The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota

Ilias Lagkouvardos, Rüdiger Pukall, Birte Abt, Bärbel U. Foesel, Jan P. Meier-Kolthoff, Neeraj Kumar, Anne Bresciani, Inés Martínez, Sarah Just, Caroline Ziegler, Sandrine Brugiroux, Debora Garzetti, Mareike Wenning, Thi P. N. Bui, Jun Wang, Floor Hugenholtz, Caroline M. Plugge, Daniel A. Peterson, Mathias W. Hornef, John F. Baines, Hauke Smidt, Jens Walter, Karsten Kristiansen, Henrik B. Nielsen, Dirk Haller, Jörg Overmann, Bärbel Stecher, Thomas Clavel

OPEN

The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota

Ilias Lagkouvardos[1], Rüdiger Pukall[2], Birte Abt[2,3], Bärbel U. Foesel[2,3], Jan P. Meier-Kolthoff[2], Neeraj Kumar[1], Anne Bresciani[4], Inés Martínez[5], Sarah Just[1], Caroline Ziegler[1], Sandrine Brugiroux[6], Debora Garzetti[6], Mareike Wenning[7], Thi P. N. Bui[8], Jun Wang[9], Floor Hugenholtz[8], Caroline M. Plugge[8], Daniel A. Peterson[10], Mathias W. Hornef[11], John F. Baines[9,12], Hauke Smidt[8], Jens Walter[5], Karsten Kristiansen[13], Henrik B. Nielsen[4], Dirk Haller[1,14], Jörg Overmann[2,3], Bärbel Stecher[3,6] and Thomas Clavel[1]*

Intestinal bacteria influence mammalian physiology, but many types of bacteria are still uncharacterized. Moreover, reference strains of mouse gut bacteria are not easily available, although mouse models are extensively used in medical research. These are major limitations for the investigation of intestinal microbiomes and their interactions with diet and host. It is thus important to study in detail the diversity and functions of gut microbiota members, including those colonizing the mouse intestine. To address these issues, we aimed at establishing the Mouse Intestinal Bacterial Collection (miBC), a public repository of bacterial strains and associated genomes from the mouse gut, and studied host-specificity of colonization and sequence-based relevance of the resource. The collection includes several strains representing novel species, genera and even one family. Genomic analyses showed that certain species are specific to the mouse intestine and that a minimal consortium of 18 strains covered 50–75% of the known functional potential of metagenomes. The present work will sustain future research on microbiota–host interactions in health and disease, as it will facilitate targeted colonization and molecular studies. The resource is available at www.dsmz.de/miBC.

Molecular techniques have revolutionized the investigation of microbial ecosystems, but, as a consequence, culture-based approaches have been neglected. Interpretation of the gigantic amount of data obtained by high-throughput sequencing requires key isolates for precise taxonomic and functional classification. Large-scale cultivation of bacteria has primarily been focused on the human gut[1,2]. Yet, mouse models allow functional insights to be gained into host–microbe interactions while testing the causal role of microbiota in health and disease. However, the mouse gut microbiota has been poorly characterized, and bacterial isolates with reference genomes from mice are scarce. Improving the availability of representative isolates from the mouse gut microbiome is essential for future studies using gnotobiotic mice, particularly because the first reference gene catalogue of the mouse gut microbiota shows very limited overlap with human gut microbial gene diversity[3]. Furthermore, the host species from which bacteria originate influences colonization processes and the effects on host physiology[4–6]. Hence, concerted actions to archive and describe bacterial isolates in a host-specific fashion together with appropriate

genomic databases are urgently needed[7,8]. Pioneering work in the 1960s focused on the isolation of gut bacteria from the mouse intestine[9], but classification was limited to phenotypic features, and the majority of strains obtained in these and subsequent studies have been lost over the years. A very few species remain available (for example, members of the so-called Altered Schaedler Flora, ASF), but only under restricted conditions.

In the present work, a comprehensive collection of mouse gut bacteria was established and made available to the scientific community, including corresponding draft genomes. Important questions in the field were also addressed: 'Is species diversity host-specific?' 'To what extent do cultured bacteria cover the ecosystem diversity as assessed through molecular methods?'

Results

A comprehensive collection of bacteria from the mouse intestine. The primary aim of the work was to establish the Mouse Intestinal Bacterial Collection (miBC) and make it available to the scientific community.

