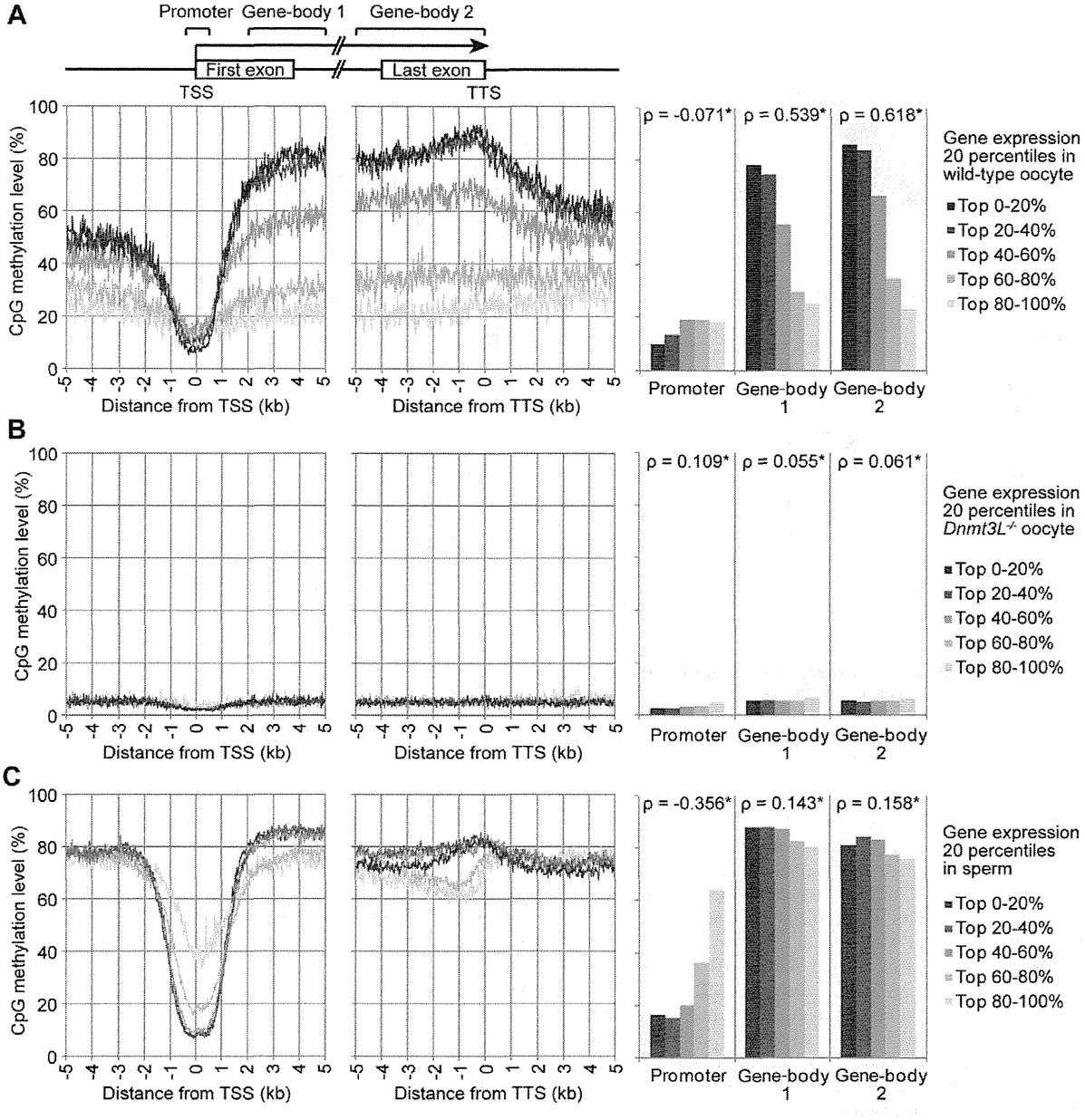

Figure 4. Relationship between gene expression and methylation in promoter and gene-body regions in mouse germ cells. The expression level of genes in wild-type oocytes (A), sperm (B), and *Dnmt3L*$^{-/-}$ oocytes (C) were divided into 5 percentile groups. The distribution of methylation is shown ±5 kb from the transcription termination site (TTS; *left*) and transcription start site (TSS; *middle*). The graphs on the right show the average methylation levels in the promoter and gene-body regions. Spearman's rank correlation coefficient (ρ) was used to test the statistical significance of the correlation between gene expression and DNA methylation levels (*: $p<1\times10^{-9}$).
doi:10.1371/journal.pgen.1002440.g004

To determine whether or not these germ cell-specific methylations are maintained after fertilization, when the genomes undergo global demethylation, the individual CGI methylation levels in blastocyst genomes were calculated. In blastocysts, all ICRs demonstrated low to moderate methylation (25.1–64.3%), whereas many gDMRs were demethylated (0–20%) (Figure 6D). Furthermore, 817 oocyte-methylated gDMRs (including *Piwil1*, despite being a non-imprinted gene locus) and 34 sperm-specific gDMRs were resistant to demethylation during early embryogenesis (≥20% methylation in blastocysts) (Figure 6D and Table S2). Among the demethylation-resistant gDMRs, a novel gDMR in the intron of *Gpr1* (Figure S10) was found to be a tissue-specific, paternally-expressed imprinted gene [40]. Bisulfite sequencing analysis showed that this gDMR was hypomethylated in *Dnmt3L*$^{-/-}$ oocytes and maternal allele-specific methylation was detected in this region in blastocysts (Figure 6E). Methylation profiles in ESCs showed that 26% (n = 213) of the demethylation-resistant gDMRs became less methylated (0–20%) whereas the other gDMRs maintained or increased DNA methylation (Figure S15). Among ICRs, only *Gnas* exon1A ICR was demethylated (7.8%),

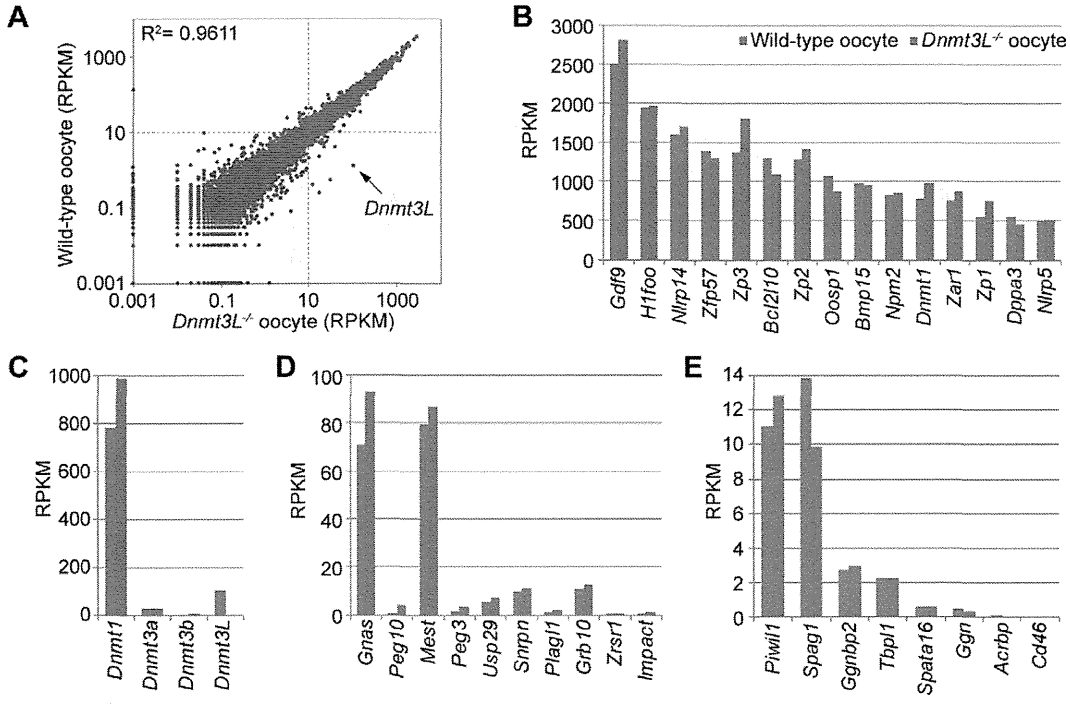

Figure 5. Comparison of gene expression profiles between wild-type and *Dnmt3L*$^{-/-}$ oocytes. (A) Scatter plot and correlation coefficient (R^2) of RPKM values of 20,854 genes in wild-type and *Dnmt3L*$^{-/-}$ oocytes. Expression levels of oocyte-specific genes (B), DNA methyltransferase genes (C), maternally-imprinted genes that are potentially necessary to establish methylation imprints (D), and male germline-specific genes that contain oocyte-specific methylated CpG islands (CGIs) (E).
doi:10.1371/journal.pgen.1002440.g005

whereas the other ICRs developed partial or high methylation levels (range, 38.6–83.1%) in ESCs (Table 2). Among other demethylation-sensitive gDMRs, which were demethylated (<20% methylation) in blastocysts, many (76%, n = 264) sperm-methylated gDMRs were re-methylated (≥20% methylation); most (81%, n = 416) of the oocyte-methylated gDMRs maintained low methylation (0–20%) in ESCs (Figure S15). Finally, out of 704 demethylation-resistant (in blastocysts) oocyte-methylated gDMRs which were informative in *Dnmt3L*$^{-/-}$ oocytes, only 4 remained hypermethylated (80–100% methylation) in the *Dnmt3L*$^{-/-}$ oocyte genome. However, almost all other oocyte-specific methylation marks at gDMRs were *Dnmt3L*-dependent (Figure 6F). These results suggest that *Dnmt3L*-mediated methylation during oogenesis regulates the establishment of most heritable oocyte-specific marks, including genomic imprints.

Discussion

To the best of our knowledge, this is the first study to generate single-base resolution maps of DNA methylomes spanning the entire genome of mouse germ cells. The oocyte maps are particularly valuable and informative because, in the past, such an analysis was prohibitive due to the need for large quantities of DNA. Recently, Smallwood et al. [38] reported large-scale DNA methylation patterns in mouse germ cells by using the RRBS method, which targets only CpG-rich regions. However, our more comprehensive results provide strong evidence that gene expression was positively correlated to *Dnmt3L*-dependent intragenic methylation in oocytes, and that methylation patterns in oocytes differed from those in sperm and non-germline cells.

The functional role of gene-body methylation has been an enigma despite its conservation in plants and animals [41–43]. Maunakea et al. [44] suggested that gene-body methylation is involved in the regulation of alternative splicing events. Although methylated gDMRs were detected in the alternative exons of *Dnmt1* and *Gnas* in mouse oocytes, loss of oocyte-specific methylation marks in the *Dnmt3L*$^{-/-}$ oocytes did not affect the expression patterns of alternatively spliced transcripts. Therefore, our results indicate that gene-body methylation is not involved in alternative splicing in oocytes.

Previously, Chotalia et al. [36] showed that transcription during the oocyte stage is required for the establishment of maternal methylation marks on an imprinted gene. The present results show that *Dnmt3L*$^{-/-}$ oocytes lost almost all of their maternal methylation imprints while maintaining a constant amount of mRNA through each ICR despite the global loss of intragenic methylation. Thus, these results strongly suggest that the establishment of genomic imprints via transcription is mediated by *Dnmt3L*-dependent intragenic methylation.

A possible mechanism for gene-body methylation involves the exposure of intragenic regions to DNA methyltransferases, considering that RNA polymerase disrupts the chromatin structure during transcription. However, not all transcripts across gDMRs corresponded to highly expressed genes in oocytes (Figure 6C). Therefore, other epigenetic marks with an open chromatin structure might also be important for DNA methylation in oocytes. For instance, a recent knockout study showed that *Kdm1b*, which encodes histone H3K4 demethylase, is required for the establishment of some maternal methylation imprints [45]. Thus, several factors, including transcriptional and epigenetic modifica-

