–5269.
12. Ai L, Chen C, Zhou F, Wang L, Zhang H, et al. (2011) Complete genome sequence of the probiotic strain *Lactobacillus casei* BD-II. J Bacteriol 193: 3160–3161.
13. Chen C, Ai L, Zhou F, Wang L, Zhang H, et al. (2011) Complete genome sequence of the probiotic bacterium *Lactobacillus casei* LC2W. J Bacteriol 193: 3419–3420.
14. Pittet V, Ewen E, Bushell BR, Ziola B (2012) Genome sequence of *Lactobacillus rhamnosus* ATCC 8530. J Bacteriol 194: 726.
15. Morita H, Toh H, Oshima K, Murakami M, Taylor TD, et al. (2009) Complete genome sequence of the probiotic *Lactobacillus rhamnosus* ATCC 53103. J Bacteriol 191: 7630–7631.
16. Doron S, Snydman DR, Gorbach SL (2005) *Lactobacillus* GG: bacteriology and clinical applications. Gastroenterol Clin North Am 34: 483–498.
17. Mahenthiralingam E, Marchbank A, Drevinek P, Garaiova I, Plummer S (2009) Use of colony–based bacterial strain typing for tracking the fate of *Lactobacillus* strains during human consumption. BMC Microbiol 9: 251.
18. Morita H, Kuwahara T, Ohshima K, Sasamoto H, Itoh K, et al. (2007) An improved DNA isolation method for metagenomic analysis of the microbial flora of the human intestine. Microbes Environ 22: 214–222.

19. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) Improved microbial gene identification with GLIMMER. Nucleic Acids Res 27: 4636–4641.
20. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 955–964.
21. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8: 785–786.
22. Grissa I, Vergnaud G, Pourcel C (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 35: W52–57.
23. Rasko DA, Myers GS, Ravel J (2005) Visualization of comparative genomic analyses by BLAST score ratio. BMC Bioinformatics 6: 2.
24. Cai H, Rodríguez BT, Zhang W, Broadbent JR, Steele JL (2007) Genotypic and phenotypic characterization of *Lactobacillus casei* strains isolated from different ecological niches suggests frequent recombination and niche specificity. Microbiology 153: 2655–2665.
25. Broadbent JR, Neeno-Eckwall EC, Stahl B, Tandee K, Cai H, et al. (2012) Analysis of the *Lactobacillus casei* supragenome and its influence in species evolution and lifestyle adaptation. BMC Genomics 13: 533.
26. Horvath P, Coûté-Monvoisin AC, Romero DA, Boyaval P, Fremaux C, et al. (2009) Comparative analysis of CRISPR loci in lactic acid bacteria genomes. Int J Food Microbiol 131: 62–70.
27. Cai H, Thompson R, Budinich MF, Broadbent JR, Steele JL (2009) Genome sequence and comparative genome analysis of *Lactobacillus casei*: Insights into their niche-associated evolution. Genome Biol Evol 1: 239–257.
28. Makarova KS, Koonin EV (2007) Evolutionary genomics of lactic acid bacteria. J Bacteriol 189: 1199–1208.
29. Schell MA, Karmirantzou M, Snel B, Vilanova D, Berger B, et al. (2002) The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. Proc. Natl. Acad. Sci U S A 99: 14422–14427.
30. Reunanen J, von Ossowski I, Hendrickx AP, Palva A, de Vos WM (2012) Characterization of the SpaCBA pilus fibers in the probiotic *Lactobacillus rhamnosus* GG. Appl Environ Microbiol 78: 2337–2344.
31. Douillard FP, Ribbera A, Järvinen HM, Kant R, Pietilä TE, et al. (2013) Comparative genomic and functional analysis of *Lactobacillus casei* and *Lactobacillus rhamnosus* strains marketed as probiotics. Appl Environ Microbiol 79: 1923–1933.
32. Lee YK, Lim CY, Teng WL, Ouwehand AC, Tuomola EM (2000) Quantitative approach in the study of adhesion of lactic acid bacteria to intestinal cells and their competition with enterobacteria. Appl Environ Microbiol 66: 3692–3697.
33. Kleerebezem M, Boekhorst J, van Kranenburg R, Molenaar D, Kuipers OP, et al. (2003) Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc Natl Acad Sci U S A 100: 1990–1995.
34. Brinster S, Furlan S, Serror P (2007) C-terminal WxL domain mediates cell wall binding in *Enterococcus faecalis* and other gram-positive bacteria. J Bacteriol 189: 1244–1253.
35. Siezen R, Boekhorst J, Muscariello L, Molenaar D, Renckens B, et al. (2006) *Lactobacillus plantarum* gene clusters encoding putative cell-surface protein complexes for carbohydrate utilization are conserved in specific gram-positive bacteria. BMC Genomics 7: 126.
36. Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, et al. (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science 293: 498–506.
37. Pridmore RD, Berger B, Desiere F, Vilanova D, Barretto C, et al. (2004) The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. Proc Natl Acad Sci U S A 101: 2512–2517.
38. Vélez MP, Petrova MI, Lebeer S, Verhoeven TL, Claes I, et al. (2010) Characterization of MabA, a modulator of *Lactobacillus rhamnosus* GG adhesion and biofilm formation. FEMS Immunol Med Microbiol 59: 386–398.