[1]ZIEL Institute for Food and Health, Core Facility NGS/Microbiome, Technical University of Munich, 85354 Freising, Germany. [2]Leibniz-Institute DSMZ – German Collection of Microorganisms and Cell Cultures, 38124 Braunschweig, Germany. [3]German Center for Infection Research (DZIF), Partner site Hannover-Braunschweig and Munich, 38124 Braunschweig, Germany. [4]Department of Systems Biology, Center for Biological Sequence Analysis, Technical University of Denmark, DK2800 Kongens Lyngby, Denmark. [5]University of Alberta, Edmonton, Alberta T6G 2R3, Canada. [6]Max von Pettenkofer-Institute, Ludwig-Maximilians-Universität München, 80336 Munich, Germany. [7]Chair for Microbial Ecology, Technical University of Munich, 85354 Freising, Germany. [8]Laboratory of Microbiology, Wageningen University, 6708 WE Wageningen, The Netherlands. [9]Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany. [10]Immunology laboratory, Johns Hopkins University School of Medicine, Baltimore 21287, USA. [11]Institute of Medical Microbiology, RWTH University Hospital, D-52074 Aachen, Germany. [12]Institute for Experimental Medicine, Christian-Albrechts-University of Kiel, 24118 Kiel, Germany. [13]Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark. [14]Chair of Nutrition and Immunology, Technical University of Munich, 85354 Freising, Germany. *e-mail: clavel.tom@gmail.com

A total of approximately 1,500 pure cultures were isolated from the intestine of mice, from which strains were selected on the basis of colony and cell morphology as well as 16S rRNA gene sequence to cover as much diversity as possible at the species level. In an effort to unify information scattered throughout the literature, we also added eight of our own previously published mouse-derived bacterial species[10–17] and four species published by others[18–21].

The diversity of miBC members based on taxonomic lineage is presented in Fig. 1. The collection contains 100 strains representing 76 different species from 26 families belonging to the phyla Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria and Verrucomicrobia (Supplementary Table 1). It is dominated by Firmicutes (74% of strains), reflecting the dominance of this phylum in the mouse intestine. All isolates were deposited at the German Collection of Microorganisms and Cell Cultures (DSMZ) to ensure long-term storage and availability. A specific list linked to the metadatabase BacDive[22] was created to allow rapid query of the collection by users (www.dsmz.de/miBC). Hence, miBC is a unique tool that can serve as a reference for cultivated fractions of the mouse gut microbiota.

Culturing effort brings novel bacterial diversity and functions to light. The second purpose of the project was to isolate and describe new members of the mouse gut microbiota.

Based on 16S rRNA gene sequencing, 15 strains were characterized by values with <97% identity to any species described so far, and draft genomes were generated. Whole proteome-based phylogenomic analysis showed that the isolates were distant from known relatives with available genomic information (Supplementary Fig. 1). Genome-derived 16S rRNA gene and *in silico* DNA:DNA hybridization (DDH) estimates (see Methods) were below species-level thresholds (98.1% 16S rRNA gene sequence identity or DDH value of 70%, respectively) (Supplementary Fig. 2, Supplementary Tables 2 and 3)[23]. This confirmed that the isolates represented novel bacterial taxa, including four species, ten genera and one family. Proposed names are indicated in quotation marks throughout the manuscript and detailed taxonomic descriptions are provided below. Microscopic pictures are shown in Supplementary Fig. 3, and enzymatic profiles in Supplementary Table 4. Detailed characterization revealed that 'Flintibacter butyricus' can grow on the amino acids glutamine and glutamate (but not leucine, asparagine, lysine, arginine and aspartic acid), producing up to 7 mM butyrate from 20 mM substrate. Moreover, miBC provides the first cultivable member of the 'Muribaculaceae' (family S24-7), which is known to be a dominant bacterial group in the mouse gut[6]. To confirm the latter observation and assess the occurrence of all novel bacteria within miBC, draft genomes were compared to metagenomic species (MGS)[24] from a comprehensive shotgun sequencing study of the mouse gut metagenome[3].