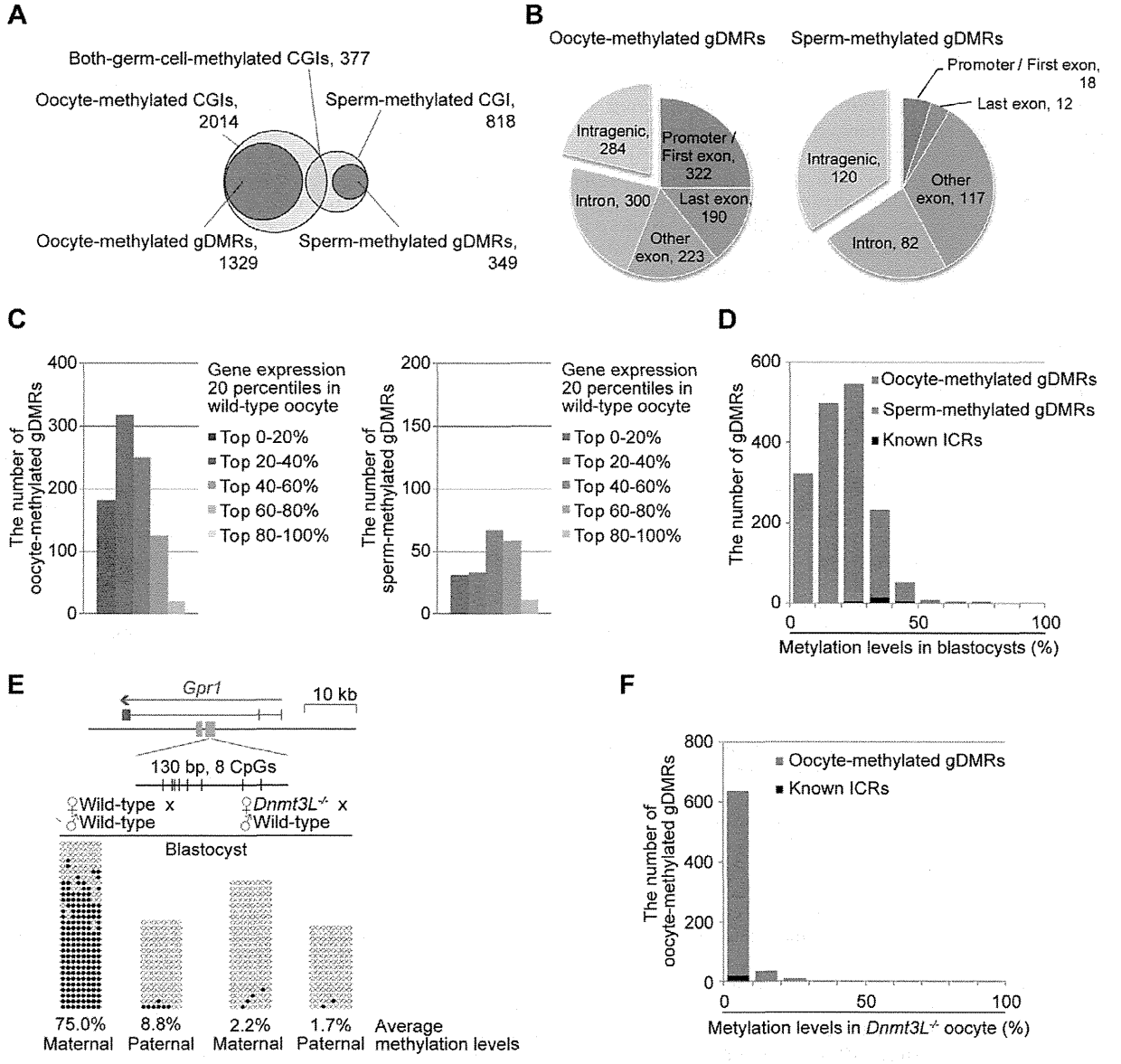

Figure 6. Identification of germline differentially methylated CGIs from DNA methylome profiles. (A) Venn-like diagram of two groups of CGIs, namely, oocyte-methylated CGIs (light pink) and sperm-methylated CGIs (light blue) and two groups of gDMRs, namely, oocyte-methylated gDMRs (red) and sperm-methylated gDMRs (blue). (B) The genomic distribution of 1329 oocyte-methylated (left) and 349 sperm-methylated gDMRs (right). The gDMRs were classified into 5 genomic locations; promoter (within 500-bp upstream from the first exon) or first exon, last exon, other exon, intron, and intergenic region. (C) The locations of the intragenic 1045 oocyte-methylated (left) and 229 sperm-methylated gDMRs (right). The gDMRs were classified into 5 gene group locations; the genes were divided into 5 percentile groups according to their expression levels in wild-type oocytes and sperm, as shown in Figure 3. (D) Histograms of the methylation levels of the gDMRs in blastocysts. The number of newly identified oocyte-specific, sperm-specific methylated gDMRs, and known ICRs are shown in black, red, and blue, respectively. (E) Bisulfite sequencing at the *Gpr1* gDMR in mouse blastocysts. (*Top*) Schematic representation of paternally-expressed *Gpr1*. The gene and gDMRs are shown in blue and green, respectively, and CpG sites are represented by vertical bars. (*Bottom*) Methylated and unmethylated CpGs are indicated by open and closed circles, respectively. The maternal and paternal alleles were distinguished by three polymorphisms between C57BL/6N and JF1 mice (G/A at 63,247,064; T/A at 63,247,072; and TA/AG at 63,247,089–63,247,090 on chromosome 1). (F) Histograms of the methylation levels of the demethylation-resistant oocyte-methylated gDMRs in *Dnmt3L*$^{-/-}$ oocytes. The number of newly identified oocyte-specific, sperm-specific methylated gDMRs, and known ICRs are shown in black, red, and blue, respectively.
doi:10.1371/journal.pgen.1002440.g006

tions, might be involved in *Dnmt3L*-mediated intragenic methylation.

The results of this study show that gene-body methylation was correlated to gene expression in sperm. However, the extent of that correlation is much less than in oocytes due to genome-wide hypermethylation, including in low-CpG-density regions. In male germline cells, global methylation acquisition begins during late embryonic development and before birth [3]. To more clearly

show this correlation, analysis of early-stage germ cells in fetal or neonatal animals might be required. Surprisingly, a positive correlation between mRNA expression and gene-body methylation was not observed in mouse ESCs. In addition, the accumulation of non-CpG methylation was not observed in mouse ESCs. These results contradict the results of another study, which showed that active transcription was associated with intragenic DNA methylation with non-CpG methylation in human ESCs [22,23]. This discrepancy might reflect the differences between human and mouse ESCs, the precise cell derivations or culture conditions [46,47]. However, further comparative studies on germ cell epigenomes from other species are required to further elucidate the functional role of epigenetic marking systems.

In this study, a large number of heritable oocyte-specific methylation marks were identified within a set of novel CpG islands [37]. The difference in the number of oocyte- and sperm-specific gDMRs reflects the fact that only 3 or 4 paternally-methylated imprinted loci were observed, as compared to approximately 20 maternally-methylated imprinted loci. The reason for the relative abundance of oocyte-specific methylated CGIs might be related to the intragenic methylation of CpG-rich regions, which are hypomethylated in sperm. The results show that most of the oocyte-specific marks are *Dnmt3L*-dependent, similar to results recently obtained by RRBS-based analysis [38]. However, whether all of these CpG-rich regions serve as imprinting methylation marks is unclear. For instance, although many genes with oocyte-specific methylation marks were identified (Figure 6B), the evidence that these genes were imprinted was lacking (*e.g.*, *Piwil1* and *Dnmt1*). These methylation marks might not be involved in the formation of a fertile oocyte but might play crucial roles in gene expression after fertilization. Furthermore, ESC methylomes showed that many gDMRs, especially sperm-specific gDMRs, acquired new methylation patterns after implantation. Methylation of these CGIs might control tissue-specific gene expression [48,49]. Partial alternation of imprinted methylation patterns in ESCs were observed in the present study, potentially caused by significant differences in the extent of the ICRs during embryo development [39]. A fuller understanding of epigenetic stability will require further methylome profiling during early embryogenesis and stem cell differentiation. The present study also identified a gDMR as a novel ICR candidate in the intron of the imprinted *Gpr1* gene. Thus, traditional promoter arrays may not identify all ICRs. However, further analyses are needed to determine which gDMRs, identified in the CpG methylome maps, are true ICRs at the imprinted *Gpr1-Zdbf2* locus [40,50].

mRNA-seq results showed that the expression levels of most genes in the wild-type and *Dnmt3L*$^{-/-}$ oocytes were similar. For instance, the expression level of almost all oocyte-specific genes, which regulate ovarian follicle formation, reproduction, and early development, were not significantly altered (Figure 5B and Table S1). These results are consistent with the findings of previous studies, which showed that *Dnmt3L*$^{-/-}$ female mice were capable of producing fertile oocytes (however, their offspring were not viable due to the lack of imprinting) [5,6]. Thus, regulation of oocyte-specific genes must be beyond the control of *Dnmt3L*-dependent cytosine methylation.

Although *Dnmt3L*$^{-/-}$ oocytes showed global hypomethylation at low to high CpG densities, some families of retrotransposons, such as LINEs and LTRs, were partially methylated at moderate to high CpG densities. Therefore, *Dnmt3L*-independent methylation might be involved in the silencing of retrotransposons and completion of oocyte meiosis. Previously, De La Fuente et al. [51] showed that *Hells* (also known as *Lsh*), which encodes a member of the sucrose non-fermenter 2 (SNF2) family of chromatin remodeling proteins, is required for DNA methylation of IAP and pericentromeric satellite repeats as well as repression of IAP retrotransposition in pachytene oocytes. Unfortunately, measurement of the methylation levels of satellite DNA, which is abundant in the pericentromeric regions, was not possible because these sequences were excluded from our analysis. However, a previous sequencing study showed that methylation levels of satellite DNA did not differ between the wild-type and *Dnmt3L*$^{-/-}$ oocytes [52]. Combined, these results suggest the presence of 2 types of oocyte methylation patterns: (i) *Dnmt3L*-mediated intragenic methylation that is essential for early embryogenesis and (ii) *Dnmt3L*-independent retroviral and pericentromeric methylation, which may be mediated by *Hells* activity, is crucial for oocyte meiosis [51]. Further studies on *Hells*-mediated oocyte methylation are required to elucidate the details of this mechanism.

Previous studies on the cytosine methylation of mtDNA have been highly controversial. A recent study by Shock et al. [53] reported cytosine methylation and hydroxymethylation in mammalian mitochondria. Our results indicated that mtDNA is unmethylated in blastocysts and ESCs, but is partially methylated in germ cells. Whether or not 5-hydroxymethylcytosine (5-hmC) exists in mitochondrial or genomic chromosomes of germ cells remains unclear. Meanwhile, rapid hydroxylation of 5-methylcytosine (5-mC) in the paternal pronucleus during zygotic development was also recently reported [54,55]. Currently, it is difficult to assess hydroxymethylation profiles in oocyte genomes due to the limited DNA recovery. Further investigation of cytosine modification during germ cell and zygote development will be required in the future to better understand this process.

The DNA methylome maps of mouse germ cells, in this study, were derived from SBS data and, therefore, accurately represent methylation levels of individual CpGs on a whole-genome level. The adaptation of the SBS method for small-scale DNA analysis, described in the present report, has the potential to enable further analyses of germline lineages. The current work examined SBS library construction using 3 methods, MethylC-seq, WBA-seq, and PBAT. MethylC-seq basically required only microograms of DNA [22,23,56], thus over amplification might cause redundancy in oocyte libraries. The latter methods allow comprehensive methylome analysis in samples with low amounts of starting DNA by avoiding DNA damage due to sodium bisulfite treatment (after adapter ligation, in the case of MethylC-Seq). Recent studies using BS sequencing have shown that methylated cytosine is abundant in the non-CpG regions of human pluripotent stem cells and mouse oocytes [22,23,39,56]; however, the function of non-CpG methylation in mammalian genomes remains unclear. The PBAT results also showed an abundance of non-CpG methylation in oocytes, with results similar to a previous sequencing study on imprinted loci [39]. However, accurate assessment of non-CpG methylation is required using increased sequencing depths because methylation levels of the non-CpG sites were much lower than those of the CpG sites. SBS library construction was conducted by WBA-seq from 2000 fully matured (metaphase II stage) oocytes; sufficient quantities for sequencing were not obtained. During oogenesis, most of the oocyte specific imprinted methylation marks were established during the GV stage. This contrasted to a previous study where a continuous increase in methylation levels was observed [38]. Further improvement of SBS methods, requiring smaller amounts of DNA, is needed to provide complete germ cell methylome maps and to elucidate the exact function of non-CpG methylation in germ cells.

In conclusion, we constructed the first extensive, high-resolution maps of DNA methylomes of mouse oocytes and sperm. These maps described the epigenetic properties of these DNA methylomes. Our data could serve as a platform for future studies to elucidate the role of epigenetic modifications in the development and functioning of germ and stem cells. Such studies are anticipated to improve our understanding of epigenetic reprogramming.

Materials and Methods

Preparation of MethylC-seq libraries

Five thousand germinal vesicle (GV)-stage oocytes were collected from the ovarian follicles of adult (7- to 9-week-old) female C57BL/6N mice (Clea Japan, Tokyo, Japan) 44–48 h after they were injected with equine chorionic gonadotropin. Three hundred blastocysts at embryonic day 3.5 were obtained from superovulated adult female C57BL/6N mice by flushing the uterus. Genomic DNA was extracted using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA). Sperm were released from the cauda epididymises of adult male C57BL/6N mice. Sperm DNA was isolated by a standard phenol-chloroform extraction procedure with dithiothreitol (DTT). Genomic DNA from 2 lines of ESCs derived from C57BL/6J mice (Clea Japan) was extracted using the DNeasy Blood & Tissue Kit (Qiagen). DNA samples were sheared into 100-bp fragments in oocytes and 200-bp fragments in other samples using the Covaris S2 focused acoustic system (Covaris, Woburn, MA). Cytosine-methylated adapters (Illumina, San Diego, CA) were ligated to DNA by using the Paired-End DNA Sample Prep Kit or ChIP-Seq DNA Sample Prep Kit (Illumina). DNA fragments were isolated by 2–3% agarose gel electrophoresis and purified using the QIAquick Gel Extraction Kit (Qiagen). Sodium bisulfite conversion was performed using the Epitect Bisulfite Kit (Qiagen).

All bisulfite-converted DNA molecules were polymerase chain reaction (PCR)-amplified as follows: 2.5 U of Hot Start Taq polymerase (TaKaRa, Tokyo, Japan), 5 μL 10× PCR buffer, 25 μM dNTPs, 1 μL of each PCR Primer PE 1.0 and 2.0 (Illumina) (50 μL final). Thermocycling parameters were: initial denaturation at 94°C for 1 min, 15–25 cycles of denaturation at 94°C for 30 s, annealing at 65°C for 30 s, and extension at 72°C for 30 s, followed by a final extension at 72°C for 5 min. PCR reaction products were purified using the QIAquick kit (Qiagen).

Preparation of whole WBA-seq libraries

Two thousand GV-stage oocytes were collected from 7- to 9-week-old female C57BL/6N mice (Clea Japan) and, 2300 GV-stage oocytes were collected from 7–15-week-old *Dnmt3L*$^{-/-}$ female mice (129SvJae×C57BL/6N hybrid genetic background) [6,57]. Genomic DNA was extracted using the QIAamp DNA Mini Kit (Qiagen), and then bisulfite-treated with Epitect Bisulfite Kit (Qiagen). Subsequently, the bisulfite-converted DNA was amplified using Epitect Whole Bisulfitome Kit (Qiagen). The collected DNA was sheared into 200-bp fragments using Covaris S2. Unmodified Paired-End adapters (Illumina) were ligated to the DNA by using the Paired-End DNA Sample Prep Kit (Illumina). DNA fragments were isolated by 2% agarose gel electrophoresis and purified using the QIAquick Kit (Qiagen). All DNA was PCR amplified and purified in the same manner as the MethylC-seq method, except the number of PCR cycles was reduced to 7.

Preparation of PBAT libraries

GV-stage oocytes (400) and blastocysts (100) were obtained from 7- to 9-week-old female C57BL/6N mice (Clea Japan), and genomic DNA was extracted using the QIAamp DNA Mini Kit (Qiagen). The isolated oocyte and blastocyst genomic DNA and 100 ng of genomic DNA from sperm, blastocysts, and ESCs containing 1:200 amount of unmethylated lambda DNA (Invitrogen, Carlsbad, CA) were bisulfite-treated using the MethylCode Bisulfite Conversion Kit (Invitrogen). Details of the PBAT method are unpublished [Miura F & Ito T, personal communication]. Briefly, bisulfite-treated DNA were double-stranded using Klenow Fragments (3′- 5′ exo-) (New England Biolabs, Ipswich, MA) with random primers containing 5′ biotin tags and Illumina PE adaptors. The biotinylated molecules (first strand) were captured using Dynabeads M280 Streptavidin (Invitrogen) and double-stranded using Klenow Fragments (3′-5′ exo-) with random primers containing Illumina PE adaptors (second strand). Finally, template DNA strands were synthesized as complementary DNA with a second strand (unmethylated C is converted to T) using Phusion Hot Start High-Fidelity DNA Polymerase (New England Biolabs) with PCR Primer PE 1.0 (Illumina).

Preparation of mRNA sequencing libraries

Total RNA from 1000 wild-type GV oocytes, 500 *Dnmt3L*$^{-/-}$ GV oocytes, sperm, and ESCs was extracted using the RNeasy Mini Kit (Qiagen) and treated with DNase I (Promega, Madison, WI). RNA-Seq libraries were constructed using the mRNA-Seq Sample Preparation Kit (Illumina).

Sequencing

The MethylC-seq for blastocysts, WBA-seq, and PBAT libraries were sequenced on a HiSeq 2000 sequencing system (Illumina); the other MethylC-seq and mRNA-seq libraries were sequenced on a Genome Analyzer II (Illumina). Sample preparation, cluster generation, and sequencing were performed using the Paired-End Cluster Generation Kit-HS and the TruSeq SBS Kit-HS for the HiSeq 2000. Similarly, the Paired-End Cluster Generation Kits v2 and v4 and 18- and 36-Cycle Sequencing Kits v3 and v4 were used for the Genome Analyzer II. All kits were from Illumina.

Gene mapping

All sequenced reads were processed using the standard Illumina base-calling pipeline (v1.4–1.7). Generated sequence tags were mapped onto the mouse genome (mm9, UCSC Genome Browser, July 2007, Build 37.1) by using the Illumina ELAND program.

MethylC-seq tags (36 or 76 nt) were mapped with a custom Perl program, as described previously [17,22]. Briefly, all cytosines in the tags were replaced by thymines. Next, these tags were aligned to 2 mouse genome reference sequences (mm9), such that the antisense strand had cytosines replaced by thymines and the sense strand had guanines replaced by adenines. Finally, all tags (32–76 nt) that mapped uniquely without any mismatches to both strands were compiled and used for further analyses.

The 76 nt WBA-seq tags were mapped as follows. All tags were converted to 2 types of reads; in 1 read ("For" read), cytosines were replaced by thymines and in the other read ("Rev" read), guanines were replaced by adenines. Both "For" and "Rev" reads were aligned to sense and antisense mm9 strands. A total of 793, 397, 948, 480, and 238 million tags were aligned in wild-type oocytes, *Dnmt3L*$^{-/-}$ oocytes, sperm, blastocysts, and ESC genomes, respectively. To avoid bias, tags mapped with multiple hits or matched chromosome M (mitochondria), chromosome Y, or 3 types of repetitive sequences (simple repeat, low complexity repeat, and satellite DNA sequences) were omitted from further analyses.

The 47 nt PBAT tags (trimmed first 4 nt and last 1 nt) were mapped as follows. All guanidines in the tags were replaced by

adenines, and these tags were aligned to sense and antisense strands mm9.

For gene-level analysis, the concentrations of the perfectly matching 35 nt (trimmed first nt) mRNA-seq tags from wild-type oocytes, *Dnmt3L*$^{-/-}$ oocytes, sperm, and ESCs were calculated for the genomic regions corresponding to those covered by the RefSeq transcript models. The expression level of 20,854 unique genes was ranked by expression levels (calculated as RPKM values) in each library (Table S1). A total of 33, 28, 23, and 25 tags were aligned in 4 mRNA-seq libraries, respectively. mRNA-seq data analysis was performed and visualized using GenomeStudio Data Analysis software (Illumina).

Methylation analysis

The percentage of individual cytosines methylated at all CpG sites covered by at least 1 read was calculated as 100×(number of aligned cytosines (methylated cytosines))/(total number of aligned cytosines and thymines (originally unmethylated cytosines)). All genomic CpG methylation data are available on our website (http://www.nodai-genome.org/mouse_en.html). The CpG and non-CpG (CpH) methylation levels determined by PBAT results were calculated as the ratio between the total read C and the total read T mapped to genomic cytosines. Bisulfite conversion failure rates were calculated by read C:T ratios from lambda DNA mapping data. The failure rates were as follows: GV oocyte, 0.009; sperm, 0.008; blastocysts, 0.011; and ESCs, 0.006. Locations of transposable elements in the mouse genome (mm9) were obtained from the UCSC Genome Browser, and the average methylation levels of the whole genome and each transposable element were recalculated from the ratio of the aligned cytosines and thymines in each sequence. Lists of 23,021 CGIs were obtained from a previous report [37]. Around the TSS and TTS (±5 kb), genomic regions were divided into 20-bp bins. For each bin, the average methylation value was calculated for each gene. The expression level of 20,854 genes was divided into 5 percentile groups ranked by RPKM values, and the average methylation level for each group was mapped onto the gene structure model. These computational analyses were performed using a custom Perl program. Supercomputing resources were provided by the Human Genome Center, Institute of Medical Science, University of Tokyo.

Statistical analysis

Correlations between gene expression ranks and average methylation levels in the promoter (±500 bp from the TTS) or gene-body regions (gene-body 1: +2 to +5 kb from the TSS; gene-body 2: 0 to −5 kb from the TTS) were calculated using Spearman's rank correlation coefficient (ρ). An R-squared value (R^2) was calculated to evaluate the correlation of RPKM values between wild-type and *Dnmt3L*$^{-/-}$ oocytes. Statistical analysis was performed using the R statistical package.

Bisulfite sequencing

To analyze the methylation of the three transposable elements (L1 LINE, B1/Alu SINE, and IAP LTR), 20 wild-type GV oocytes were obtained from adult female C57BL/6N mice. Bisulfite sequencing conditions and primer sets for the three transposable elements were described, previously [52]. To analyze the methylation of the *Gpr1* locus, 10 blastocysts were obtained from BJF1 (C57BL/6N×JF1) and *Dnmt3L*$^{mat-/-}$ (*Dnmt3L*$^{-/-}$ ×JF1) mice [6,57]. Genomic DNA from blastocysts was isolated using the QIAamp DNA Mini Kit (Qiagen) and treated with sodium bisulfite with the Epitect Bisulfite Kit (Qiagen). The *Gpr1* gDMR sequence was amplified with 2 rounds of nested PCR. The first-round PCR reaction contained 1 U of Hot Start Taq polymerase (TaKaRa), 1× PCR buffer, 200 μM dNTPs, 1 μM forward primer, and 1 μM reverse primer (20 μL final). Thermocycling parameters were as follows: initial denaturation at 94°C for 1 min, 35 cycles of denaturation at 94°C for 30 s, annealing at 50°C for 30 s, and extension at 72°C for 30 s, followed by a final extension at 72°C for 5 min. Subsequently, 2 μL of the product was used as the input for the second-round PCR, which was performed in the same manner. Primer sets for the nested PCR were as follows: Gpr1-BSF1 (5′-GATTAGATTAGGTTAGTTTGGAA-3′) and Gpr1-BSR1 (5′-ACTAAAACACTAATCACCAAATA-3′) for the first round; Gpr1-BSF2 (5′-AGATTAGGTTAGTTTGGAATT-3′) and Gpr1-BSR2 (5′-AACACTAATCACCAAATAATTC-3′) for the second round. The second-round PCR product was subcloned and sequenced, as described previously [50]. The percentage methylation was calculated as 100×(number of methylated CpG dinucleotides)/(total number of CpGs). At least 10 clones from each parental allele were sequenced. Sequence data were analyzed using the QUMA quantification tool for methylation analysis [58].

Accession number

The MethylC-seq, WBA-seq, PBAT, and mRNA-seq data in this study have been deposited in the DNA Data Bank of Japan (DDBJ) under accession number DRA000484.

Supporting Information

Figure S1 Schematic of the SBS library construction procedure. MethylC-Seq libraries were generated by ligation of methylated sequencing adapters to fragmented genomic DNA followed by gel purification, sodium bisulfite conversion, and PCR amplification (*left*). WBA-seq libraries were generated by ligation of unmodified sequencing adapters to bisulfite-modified (amplified using EpiTect Whole Bisulfitome Kits) and fragmented genomic DNA followed by gel purification and PCR amplification (*middle*). PBAT libraries were generated by double-stranded DNA synthesis from bisulfite-treated (single-stranded) DNA with random primers containing sequencing adapters (*right*).
(TIF)

Figure S2 The percent of the oocyte and sperm genomes covered by differing minimum numbers of MethylC-seq and WBA-seq reads.
(TIF)

Figure S3 Sequencing bias towards mitochondrial and repetitive DNA sequences. (A) Average read depths for autosomal chromosomes and chromosome M (mitochondria) of mouse oocyte and sperm genomes. Occupancy of transposable elements in reads from SBS libraries before (B) and after (C) filtering the biased reads. (D) Genomic CpG coverage of SBS reads for each chromosome of mouse oocyte (orange: MethylC-seq, red: combined between MethylC-seq and WBA-seq) and sperm genomes (blue).
(TIF)

Figure S4 Average CpG methylation levels in genomic chromosomal DNA and mitochondrial DNA.
(TIF)

Figure S5 High-resolution DNA methylome map on mouse X inactivation center region in chromosome X (100,200,000–101,200,000). GenomeStudio view of Refseq's positions, repetitive element, CpG methylation map, CpG densities, CGI positions, and CGI methylation map were shown. Red, purple, blue, green,

and khaki dots and boxes represent the methylation levels at individual CpGs and CGIs in wild-type oocyte, *Dnmt3L*$^{-/-}$ oocyte, sperm, blastocyst, and ESC genomes, respectively, as shown in Figure 1.
(TIF)

Figure S6 DNA methylome maps of each chromosome of mouse germ cells. The methylation levels of each chromosome in wild-type oocytes, *Dnmt3L*$^{-/-}$ oocytes, and sperm in 10 kb windows (excluding mitochondrial chromosome, chromosome Y, and unplaced contigs). Red, purple, and blue lines represent the methylation levels in wild-type oocytes, *Dnmt3L*$^{-/-}$ oocytes, and sperm, respectively. Red and blue boxes represent oocyte-methylated and sperm-methylated gDMRs, and red and blue pins indicate maternal and maternal ICRs, respectively.
(TIF)

Figure S7 Methylation profiling of transposable elements in mouse germ cells. (A) CpG methylation levels are plotted as a function of CpG densities for L1 LINE, B1/Alu SINE, and LTR/ERVK retrotransposons (approximately 10% of the latter are intracisternal A particle (IAP) LTRs). Data for high CpG densities including less than 100 genomic CpGs were not plotted. (B) Bisulfite sequencing of L1 LINE, B1/Alu SINE, and IAP LTR retrotransposons. Methylated and unmethylated CpGs are indicated by open and closed circles, respectively.
(TIF)

Figure S8 Transcriptome and DNA methylome profiling at *H19-Igf2*, GenomeStudio view of mRNA-seq data (*top*) and CpG methylation map (*bottom*) of the genomic region spanning each locus. The blue shaded areas show the extent of the paternally-methylated gDMR.
(TIF)

Figure S9 Transcriptome and DNA methylome profiling at *Dnmt1*. The red shaded areas show the extent of the maternally-methylated gDMR.
(TIF)

Figure S10 Transcriptome and DNA methylome profiling at *Gpr1-Zdbf2*. The blue and red shaded areas show the extent of the paternally- and maternally-methylated gDMRs, respectively.
(TIF)

Figure S11 Quantification of the ratio of methylated (total number of read C) versus unmethylated cytosines (total number of read T) by PBAT results. Bar charts represent cytosine methylation ratio (A) at CpG (*left*), CpHpG (*middle*), and CpHpH (*right*) contexts and bisulfite-conversion failure rate (B) calculated by C:T ratio from lambda DNA mapping data. Total number of mapped reads is shown on these charts (*Top*).
(TIF)

Figure S12 Relationship between gene expression and intragenic methylation in ESCs. (A) The expression level of genes in ESCs was divided into 5 percentile groups. The distribution of methylation is shown ±5 kb from the transcription termination site (TTS; *left*) and transcription start site (TSS; *middle*). The graphs on the right show the average methylation levels in the promoter and gene-body regions. Spearman's rank correlation coefficient (ρ) was used to test the statistical significance of the correlation between gene expression and DNA methylation levels (*: $p<1\times10^{-9}$).
(TIF)

Figure S13 Expression profiles of DNA methyltransferase gene families. Red, purple, blue, and khaki bars represent RPKM values of individual genes in wild-type oocytes, *Dnmt3L*$^{-/-}$ oocytes, sperm, and ESCs.
(TIF)

Figure S14 Expression profiles of pluripotency-associated genes among wild-type oocytes, *Dnmt3L*$^{-/-}$ oocytes, sperm, and ESCs.
(TIF)

Figure S15 Histograms of the methylation levels of the demethylation-resistant (*left*) and demethylation-sensitive gDMRs (*right*) in ESCs. The number of oocyte-specific and sperm-specific methylated gDMRs is shown in red and blue, respectively.
(TIF)

Table S1 Gene transcript profiling for germ cells, blastocysts, and embryonic stem cells by mRNA-seq.
(XLSX)

Table S2 DNA methylation profiles of 23,021 CGIs.
(XLSX)

Table S3 Average DNA methylation profiles of 646 gDMRs determined by SBS and RRBS methods.
(XLSX)

Acknowledgments

We thank Terumi Horiuchi, Etsuko Sekimori, and Satoshi Sano for their assistance with analysis of the Illumina data. We are grateful to Fumihito Miura and Takashi Ito for their technical advice with the PBAT method and Takahiro Arima, Hitoshi Hiura, and Tom Moore for their helpful discussions and generous support.