A BLAST search revealed that 5 of the 15 isolates had many strong matches with the mouse gene catalogue (Supplementary Fig. 4 and Supplementary Table 5). The other ten isolates had <200 gene hits (Supplementary Table 5), suggesting that they either represent subdominant populations in this cohort of mice, or are not detectable in the mouse gut using the most advanced molecular methods. The five novel isolates well represented in the mouse catalogue showed a high gene coverage (85–99% of genes at 95% sequence identity), allowing assessment of their prevalence in the faeces of 184 mice (Supplementary Fig. 4). Two species were found in the majority of mice: 'Flintibacter butyricus' (98% prevalence) and 'Enterorhabdus muris' (76%). Only 'Flintibacter butyricus' was present in all eight mouse strains from all six animal facilities (five commercial suppliers), suggesting a widespread colonization of the mouse gut by this amino-acid-degrading butyrate producer.

In summary, the large-scale culture-based work carried out to create miBC demonstrates the relevance of isolating and characterizing bacteria, as it allows the description of novel species that can be dominant and/or functionally important.

miBC captures the majority of currently cultured bacteria in the mouse gut. The third aim of the work was to test the relevance of miBC by assessing the coverage of high-throughput 16S rRNA gene amplicon inventories. Therefore, we used caecal samples from 93 mice housed in various facilities in Europe and America (Supplementary Tables 6 and 7).

Phylogenetic profiles of dominant bacterial populations clustered according to animal facilities (Fig. 2a), characterized by significant differences in taxa richness, in the relative abundance of Bacteroidetes and Firmicutes (Fig. 2b) and by the presence of indicator species (Supplementary Table 8). This emphasizes the rationale of working with a multi-centre data set to test the representativeness of the collection. As we isolated the first cultured member of family S24-7 and assessed its host-specificity (next section), we investigated the distribution of reads classified as S24-7 members. They were detected in 86 of 93 mice (Supplementary Fig. 5), indicating a widespread distribution, in contrast to the MGS data presented above (Supplementary Fig. 4 and Supplementary Table 5).

Local BLAST analysis of the 772 detected operational taxonomic units (OTUs) against miBC revealed 46 hits at the species level (≥97% identity), and 105 at the genus level (≥95%), representing 23.5% of all sequence reads (Fig. 2c). Despite the relatively stringent quality-check and filtering used for generating the final 16S rRNA gene short reads data set and corresponding OTU table, these percentages may underestimate the cultured fractions due to the presence of spurious OTUs in the high-throughput sequence data. Looking at the most common and dominant molecular species (referred to as core OTUs from here on), the 42 OTUs detected in ≥50% of the mice at a mean relative abundance of ≥0.5% represented 56.8% of the total reads. Nine of these OTUs matched miBC strains at the genus level (Fig. 2d), accounting for 30% of the core reads. When searching for additional cultured bacteria (from environments other than the mouse gut) that represented any OTUs with no match to miBC, 21 and 48 were identified at the species and genus levels, respectively, accounting for only 7.5% extra sequence abundance (Fig. 2c). Of the 33 core OTUs with no match to miBC, six matched other cultured bacteria (one at the species and five at the genus level; Fig. 2d), accounting for 5.7% of the core reads.

These data indicate that the created bacterial collection already covers important dominant lineages, although the majority of mouse gut bacteria remain to be cultured.

Gut bacterial species diversity is host-specific. Published work shows that the origin of bacteria influences colonization processes and host physiology[4,5], but there are only a few reports on the host specificity of gut bacterial taxa in human versus mouse[3,25]. The fourth aim of the project was to investigate the presence of miBC species in different ecosystems, especially gut environments, based on the entire pool of samples in the Sequence Read Archive (SRA) using IMNGS (www.imngs.org). Raw output data are provided in Supplementary Table 9.