Author Contributions

Conceived and designed the experiments: H Kobayashi, Y Suzuki, T Kono. Performed the experiments: H Kobayashi, T Sakurai. Analyzed the data: H Kobayashi, M Imai, Y Suzuki. Contributed reagents/materials/analysis tools: H Kobayashi, T Sakurai, M Imai, N Takahashi, A Fukuda, O Yayoi, S Sato, K Nakabayashi, K Hata, Y Sotomaru, Y Suzuki. Wrote the paper: H Kobayashi, T Kono.

References

1. Li E (2002) Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet 3: 662–673.
2. Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 33 Suppl: 245–254.
3. Lees-Murdock DJ, Walsh CP (2008) DNA methylation reprogramming in the germ line. Epigenetics 3: 5–13.
4. Sasaki H, Matsui Y (2008) Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nat Rev Genet 9: 129–140.
5. Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints. Science 294: 2536–2539.
6. Hata K, Okano M, Lei H, Li E (2002) Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development 129: 1983–1993.
7. Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, et al. (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature 429: 900–903.
8. Kono T, Obata Y, Wu Q, Niwa K, Ono Y, et al. (2004) Birth of parthenogenetic mice that can develop to adulthood. Nature 428: 860–864.
9. Hirasawa R, Chiba H, Kaneda M, Tajima S, Li E, et al. (2008) Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. Genes Dev 22: 1607–1616.
10. Bourc'his D, Bestor TH (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature 431: 96–99.
11. Webster KE, O'Bryan MK, Fletcher S, Crewther PE, Aapola U, et al. (2005) Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. Proc Natl Acad Sci U S A 102: 4068–4073.

12. Hata K, Kusumi M, Yokomine T, Li E, Sasaki H (2006) Meiotic and epigenetic aberrations in Dnmt3L-deficient male germ cells. Mol Reprod Dev 73: 116–122.
13. Shoji M, Tanaka T, Hosokawa M, Reuter M, Stark A, et al. (2009) The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. Dev Cell 17: 775–787.
14. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Takamatsu K, Chuma S, et al. (2010) MVH in piRNA processing and gene silencing of retrotransposons. Genes Dev 24: 887–892.
15. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, et al. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452: 215–219.
16. Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, et al. (2008) A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat Biotechnol 26: 779–785.
17. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133: 523–536.
18. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454: 766–770.
19. Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, et al. (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol 27: 361–368.
20. Brunner AL, Johnson DS, Kim SW, Valouev A, Reddy TE, et al. (2009) Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. Genome Res 19: 1044–1056.
21. Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, et al. (2009) Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat Biotechnol 27: 353–360.
22. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462: 315–322.
23. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, et al. (2010) Dynamic changes in the human methylome during differentiation. Genome Res 20: 320–331.
24. Popp C, Dean W, Feng S, Cokus SJ, Andrews S, et al. (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 463: 1101–1105.
25. Serre D, Lee BH, Ting AH (2010) MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res 38: 391–399.
26. Li Y, Zhu J, Tian G, Li N, Li Q, et al. (2010) The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol 8: e1000533. doi:10.1371/journal.pbio.1000533.
27. Gu H, Bock C, Mikkelsen TS, Jager N, Smith ZD, et al. (2010) Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. Nat Methods 7: 133–136.
28. Howlett SK, Reik W (1991) Methylation levels of maternal and paternal genomes during preimplantation development. Development 113: 119–127.
29. Farthing CR, Ficz G, Ng RK, Chan CF, Andrews S, et al. (2008) Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. PLoS Genet 4: e1000116. doi:10.1371/journal.pgen.1000116.
30. Barhanin J, Lesage F, Guillemare E, Fink M, Lazdunski M, et al. (1996) K(V)LQT1 and lsK (minK) proteins associate to form the I(Ks) cardiac potassium current. Nature 384: 78–80.
31. Shin J, Bossenz M, Chung Y, Ma H, Byron M, et al. (2010) Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice. Nature 467: 977–981.
32. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, et al. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 39: 457–466.
33. Edwards JR, O'Donnell AH, Rollins RA, Peckham HE, Lee C, et al. (2010) Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. Genome Res 20: 972–980.
34. Arnaud P, Hata K, Kaneda M, Li E, Sasaki H, et al. (2006) Stochastic imprinting in the progeny of Dnmt3L−/− females. Hum Mol Genet 15: 589–598.
35. Minami N, Tsukamoto S (2006) Role of oocyte-specific genes in the development of mammalian embryos. Reproductive Medicine and Biology 5: 175–182.
36. Chotalia M, Smallwood SA, Ruf N, Dawson C, Lucifero D, et al. (2009) Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev 23: 105–117.
37. Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr ARW, James KD, et al. (2010) Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. PLoS Genet 6: e1001134. doi:10.1371/journal.ppat.1001134.
38. Smallwood SA, Tomizawa SI, Krueger F, Ruf N, Carli N, et al. (2011) Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet 43: 811–814.
39. Tomizawa S, Kobayashi H, Watanabe T, Andrews S, Hata K, et al. (2011) Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. Development 138: 811–820.
40. Hiura H, Sugawara A, Ogawa H, John RM, Miyauchi N, et al. (2010) A tripartite paternally methylated region within the Gpr1-Zdbf2 imprinted domain on mouse chromosome 1 identified by meDIP-on-chip. Nucleic Acids Res 38: 4929–4945.
41. Hellman A, Chess A (2007) Gene body-specific methylation on the active X chromosome. Science 315: 1141–1143.
42. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, et al. (2010) Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A 107: 8689–8694.
43. Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328: 916–919.
44. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, et al. (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466: 253–257.
45. Ciccone DN, Su H, Hevi S, Gay F, Lei H, et al. (2009) KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. Nature 461: 415–418.
46. Ginis I, Luo Y, Miura T, Thies S, Brandenberger R, et al. (2004) Differences between human and mouse embryonic stem cells. Dev Biol 269: 360–380.
47. Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, et al. (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 448: 196–199.
48. Song F, Smith JF, Kimura MT, Morrow AD, Matsuyama T, et al. (2005) Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proc Natl Acad Sci U S A 102: 3336–3341.
49. Shiota K, Kogo Y, Ohgane J, Imamura T, Urano A, et al. (2002) Epigenetic marks by DNA methylation specific to stem, germ and somatic cells in mice. Genes Cells 7: 961–969.
50. Kobayashi H, Yamada K, Morita S, Hiura H, Fukuda A, et al. (2009) Identification of the mouse paternally expressed imprinted gene Zdbf2 on chromosome 1 and its imprinted human homolog ZDBF2 on chromosome 2. Genomics 93: 461–472.
51. De La Fuente R, Baumann C, Fan T, Schmidtmann A, Dobrinski I, et al. (2006) Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. Nat Cell Biol 8: 1448–1454.
52. Kaneda M, Hirasawa R, Chiba H, Okano M, Li E, et al. (2010) Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation. Genes to Cells 15: 169–179.
53. Shock LS, Thakkar PV, Peterson EJ, Moran RG, Taylor SM (2011) DNA methyltransferase 1, cytosine methylation, and cytosine hydroxymethylation in mammalian mitochondria. Proc Natl Acad Sci U S A 108: 3630–3635.
54. Iqbal K, Jin SG, Pfeifer GP, Szabo PE (2011) Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci U S A 108: 3642–3647.
55. Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, et al. (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun 2: 241.
56. Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, et al. (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature 471: 68–73.
57. Koide T, Moriwaki K, Uchida K, Mita A, Sagai T, et al. (1998) A new inbred strain JF1 established from Japanese fancy mouse carrying the classic piebald allele. Mamm Genome 9: 15–19.
58. Kumaki Y, Oda M, Okano M (2008) QUMA: quantification tool for methylation analysis. Nucleic Acids Res 36: W170–175.

DNA Research **20**, 241–253, (2013)
Advance Access publication on 9 April 2013
doi:10.1093/dnares/dst006

Robustness of Gut Microbiota of Healthy Adults in Response to Probiotic Intervention Revealed by High-Throughput Pyrosequencing

Seok-Won Kim[1], Wataru Suda[1], Sangwan Kim[1], Kenshiro Oshima[1], Shinji Fukuda[2,3,4], Hiroshi Ohno[3,4], Hidetoshi Morita[5], and Masahira Hattori[1,*]

Center for Omics and Bioinformatics, The Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwanoha 5-1-5, Kashiwa, Chiba 277-8561, Japan[1]; Institute for Advanced Biosciences, Keio University, Mizukami 246-2, Kakuganji, Tsuruoka City, Yamagata 997-0052, Japan[2]; Laboratory for Epithelial Immunobiology, RIKEN Research Center for Allergy and Immunology, Yokohama, Japan[3]; Graduate School of Nanobioscience, Yokohama City University, Yokohama 230-0045, Japan[4] and School of Veterinary Medicine, Azabu University, Fuchinobe 1-17-71, Chuo-ku, Sagamihara, Kanagawa 252-5201, Japan[5]

*To whom correspondence should be addressed. Tel. +81 4-7136-4070. Fax. +81 4-7136-4084.
Email: hattori@k.u-tokyo.ac.jp

Edited by Dr Katsumi Isono
(Received 12 December 2012; accepted 16 February 2013)

Abstract

Probiotics are live microorganisms that potentially confer beneficial outcomes to host by modulating gut microbiota in the intestine. The aim of this study was to comprehensively investigate effects of probiotics on human intestinal microbiota using 454 pyrosequencing of bacterial 16S ribosomal RNA genes with an improved quantitative accuracy for evaluation of the bacterial composition. We obtained 158 faecal samples from 18 healthy adult Japanese who were subjected to intervention with 6 commercially available probiotics containing either *Bifidobacterium* or *Lactobacillus* strains. We then analysed and compared bacterial composition of the faecal samples collected before, during, and after probiotic intervention by Operational taxonomic units (OTUs) and UniFrac distances. The results showed no significant changes in the overall structure of gut microbiota in the samples with and without probiotic administration regardless of groups and types of the probiotics used. We noticed that 32 OTUs (2.7% of all analysed OTUs) assigned to the indigenous species showed a significant increase or decrease of ≥10-fold or a quantity difference in >150 reads on probiotic administration. Such OTUs were found to be individual specific and tend to be unevenly distributed in the subjects. These data, thus, suggest robustness of the gut microbiota composition in healthy adults on probiotic administration.

Key words: probiotics; gut microbiota; 16S ribosomal RNA gene; pyrosequencing

1. Introduction

Probiotics are defined as live bacterial strains conferring various benefits to the consumer by modulating the intestinal ecosystem, thereby potentially promoting host health and improving host disease risk.[1–11] Various probiotic strains have been industrially developed and marketed as a variety of products and applications such as fermented foods and supplements, including yogurt[12–15] Most probiotics taxonomically belong to two genera, *Bifidobacterium* and *Lactobacillus*, that originate from various environments, including the human intestine, and both species are generally regarded as safe.[16–18]

The interaction between administrated probiotics and indigenous microbiota is one of the most attractive and important research areas, particularly because gut microbiota have been shown to be profoundly

associated with various host physiology states, including disease, diet, and age through the shift of bacterial composition, as well as metabolic and nutritional processes.[19–23] The ability of probiotics to survive through the intestine and to modulate gut microbiota is a critical factor in determining their potential for health-related outcomes.

There have been a large number of probiotic intervention studies to assess the impact of probiotics on gut microbiota in healthy adults,[24–34] infants, and children,[35,36] and in clinical trials on patients with a variety of diseases.[37,38] Most probiotic intervention studies were carried out by comparison between probiotic-treated groups and placebo controls and examined only one or two samples from periods before and during intervention or post-intervention for each subject. These experimental designs make evaluation of results obscure from a statistical viewpoint due to the high inter-individual variability of gut microbiota.[4] In addition, most of the analyses focussed on the composition of specific bacterial species or groups by conventional methods such as culturing, quantitative polymerase chain reaction (qPCR), fluorescence *in situ* hybridization, denaturing gradient gel electrophoresis, or terminal-restriction fragment length polymorphism based on the bacterial 16S ribosomal RNA gene (16S). These conventional methodologies may also overlook subtle changes in bacterial community structure and change of species other than targeted species. Thus, the effect of probiotic administration on the overall structure of gut microbiota is largely unknown.

Recently, a high-throughput sequencing-based analysis has been conducted for gut microbiota fed with a probiotic yogurt that provided new insights into probiotics research by utilizing a large-scale dataset.[39] However, much more data are required to understand the impact of probiotics on gut microbiota. Recent advances in sequencing technology have enabled us to elucidate complex bacterial communities, including human gut microbiota.[40,41] Particularly, 454 pyrosequencing of bacterial 16S gene tags coupled with bioinformatics provides a high-throughput and cost-effective approach for the comprehensive analysis of bacterial communities at the species level.[42–48]

In this study, we developed an analysis pipeline for bacterial communities based on barcoded 454 pyrosequencing of 16S gene tags using modified PCR primers that improved the quantitative accuracy of inferred species composition in human gut microbiota. Using this pipeline, we analysed faecal samples longitudinally collected from individuals with and without probiotic administration to evaluate the effect of probiotics on gut microbiota with respect to species richness and diversity. The results revealed the robustness and stability of gut microbiota of healthy adults in response to probiotic administration.

2. Materials and methods

2.1. *Subjects, faecal sample collection, and probiotic intervention*

Eighteen healthy volunteers (age: 22 ± 3.16 yrs, 6 male, 12 female) were recruited through Azabu University, Kanagawa, Japan (Supplementary Table S1). All subjects were informed of the purpose of this study. This study was approved by the ethical committee of Azabu University, and written consent was obtained from all subjects. No subjects were treated with antibiotics during faecal sample collection. The subjects were divided into six groups (three subjects per group), and each group consumed six different commercially available probiotics supplied from Yakult Honsha Co., Ltd, Kagome Co., Ltd, Morinaga Milk Industry Co., Ltd, Takanashi Milk Products Co., Ltd, Meiji Co., Ltd, and Danone Japan Co., Ltd, respectively (Supplementary Table S1). The number of each bacterial strain contained in the probiotic products was estimated as the genome equivalent by qPCR of 16S ribosomal RNA genes using 27Fmod-338R, followed by pyrosequencing of the 16S amplicons (see below). The genome equivalent per gram or millilitre and the total genome equivalent of each bacterial strain in one probiotic product are summarized in Supplementary Table S1. Three subjects in each group consumed the same probiotics daily for 8 weeks according to the schedule of sampling and probiotic intervention (Supplementary Fig. S1). Faecal samples from 4 weeks before (S00) and 8 weeks during probiotic intervention (S01–S04), and 8 weeks after cessation of probiotic intervention (S05–S08), were collected every 2 weeks from each subject. In total, we collected 158 faecal samples from the 18 subjects because we could not collect 1 sample each from 4 of the subjects.

2.2. *Recovery of bacteria from faecal samples*

Freshly collected faeces (1.0 g) were suspended in 20% glycerol (Wako Pure Chemical Industries, Ltd) and phosphate buffered saline solution (Life Technologies Japan, Ltd, Tokyo, Japan), frozen in liquid nitrogen, and stored at −80°C until ready for use. Bacterial pellets were prepared from frozen faecal samples as described previously.[49]

2.3. *DNA isolation from bacteria*

Faecal DNA was isolated and purified according to the literature, with minor modifications.[49] The bacterial pellet was suspended and incubated with 15 mg/ml lysozyme (Sigma-Aldrich Co., LCC) at

37°C for 1 h in TE10. Purified achromopeptidase (Wako Pure Chemical Industries, Ltd) was added at a final concentration of 2000 units/ml and then incubated at 37°C for 30 min. The suspension was treated with 1% (wt/vol) sodium dodecyl sulphate and 1 mg/ml proteinase K (Merck Japan) and incubated at 55°C for 1 h. The lysate was treated with phenol/chloroform/isoamyl alcohol (Life Technologies Japan, Ltd). DNA was precipitated by adding ethanol and pelleted by centrifugation at 3,300 *g* at 4°C for 15 min. The DNA pellet was rinsed with 75% ethanol, dried, and dissolved in 10 mM Tris-HCl/1 mM EDTA (TE). DNA samples were purified by treating with 1 mg/ml RNase A (Wako Pure Chemical Industries, Ltd) at 37°C for 30 min and precipitated by adding equal volumes of 20% polyethylene glycol solution (PEG6000-2.5M NaCl). DNA was pelleted by centrifugation at 8,060 *g* at 4°C, rinsed with 75% ethanol, and dissolved in TE.

2.4. *454 barcoded pyrosequencing of 16S rRNA gene*

The V1–V2 region of the 16S rRNA gene was amplified using forward primer (5′-CCATCTCATCCCTGCG TGTCTCCGACTCAGNNNNNNNNNNagrgtttgatymtggc tcag-3′) containing the 454 primer A, a unique 10-bp barcode sequence for each sample (indicated in N), and 27Fmod (5′-agrgtttgatymtggctcag) in which the third base A in the original primer 27F was changed to R, and reverse primer (5′-CCTATCCCCTGTGTGC CTTGGCAGTCTCAGtgctgcctcccgtaggagt-3′) containing the 454 primer B and reverse primer 338R (5′-tgctgcctcccgtaggagt). PCR was performed in 1 × Ex Taq PCR buffer (50 μl), deoxynucleoside triphosphate (2.5 mM), Ex Taq polymerase (Takara Bio, Inc., Shiga), each primer (10 μM), and 40 ng of extracted DNA under conditions of 2 min at 96°C, 20 cycles of 96°C for 30 s, 55°C for 45 s, and 72°C for 1 min, and a final extension of 72°C for 10 min on a 9700 PCR system (Life Technologies Japan, Ltd, Tokyo, Japan). PCR products of approximately 370 bp were confirmed by agarose gel electrophoresis, purified by AMPure XP magnetic purification beads (Beckman Coulter, Inc., Brea, CA, USA), and quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies Japan, Ltd, Tokyo, Japan). Mixed samples were prepared by pooling approximately equal amounts of PCR amplicons from each sample and subjected to 454 GS FLX Titanium or 454 GS JUNIOR (Roche Applied Science) sequencing according to the manufacturer's instructions.

2.5. *Analysis pipeline for 454 barcoded pyrosequencing of 16S PCR amplicons*

We developed an analysis pipeline for 454 barcoded pyrosequencing of PCR amplicons of the V1-2 region amplified by 27Fmod-338R primers. First, 16S reads were assigned to each sample based on the barcode sequence information. Second, 16S reads that did not have PCR primer sequences at both sequence termini and those with an average quality value < 25 were filtered out. Third, 16S reads containing possible chimaeric sequences that had BLAST match lengths of $< 90\%$ with reference sequences in the database were removed. Reads removed in these processes accounted for about 35% of all reads, most of which represented reads lacking PCR primer sequences (Supplementary Table S2). Finally, filter-passed reads were obtained for further analysis by trimming off both primer sequences.

All 3000 filter-passed reads of the 16S V1-2 sequences obtained from each subject were deposited in DDBJ/GenBank/EMBL with accession numbers DRA000869–DRA000886.

2.6. *Assessment of the quantitative accuracy of 16S data using artificial bacterial communities*

Two artificial bacterial communities (designated 'mock01' and 'mock02') were constructed by mixing genomic DNA from 10 and 11 different human gut-associated bacterial strains with an appropriate ratio, respectively (Supplementary Table S3). Genome sequences of these microbes were completely sequenced and are publicly available. From these communities, we amplified the V1-2 region by PCR using 27F-338R and 27Fmod-338R primers, the V5-6 region by 787F-1061R primers, and the V1-9 region by 27F-1492R primers. V1-2 and V5-6 amplicons were subjected to 454 pyrosequencing, and V1-9 amplicons were cloned in *Escherichia coli*, and 3000 clones were sequenced by the Sanger method, and the products were analysed with the ABI3730xl (Life Technologies Japan, Ltd, Tokyo). We also performed duplicate qPCR experiments targeting a specific genomic region of the bacterial strains in the two mock communities. All filter-passed 16S *de novo* sequences and qPCR data were then analysed by principal component analysis (PCA) to compare and assess the quantitative accuracy (Fig. 1).

The error rate of the filter-passed sequences using 27Fmod-338R primers obtained from the two mock communities was estimated by aligning the 16S V1-2 *de novo* sequences with the reference 16S sequences in the two mock communities (Supplementary Table S4).

2.7. *Data analysis*

2.7.1. Database Two databases were constructed for the analysis of 16S sequences. One is the 16S rRNA gene sequence database constructed

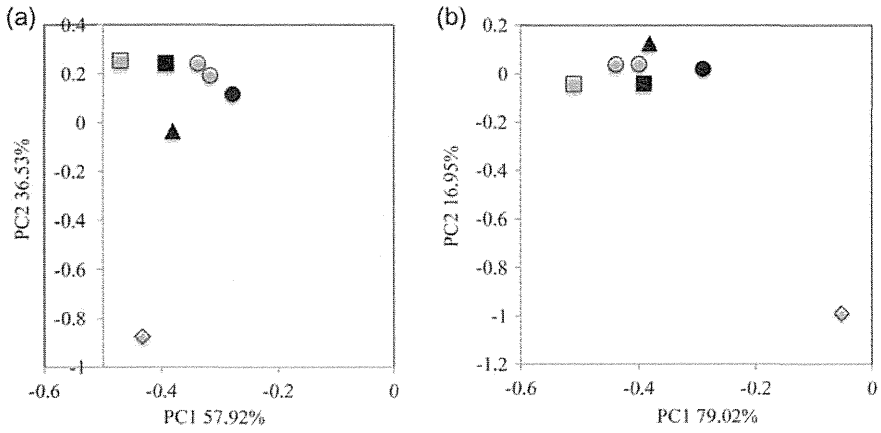

Figure 1. Assessment of the quantitative accuracy of the analysis of the bacterial composition of two mock communities by various methods. PCA analysis of the data was obtained from various methods using mock01 (a) and mock02 (b). Closed circle: expected, open circle: duplicate qPCR, closed square: pyrosequencing of 16S V1-2 region using 27Fmod, open square: pyrosequencing of 16S V1-2 region using 27F, closed triangle: pyrosequencing of 16S V5-6 region, open diamond: Sanger sequencing of nearly full-length 16S clone.

by collecting 16S sequences of $\geq$1200 bp of bacterial isolates from the Ribosomal Database Project v. 10.27. Another database is the reference genome database constructed by collecting genome sequences from the NCBI FTP site (ftp://ftp.ncbi.nih.gov/genbank/, Dec 2011) that includes 1482 complete and 605 draft bacterial genomes.

2.7.2. Operational taxonomic unit (OTU) and UniFrac distance analysis We used 3000 filter-passed reads of 16S sequences for operational taxonomic unit (OTU) and UniFrac distance analysis for each sample. For OTU analysis, clustering of 16S reads was done using a 96% pairwise-identity cutoff with the UCLUST program (www.drive5.com). Representative sequences for each OTU were assigned to bacterial species by BLAST search with a 96% pairwise-identity cutoff against the two databases mentioned above. UniFrac distance analysis was used to determine the dissimilarity (distance) between two communities based on the fraction of branch length shared between two communities within a phylogenetic tree constructed from 16S sequence datasets.[44]

2.7.3. Other Estimation of OTU numbers by extrapolation (Chao1 and ACE) was calculated with the vegan package (v2.0-5) for R (v2.15.2).

3. Results and discussion

3.1. Quantitative accuracy of 16S data produced by 454 pyrosequencing

Pyrosequencing of PCR amplicons of bacterial 16S short variable regions is the most popular and a high-throughput approach to infer and characterize the species composition in bacterial communities.[42,45,46,48] The 454 pyrosequencing platform, which can produce over 400 bases per read, is also superior to shorter read-length sequencers with respect to sequence accuracy for single-end sequencing.[50,51] However, this PCR-based method has a problem particularly in quantification of the composition of the genus *Bifidobacterium*, a dominant species in human gut microbiota because the 16S sequence of *Bifidobacterium* has a few base mismatches with the commonly used PCR primer 27F (or 8F), underestimating this genus in the community.[52–55] To improve this, we modified primer 27F to 27Fmod by changing the third base G to R (G or A) in 27F-YM[53] that perfectly matched with the annealing site of the *Bifidobacterium* 16S gene (see Materials and methods).

To assess the 16S data using 27Fmod, we compared various 16S sequence and qPCR data obtained from two mock communities (Supplementary Table S3) that are useful to evaluate the quantitative accuracy of 16S-based data and the sequencing error rate.[56,57] Quantitative accuracy of the overall bacterial composition was evaluated by comparing the similarity of each data to the expected ('Expected') using PCA (Fig. 1). From the PCA data, Euclidean distance was calculated for evaluation of the similarity of each data with the 'Expected'. The results revealed that the order of their similarities with the 'Expected' was the qPCR data $\geq$ the V1-2 data using 27Fmod > the V5-6 data > the V1-2 data using 27F >> the data of Sanger sequencing-based full-length V1-9, indicating that the use of 27Fmod greatly improved the quantitative accuracy for evaluation of the overall

bacterial composition (Supplementary Table S5). This improvement was largely dependent on the improved estimation of the *Bifidobacterium* content by the use of 27Fmod. The average relative *Bifidobacterium* content in the two mock communities estimated from the data of V1-2 using 27F was only 1.5% of the 'Expected' (100%), whereas the use of 27Fmod increased the relative content to 61% that was also better than that estimated from the data of V5-6 and Sanger full-length analyses (Supplementary Fig. S2). Because qPCR can be used only when genomes of all bacteria in a given community are known, or only for a limited number of specific known species, we concluded that 454 pyrosequencing of the V1-2 region using 27Fmod-338R provided more quantitatively accurate data for bacterial composition in human gut microbiota than that using the conventional 27F primer.