We first looked at the coverage of main SRA sample categories (environment, $n = 17,884$; host gut, $n = 21,755$; other host habitats, $n = 15,414$) based on the entire diversity in miBC. The percentage of data sets where all species in the collection represented >1% total reads was highest in host gut-derived samples (Fig. 3a), supporting the common sense that bacterial occurrence depends on major environmental features in different ecosystems. The resolution of analysis was then refined to mouse and human gut samples. Almost one-third of the 6,001 mouse samples (28.6%) were
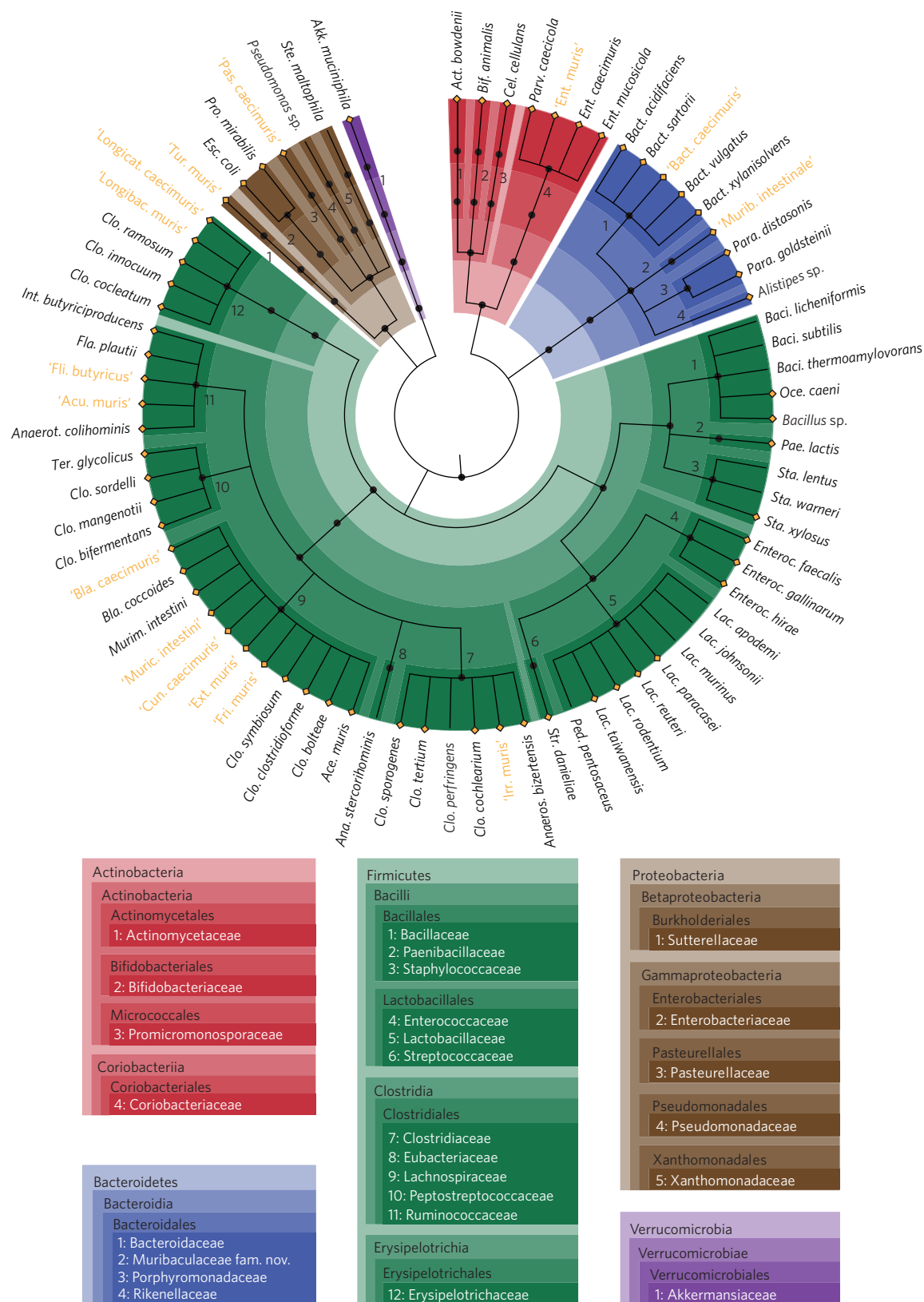
Figure 1 | Diversity of the miBC. The cladogram illustrates the taxonomic classification of all 76 species in miBC down to the family level. The cladogram is colour-coded according to phyla, as shown in the boxes. Novel taxa described in the present study with their candidate names appear in gold. Species for which draft genomes were generated are indicated with diamonds before species names.