We estimated the average error rate of filter-passed V1-2 data using 27Fmod-338R by aligning the V1-2 and reference 16S sequences of bacterial strains used in the two mock communities. The error rate was estimated to be 0.58 and 0.66% for mock01 and mock02 by local alignment, respectively (Supplementary Table S4). These error rates are similar to the previously published data,[43,45,50] but lower than in another study.[58] The latter may be due to differences in the examined alignment length and between local and global alignments. Errors in 454 pyrosequencing data can be the primary cause for overestimation of the OTU number that is an issue which needs to be improved for accurate estimation of species richness in bacterial communities.[59,60] We compared OTU numbers generated from clustering of various qualities of 16S reads with a 96% and a 97% pair-wise identity cutoff. For this comparison, we made and used three datasets: only primer check-passed reads having the highest error rates, filter-passed reads, and selected filter-passed reads having the lowest error rates. The results indicated that a 96% cutoff clustering of error-rich reads and a 97% cutoff clustering of filter-passed reads gave the worse results than a 96% cutoff clustering of filter-passed and selected filter-passed reads (Supplementary Fig. S3). A 97% cutoff was defined for clustering of highly accurate Sanger full-length 16S sequences.[61] Therefore, in clustering of pyrosequencing data having higher error rate than Sanger data, the use of a cutoff identity lower than 97% and a lower number of reads are reasonable to reduce overestimation of the OTU number. A 96% cutoff clustering of filter-passed reads gave similar OTU numbers up to 30–50 reads to those of filter-passed reads having the lowest error rates. These read numbers are approximately three to five times the number of input strains. After several trials testing the mock communities, we decided to use 3000–5000 reads per sample for clustering with a 96% cutoff for the analysis of human gut microbiota. Indeed, OTU numbers using a 96% cutoff clustering of 3000 reads decreased about 15% when compared with those using a 97% cutoff clustering.

3.2. *Species richness and diversity in human faecal microbiota with probiotic intervention*

We randomly selected 3000 reads of 16S V1-2 sequences from all filter-passed reads for each sample (Supplementary Table S2) and used 474 000 reads in total from 158 faecal DNA samples of 18 subjects for the analysis of species richness and composition in human gut microbiota. Clustering of all reads with a 96% pairwise-identity cutoff gave a total of 2758 OTUs. After removing the minority OTUs having $<0.1\%$ abundance in any samples, 1175 OTUs having $\geq 0.1\%$ abundance in at least 1 sample, accounting for 99.1% of all 16S reads, were used for further analysis.

3.2.1. Detection of administrated probiotic strains in faecal samples We investigated whether administrated strains contained in the probiotic products can be detected in faecal DNA. We sequenced the 16S V1-2 region of all bacterial strains contained in probiotic products used in this study. The BLAST search to the databases indicated that except for the *Bifidobacterium longum* strain used in Group III, the 16S sequences of all strains in the probiotic products significantly differed from those of the indigenous species phylogenetically closest to the probiotic strains. The 16S sequence of the *B. longum* strain used in Group III was almost identical to that of an indigenous *Bifidobacterium* species, so that we used a distinguishable additive *Lactococcus lactis* strain in this product for the detection of administrated bacteria in Group III samples. The 16S sequences of these probiotic strains were included in the databases constructed in this study, and the 16S reads assigned to administrated strains had the average similarity between 99.4 and 99.9% identities with the reference sequences (data not shown). The 16S reads assigned to probiotic or additive strains were detected in samples (S01–S04) during probiotic intervention [designated 'Pro(+)'] at various frequencies, but almost none were detected in samples (S00 and S05–S08) without probiotic administration [designated 'Pro(−)'] (Supplementary Table S6). The administrated probiotic strains were shown to be more frequently detected in samples during the intervention than in the pre- and post-intervention periods using different detection methods such as culturing, targeted PCR, and hybridization.[24,26–28,30,32,33,62] In

the present study, two probiotic *Lactobacillus* and one additive *Lactococcus* strains were detected in post-intervention samples in three subjects with a minimum count, respectively. The similarity of three 16S sequences was 99.4, 99.7, and 100% identity with those of administrated *Lactobacillus* and *Lactococcus* strains, indicating that these are administrated strains. The survival of some probiotic strains in the post-intervention period was also reported previously.[28,30] Our data suggested that some probiotic strains seem to be able to persistently colonize the intestine and their survivability may be related to metabolic activity in the intestine.[63,64] Probiotic *Bifidobacterium* strains were not detected in any Pro(−) samples. However, we found two distinct 16S sequences both assigned to *Bifidobacterium animalis* in two subjects APr37 and APr39. One showed a high similarity of >98% identity with the 16S sequence of the administrated *B. animalis* and was detected with high frequency only in the Pro(+) samples, whereas another showed a low similarity of 96.5–97.4% identity (a mean of 97.2%) with low frequency in both the Pro(−) and Pro(+) samples. These data suggest the presence of unknown indigenous species phylogenetically close to, but distinct from, probiotic *B. animalis* in human gut microbiota. The total number of bacteria contained in each probiotic product was varied between 10^9 and 10^{10}, showing no large difference in quantity among them (Supplementary Table S1). No clear correlation was also observed between the number of bacteria in the products and the frequency in detection of the administrated strains in the Pro(+) samples. From these observations, the frequency of administrated bacteria detected in faeces may not be largely affected by their amounts in the products. Therefore, detection of *Lactobacillus brevis* and *Lactobacillus delbrueck* at relatively low level in faeces cannot be simply explained by the difference in a dose, but could be considered the association with several factors such as their survivability in the intestine, diet, or physiological conditions of subjects.

3.2.2. Change of species richness in samples with and without probiotics We analysed species richness (OTU number) in the Pro(+) and Pro(−) samples. Supplementary Figure S4 shows the change in OTU numbers for every sample in each subject, indicating that OTU numbers vary dramatically for every sample. Most of the variation can be attributed to single OTUs representing the minority species. We averaged the OTU numbers of the Pro(−) and Pro(+) samples and compared them for subject, group, type of probotics (*Lactobacillus* and *Bifidobacterium*), and all combined samples, respectively (Fig. 2). The average OTU numbers in 6 out of 18 subjects were decreased in the range of the ratio of 0.83–0.95 in the Pro(+) samples when compared with the Pro(−) samples, whereas those in other 12 subjects were increased in the range of the ratio of

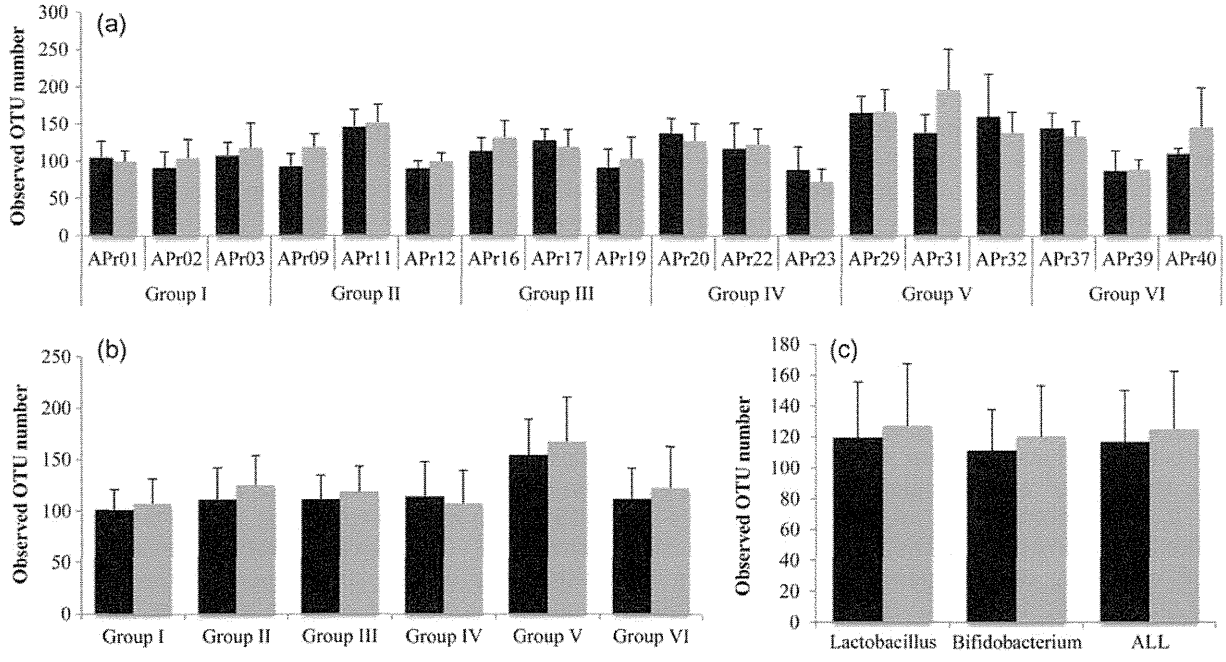

Figure 2. Change in OTU number in faecal microbiota with and without probiotic administration. (a) Individual, (b) group, (c) type of probiotics. Black bar indicates Pro(−) samples. Grey bar indicates Pro(+) samples. The error bars represent standard deviation.

1.01–1.43. For group, only Group IV showed a decrease in the average OTU number in the Pro(+) samples with the ratio of 0.94. For type of probiotics and all samples, the average OTU numbers in the Pro(+) samples were slightly more abundant (approximately 1.07-fold) than those in the Pro(−) samples, but no statistical significance was observed in any dataset. The increase in OTU number in the Pro(+) samples was largely due to the minority species (Supplementary Fig. S4), whereas the abundance of the majority species (OTUs containing ≥10 reads) was almost constant over time. We performed the same analysis using different sets of 3000 reads for each subject. The analysis reproducibly showed the similar pattern and the degree of the change in OTU numbers to which the minority species is largely attributed (data not shown). These data indicate that administration of probiotics tends to increase species richness in faecal microbiota that may be beneficial for the consumer because the species richness in faecal microbiota of subjects afflicted with disease such as inflammatory bowel disease is significantly reduced when compared with that of healthy subjects.[65]

3.2.3. Change of species composition in samples with and without probiotics We obtained the average weighted and unweighted UniFrac distances within Pro(−), within Pro(+), and between Pro(−) and Pro(+) samples for every group, probiotic types, and all subjects, respectively (Fig. 3). High UniFrac distance implies high variability of microbiota structure within and between samples. If the difference between any pair of the three distances is statistically significant, it can be considered that probiotic administration significantly affected the overall microbiota composition. We found the largest difference between weighted UniFac distances of the Pro(+)

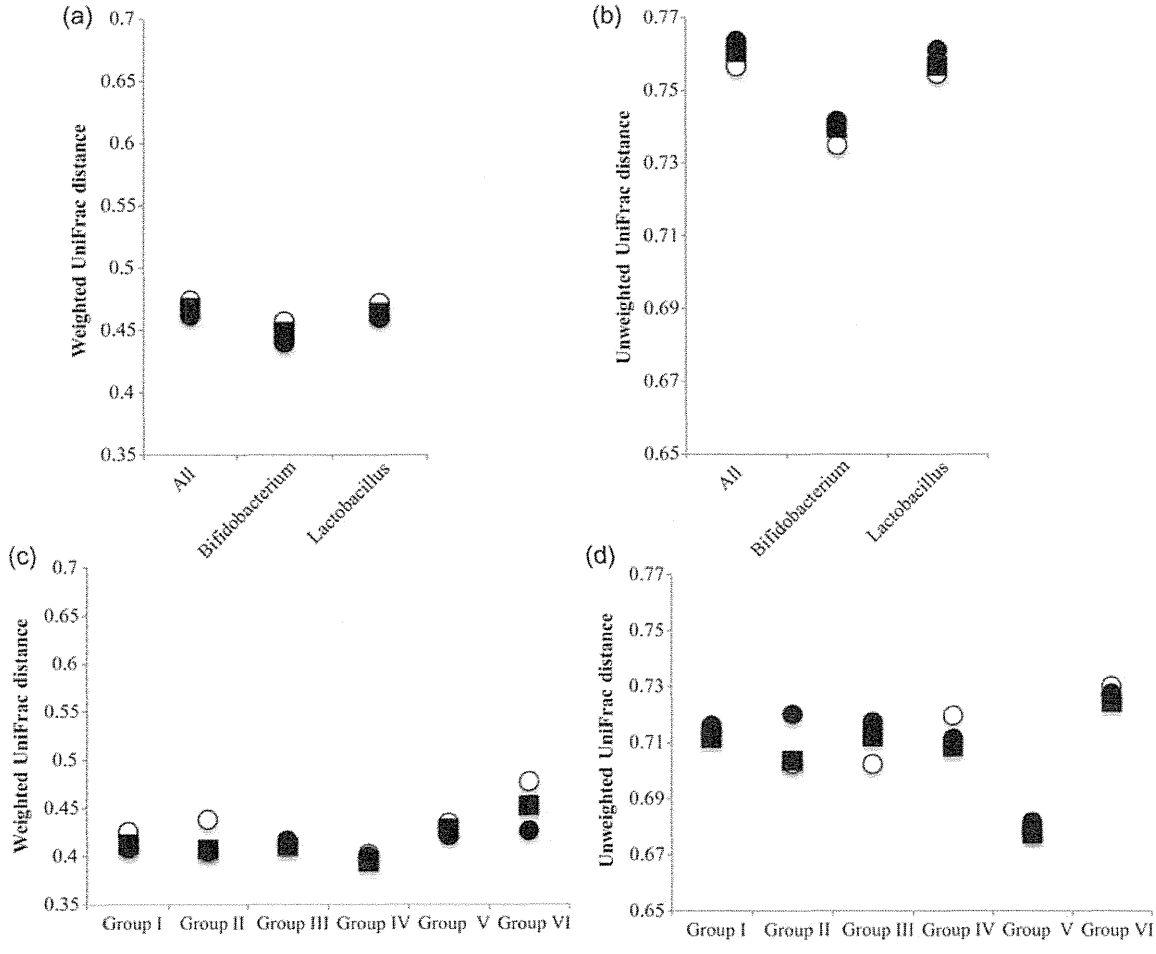

Figure 3. Average UniFrac distance within Pro (−) and Pro(+) and between Pro(−) and Pro(+) for each group, type of probitotics, and all subjects. Average UniFrac distance between any pair of the three distances for type of probiotics and all subjects (a and b), and each group (c and d). Open circle, closed circle, and closed square indicate average UniFrac distance within Pro (−), within Pro (+), and between Pro(−) and Pro(+) samples, respectively.

and Pro(−) samples in Group VI. However, statistical evaluation of this difference by the Student's *t*-test showed no significance (*P*-value > 0.05) for 781 out of 1000 times (Supplementary Table S7). These data imply high stability of gut microbiota to probiotic administration for all subjects examined. We also analysed UniFrac distances of intra-subject gut microbiota (Fig. 4). Although 5 subjects (APr02, 12, 16, 37, and 39) showed a significant difference in the UniFrac distances between Pro(−) and Pro(+) samples, the results showed that both weighted and unweighted distances between Pro(−) and Pro(+) of all intra-subjects were significantly lower than the average distance of the 18 unrelated subjects. The Welch's *t*-test for these differences showed statistical significance (Supplementary Table S8). We also performed the UniFrac distance analysis using different 16S datasets of 5000 reads for group, type of probiotics, all subjects, and intra-subject. The results similarly showed no statistical significance in differences between any pair of the 3 UniFrac distances and the significantly lower UniFrac distance of each intra-subject than that of the 18 unrelated subjects (data not shown). Thus, these data suggested that the perturbation of microbiota elicited by probiotics in an intra-subject did not overcome the inter-subject variations of gut microbiota, supporting high intra-specificity and stability of gut microbiota.[66,67] This robustness of gut microbiota of adults is in contrast with the profound effect of antibiotic administration on adult gut microbiota[68] and the observed response of gut microbiota of infants fed with probiotics, in which the infant gut microbiota composition was considerably affected by probiotics.[36] A short-term dietary intervention study showed that in controlled feeding of the same diet to subjects over 10 days, a marked change was observed within 1 day after the intervention initiation.[69] In the present study, no significant difference was observed between samples before (S00) and first samples (S01) after the intervention initiation (data not shown). It would be valuable to analyse faecal samples collected within a few days after administration of probiotics for evaluation of the short-term effect of probiotics.

4. Identification of bacterial species showing significant increase or decrease by probiotic administration

Although our results suggested that administration of probiotics had almost no effect on the overall structure of gut microbiota, it is possible to identify bacterial species largely responding to the administrated probiotics at the OTU/species level. We surveyed OTUs showing an increase or a decrease between the Pro(+) and Pro(−) samples by comparing the number of 16S reads for each OTU. We first enumerated the OTUs showing ≥2-fold change between the

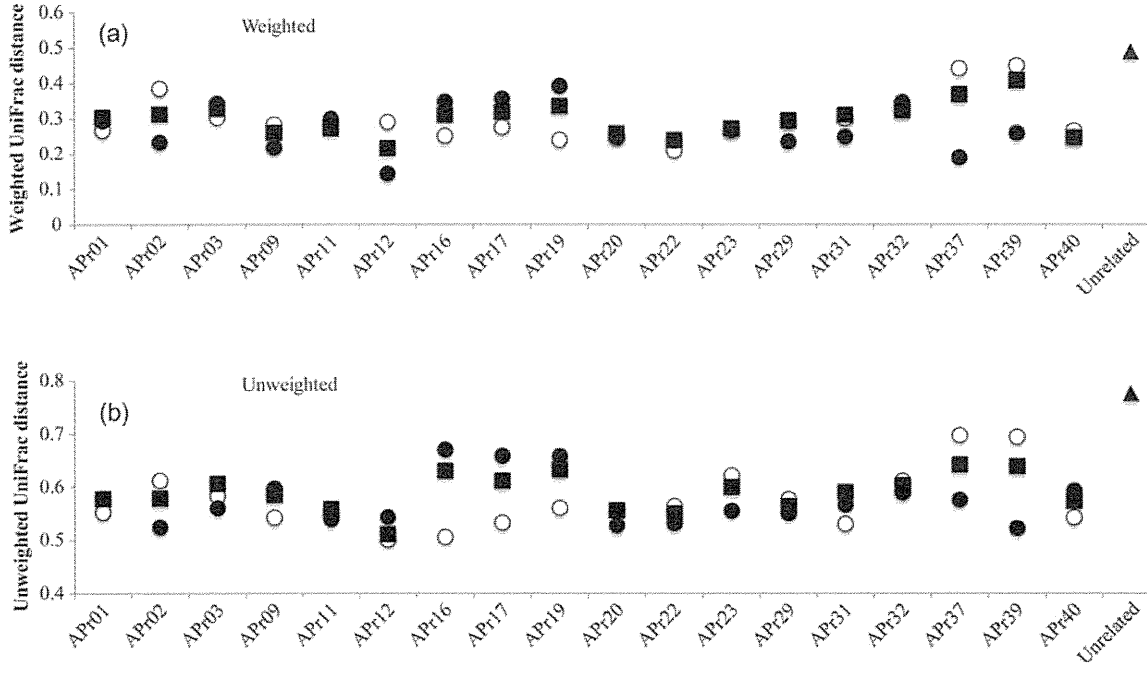

Figure 4. Average UniFrac distance within Pro(−) and Pro(+) and between Pro(−) and Pro(+) for each subject. Open circles, closed circles, and closed squares indicate average UniFrac distance within Pro(−), within Pro(+), and between Pro(−) and Pro(+) samples, respectively. Closed triangles indicate average UniFrac distance between samples (S00) of 18 unrelated individuals.

Pro(−) and Pro(+) samples for each subject, and the quantity difference was also obtained by subtracting the 16S read number of the Pro(+) samples from that of the Pro(−) samples. This is because OTUs showing a high quantity difference, but less fold change may also have substantial influence on gut microbiota composition. We found several OTUs significantly changed by probiotic administration, including OTUs assigned to both the indigenous and administrated strains (Fig. 5). We listed 88 OTUs (7.5% of all analysed 1175 OTUs) showing significant change of ≥3-fold, among which 30 OTUs changed by ≥10-fold (Supplementary Fig. S5). We excluded 6 OTUs assigned to the administrated strains from the 30 OTUs and obtained 24 OTUs assigned to the indigenous species, including OTU00072 assigned to *Streptococcus salivarius* that showed significant change in 2 subjects (Supplementary Table S9). We also found seven OTUs showing significant difference in quantity between both samples (Supplementary Table S10). Of the combined 32 OTUs (2.7%), 18 were increased and 14 were decreased by probiotic administration. Many of the OTUs showing a significant increase were assigned to minority species in the Pro(−) samples, but some increased up to nearly 7% in abundance (e.g. OTU00372 assigned to *Eubacterium rectale*). On the other hand, the OTUs showing a significant decrease were almost undetected in the Pro(+) samples. Phylum-level species assignment showed that species belonging to the phylum *Firmicutes* were most largely affected by both probiotics, and all species belonging to the phylum *Bacteroidetes* were affected only by *Lactobacillus* probiotics (Table 1). The 32 OTUs were assigned to 27 indigenous species, among which 4 species (*Clostridium clostridioforme*, *Eubacterium eligens*, *E. rectale*, and *Faecalibacterium prausnitzii*) were assigned by 8 different OTUs and 1 species (*S. salivarius*) was assigned by the 2 same OTUs as described above. All these species except for *S. salivarius* were found to show significant change only in one subject, indicating that response of the indigenous species to probiotics is highly individual specific (Supplementary Fig. S6). Two different OTUs (OTU02677 and OTU02748) assigned to *F. prausnitzii*, of which the reduction is known to be correlated with inflammatory bowel disease,[70] were found to both decrease and increase in the same subject (APr40) by probiotic administration, suggesting that these two phylogenetically close species may have the diversity of response to probiotic action. We also examined distribution of the 32 OTUs in the subjects. The results revealed that 4 subjects

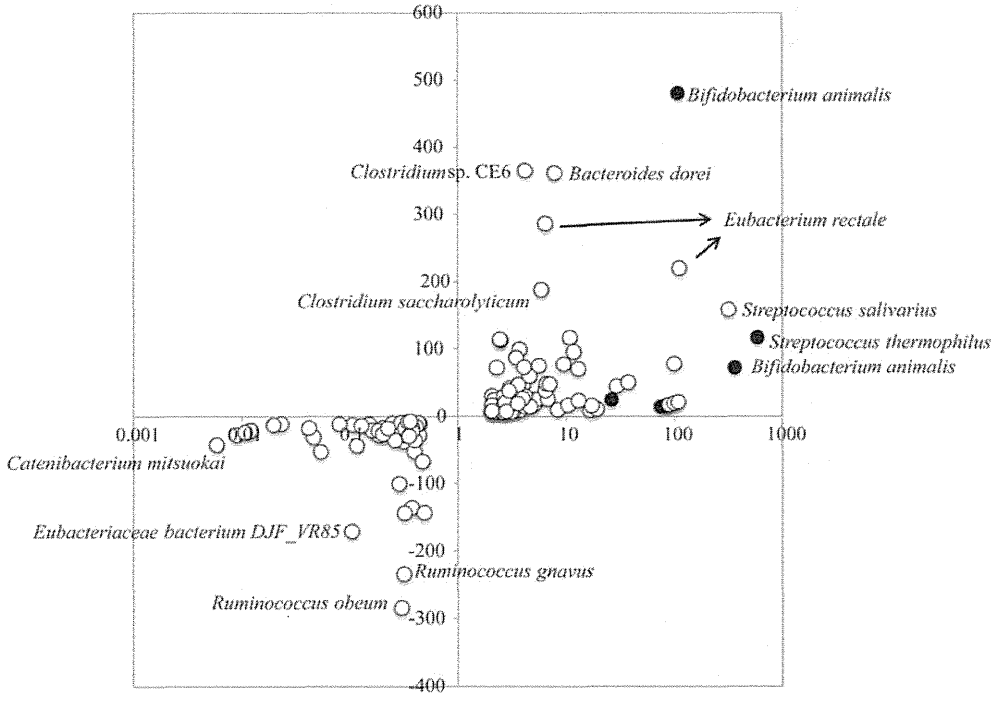

Figure 5. OTUs showing ≥2-fold change and their difference in quantity between the Pro(−) and Pro(+) samples. The *x*-axis represents the scale of fold change between the Pro(+) and Pro(−) samples. The *y*-axis represents the difference (number of reads) in quantity between the Pro(+) and Pro(−) samples. Closed and open circles indicate the administrated probiotic and indigenous species, respectively.

Table 1. Phylum-level species assignment of OTUs showing significant fold change or quantity difference by administration of probiotics

Type of probiotics	Change	[a]Number of varied OTUs	Fold change (≥10-fold)				Number of varied OTUs	Difference (≥150 reads)	
			Firmicutes	Actinobacteria	Bacteroidetes	Unclassified bacterium		Firmicutes	Bacteroidetes
Lactobacillus	Increase	9	7	0	1	1	3	2	1
	Decrease	7	3	1	3	0	1	1	0
	Total	16	10	1	4	1	4	3	1
Bifidobacterium	Increase	5	5	0	0	0	1	1	0
	Decrease	4	4	0	0	0	2	2	0
	Total	9	9	0	0	0	3	3	0
All	Increase	14	12	0	1	1	4	3	1
	Decrease	11	7	1	3	0	3	3	0
	Total	25	19	1	4	1	7	6	1

[a]Administrated probiotic strains were excluded, and only OTUs with a P-value < 0.05 are shown.

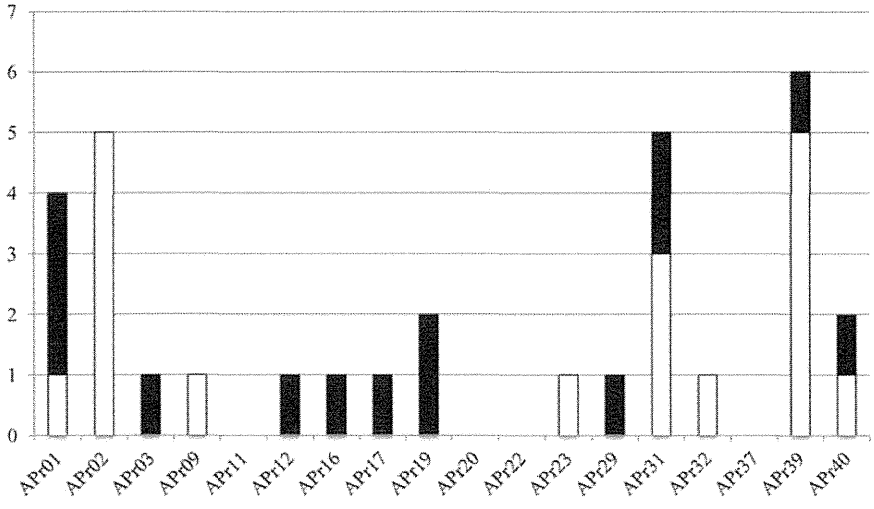

Figure 6. Distribution of 32 OTUs showing a significant change in 18 subjects. The y-axis indicates the number of OTUs showing significant change between the Pro(−) and Pro(+) samples in each subject (see Supplementary Tables S9 and S10). Open and closed bars indicate increased and decreased OTUs, respectively.

(APr11, 20, 22, and 37) did not have such OTUs and 8 subjects had only 1 OTU, whereas 4 subjects (APr01, 02, 31, and 39) had more than 4 OTUs showing significant change (Fig. 6), suggesting their uneven distribution in the 18 subjects. These data imply existence of the sensitive and less sensitive responders to probiotic action and if so, it would be interesting to investigate the relation between gut microbiota type and its response to probiotics.

In summary, we analysed changes of the gut microbiota composition of healthy adults fed with probiotics using the 454 pyrosequencing platform with the improved quantitative accuracy for evaluation of the overall bacterial composition. The present study using large datasets enabled us to more comprehensively and precisely evaluate the effect of probiotics on gut microbiota than the previous probiotic intervention researches in which the analysis exclusively focussed on only several limited bacterial species using conventional methods. Our data further support the high inter-subject variability and the high intra-subject stability that is the current common view for the feature of adult gut microbiota. A recent study of gut microbiota in twins demonstrated that probiotics had almost no effect on the community structure, but affected the gene expression of microbiota.[39] To more deeply understand the potential function of probiotics, the analysis of bacterial and host cell's transcriptome and intestinal metabolome is required.

Acknowledgements: We thank Dr Todd D. Taylor for critical reading of the manuscript, and K. Furuya, C. Shindo, H. Inaba, E. Iioka, Y. Takayama, E. Ohmori, M. Kiuchi, Y. Hattori (The University of Tokyo), and A. Nakano (Azabu University) for technical support.

Supplementary data: Supplementary Data are available at www.dnaresearch.oxfordjournals.org.

Funding

This work was supported in part by the global COE project of 'Genome Information Big Bang' from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan (to M.H. and K.O.), a research project grant from Azabu University to H.M. and by a grant from the Core Research for Evolutional Science and Technology (CREST) program of the Japan Science and Technology Agency (JST) to K.O.

References

1. Preidis, G.A. and Versalovic, J. 2009, Targeting the human microbiome with antibiotics, probiotics, and prebiotics: gastroenterology enters the metagenomics era, *Gastroenterology*, **136**, 2015–31.
2. Patel, R.M. and Lin, P.W. 2010, Developmental biology of gut-probiotic interaction, *Gut Microbes*, **1**, 186–95.
3. Gerritsen, J., Smidt, H., Rijkers, G.T. and de Vos, W.M. 2011, Intestinal microbiota in human health and disease: the impact of probiotics, *Genes Nutr.*, **6**, 209–40.
4. Sanders, M.E., Heimbach, J.T., Pot, B., et al. 2011, Health claims substantiation for probiotic and prebiotic products, *Gut Microbes*, **2**, 127–33.
5. Aureli, P., Capurso, L., Castellazzi, A.M., et al. 2011, Probiotics and health: an evidence-based review, *Pharmacol. Res.*, **63**, 366–76.
6. Rauch, M. and Lynch, S.V. 2012, The potential for probiotic manipulation of the gastrointestinal microbiome, *Curr. Opin. Biotechnol.*, **23**, 192–201.
7. Fujimura, K.E., Slusher, N.A., Cabana, M.D. and Lynch, S.V. 2010, Role of the gut microbiota in defining human health, *Expert Rev. Anti Infect. Ther.*, **8**, 435–54.
8. Deshpande, G.C., Rao, S.C., Keil, A.D. and Patole, S.K. 2011, Evidence-based guidelines for use of probiotics in preterm neonates, *BMC Med.*, **9**, 92.
9. Bron, P.A., van Baarlen, P. and Kleerebezem, M. 2011, Emerging molecular insights into the interaction between probiotics and the host intestinal mucosa, *Nat. Rev. Microbiol.*, **10**, 66–78.
10. Thomas, D.W., Greer, F.R., American Academy of Pediatrics Committee on Nutrition; American Academy of Pediatrics Section on Gastroenterology, Hepatology, and Nutrition. 2010, Probiotics and prebiotics in pediatrics, *Pediatrics*, **126**, 1217–31.
11. Indrio, F. and Neu, J.N. 2011, The intestinal microbiome of infants and the use of probiotics, *Curr. Opin. Pediatr.*, **23**, 145–50.
12. Saxelin, M., Tynkkynen, S., Mattila-Sandholm, T. and de Vos, W.M. 2005, Probiotic and other functional microbes: from markets to mechanisms, *Curr. Opin. Biotechnol.*, **16**, 204–11.
13. Nagpal, R., Kumar, A., Kumar, M., Behare, P.V., Jain, S. and Yadav, H. 2012, Probiotics, their health benefits and applications for developing healthier foods: a review, *FEMS Microbiol. Lett.*, **334**, 1–15.
14. Bisanz, J.E. and Reid, G. 2011, Unraveling how probiotic yogurt works, *Sci. Transl. Med.*, **3**, 106ps41.
15. Bron, P.A. and Kleerebezem, M. 2011, Engineering lactic acid bacteria for increased industrial functionality, *Bioeng. Bugs*, **2**, 80–7.
16. Kleerebezem, M. and Vaughan, E.E. 2009, Probiotic and gut lactobacilli and bifidobacteria: molecular approaches to study diversity and activity, *Annu. Rev. Microbiol.*, **63**, 269–90.
17. Ventura, M., O'Flaherty, S., Claesson, M.J., et al. 2009, Genome-scale analyses of health-promoting bacteria: probiogenomics, *Nat. Rev. Microbiol.*, **7**, 61–71.
18. Snydman, D.R. 2008, The safety of probiotics, *Clin. Infect. Dis.*, **46**, S104–11.
19. Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K. and Knight, R. 2012, Diversity, stability and resilience of the human gut microbiota, *Nature*, **489**, 220–30.
20. Clemente, J.C., Ursell, L.K., Parfrey, L.W. and Knight, R. 2012, The impact of the gut microbiota on human health: an integrative view, *Cell*, **148**, 1258–70.
21. Nicholson, J.K., Holmes, E., Kinross, J., et al. 2012, Host-gut microbiota metabolic interactions, *Science*, **336**, 1262–7.
22. Walter, J. and Ley, R. 2011, The human gut microbiome: ecology and recent evolutionary changes, *Annu. Rev. Microbiol.*, **65**, 411–29.
23. Hooper, L.V., Littman, D.R. and Macpherson, A.J. 2012, Interactions between the microbiota and the immune system, *Science*, **336**, 1268–73.
24. Tannock, G.W., Munro, K., Harmsen, H.J., Welling, G.W., Smart, J. and Gopal, P.K. 2000, Analysis of the fecal microflora of human subjects consuming a probiotic product containing *Lactobacillus rhamnosus* DR20, *Appl. Environ. Microbiol.*, **66**, 2578–88.
25. García-Albiach, R., Pozuelo de Felipe, M.J., Angulo, S., et al. 2008, Molecular analysis of yogurt containing *Lactobacillus delbrueckii* subsp. bulgaricus and *Streptococcus thermophilus* in human intestinal microbiota, *Am. J. Clin. Nutr.*, **87**, 91–6.
26. Alvaro, E., Andrieux, C., Rochet, V., et al. 2007, Composition and metabolism of the intestinal microbiota in consumers and non-consumers of yogurt, *Br. J. Nutr.*, **97**, 126–33.
27. Rochet, V., Rigottier-Gois, L., Levenez, F., et al. 2008, Modulation of *Lactobacillus casei* in ileal and fecal samples from healthy volunteers after consumption of a fermented milk containing *Lactobacillus casei* DN-114 001Rif, *Can. J. Microbiol.*, **54**, 660–7.
28. Rochet, V., Rigottier-Gois, L., Ledaire, A., et al. 2008, Survival of *Bifidobacterium animalis* DN-173 010 in the faecal microbiota after administration in lyophilized form or in fermented product – a randomised study in healthy adults, *J. Mol. Microbiol. Biotechnol.*, **14**, 128–36.
29. Ouwehand, A.C., Bergsma, N., Parhiala, R., et al. 2008, Bifidobacterium microbiota and parameters of immune function in elderly subjects, *FEMS Immunol. Med. Microbiol.*, **53**, 18–25.

30. Firmesse, O., Mogenet, A., Bresson, J.L., Corthier, G. and Furet, J.P. 2008, *Lactobacillus rhamnosus* R11 consumed in a food supplement survived human digestive transit without modifying microbiota equilibrium as assessed by real-time polymerase chain reaction, *J. Mol. Microbiol. Biotechnol.*, **14**, 90–9.
31. Lahtinen, S.J., Tammela, L., Korpela, J., et al. 2009, Probiotics modulate the Bifidobacterium microbiota of elderly nursing home residents, *Age (Dordr)*, **31**, 59–66.
32. Savard, P., Lamarche, B., Paradis, M.E., Thiboutot, H., Laurin, É. and Roy, D. 2011, Impact of *Bifidobacterium animalis* subsp. lactis BB-12 and, *Lactobacillus acidophilus* LA-5-containing yoghurt, on fecal bacterial counts of healthy adults, *Int. J. Food Microbiol.*, **149**, 50–7.
33. Yamano, T., Iino, H., Takada, M., Blum, S., Rochat, F. and Fukushima, Y. 2006, Improvement of the human intestinal flora by ingestion of the probiotic strain *Lactobacillus johnsonii* La1, *Br. J. Nutr.*, **95**, 303–12.
34. Engelbrektson, A.L., Korzenik, J.R. and Sanders, M.E. 2006, Analysis of treatment effects on the microbial ecology of the human intestine, *FEMS Microbiol. Ecol.*, **57**, 239–50.
35. Marzotto, M., Maffeis, C., Paternoster, T., et al. 2006, *Lactobacillus paracasei* A survives gastrointestinal passage and affects the fecal microbiota of healthy infants, *Res. Microbiol.*, **157**, 857–66.
36. Cox, M.J., Huang, Y.J., Fujimura, K.E., et al. 2010, *Lactobacillus casei* abundance is associated with profound shifts in the infant gut microbiome, *PLoS One*, **5**, e8745.
37. Culligan, E.P., Hill, C. and Sleator, R.D. 2009, Probiotics and gastrointestinal disease: successes, problems and future prospects, *Gut Pathog.*, **1**, 19.
38. Gareau, M.G., Sherman, P.M. and Walker, W.A. 2010, Probiotics and the gut microbiota in intestinal health and disease, *Nat. Rev. Gastroenterol. Hepatol.*, **7**, 503–14.
39. McNulty, N.P., Yatsunenko, T., Hsiao, A., et al. 2011, The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins, *Sci. Transl. Med.*, **3**, 106ra106.
40. Metzker, M.L. 2010, Sequencing technologies—the next generation, *Nat. Rev. Genet.*, **11**, 31–46.
41. Qin, J., Li, R., Raes, J., et al. 2010, A human gut microbial gene catalogue established by metagenomic sequencing, *Nature*, **464**, 59–65.
42. Huse, S.M., Dethlefsen, L., Huber, J.A., Mark, WD., Relman, D.A. and Sogin, M.L. 2008, Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing, *PLoS Genet.*, **4**, e1000255.
43. Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L. and Welch, D.M. 2007, Accuracy and quality of massively parallel DNA pyrosequencing, *Genome Biol.*, **8**, R143.
44. Hamady, M., Lozupone, C. and Knight, R. 2010, Fast Unifrac: facilitating high-throughput phylogenetic analysis of microbial communities including analysis of pyrosequencing and PhyloChip data, *ISME J.*, **4**, 17–27.
45. Droege, M. and Hill, B. 2008, The Genome Sequencer FLX System—longer reads, more applications, straight forward bioinformatics and more complete data sets, *J. Biotechnol.*, **136**, 3–10.
46. Andersson, A.F., Lindberg, M., Jakobsson, H., Bäckhed, F., Nyrén, P. and Engstrand, L. 2008, Comparative analysis of human gut microbiota by barcoded pyrosequencing, *PLoS One*, **3**, e2836.
47. Kuczynski, J., Lauber, C.L., Walters, W.A., et al. 2011, Experimental and analytical tools for studying the human microbiome, *Nat. Rev. Genet.*, **13**, 47–58.
48. Hamady, M., Walker, J.J., Harris, J.K., Gold, N.J. and Knight, R. 2008, Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex, *Nat. Methods*, **5**, 235–7.
49. Morita, H., Kuwahara, T., Ohshima, K., et al. 2007, An improved isolation method for metagenomic analysis of the microbial flora of the human intestine, *Microbes Environ.*, **22**, 214–22.
50. Claesson, M.J., Wang, Q., O'Sullivan, O., et al. 2010, Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions, *Nucleic Acids Res.*, **38**, e200.
51. Zhou, H.W., Li, D.F., Tam, N.F., et al. 2011, BIPES, a cost-effective high-throughput method for assessing microbial diversity, *ISME J.*, **5**, 741–9.
52. Hattori, M. and Taylor, T.D. 2009, The human intestinal microbiome: a new frontier of human biology, *DNA Res.*, **16**, 1–12.
53. Frank, J.A., Reich, C.I., Sharma, S., Weisbaum, J.S., Wilson, B.A. and Olsen, G.J. 2008, Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes, *Appl. Environ. Microbiol.*, **74**, 2461–70.
54. Hill, J.E., Fernando, W.M., Zello, G.A., Tyler, R.T., Dahl, W.J. and Van Kessel, A.G. 2010, Improvement of the representation of bifidobacteria in fecal microbiota metagenomic libraries by application of the cpn60 universal primer cocktail, *Appl. Environ. Microbiol.*, **76**, 4550–2.
55. Palmer, C. 2007, Development of the human infant intestinal microbiota, *PLoS Biol.*, **5**, 1556–73.
56. Haas, B.J., Gevers, D., Earl, A.M., et al. 2011, Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons, *Genome Res.*, **21**, 494–504.
57. Schloss, P.D., Gevers, D. and Westcott, S.L. 2011, Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies, *PLoS One*, **6**, e27310.
58. Gilles, A., Meglécz, E., Pech, N., Ferreira, S., Malausa, T. and Martin, J.F. 2011, Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing, *BMC Genomics*, **12**, 245.
59. Quince, C., Lanzén, A., Curtis, T.P., et al. 2009, Accurate determination of microbial diversity from 454 pyrosequencing data, *Nat. Methods*, **6**, 639–41.
60. Diaz, P.I., Dupuy, A.K., Abusleme, L., et al. 2012, Using high throughput sequencing to explore the biodiversity in oral bacterial communities, *Mol. Oral Microbiol.*, **27**, 182–201.
61. Schloss, P.D. and Handelsman, J. 2005, Introducing DOTUR, a computer program for defining operational