characterized by a cumulative abundance coverage >50% (that is, the sum of sequences matching that of miBC species at >97% identity was more than 50% total reads per sample). This was nearly twice as much as for human samples (16.6% of 11,705) (Fig. 3b). These human samples seem to contain relatively high proportions of short 16S rRNA gene amplicons similar to 16S rRNA genes of
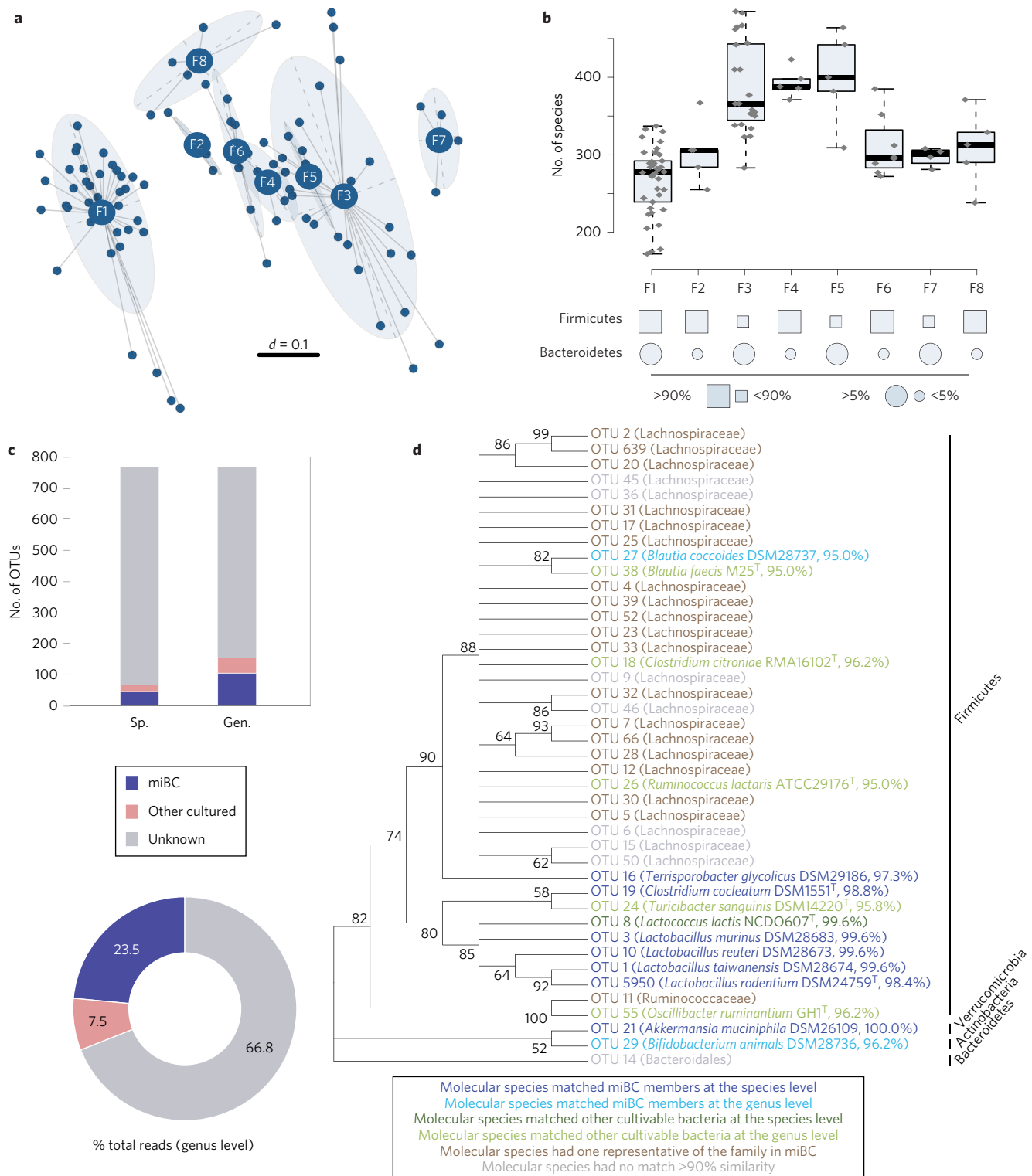
Figure 2 | Representativeness of miBC based on 16S rRNA gene sequence analysis. Mouse caecal samples (n = 93) were analysed by high-throughput sequencing of 16S rRNA gene amplicons (V5/6 region, ~250 bp). **a**, Multidimensional plot of generalized UniFrac distances. Animal facilities (F1–F8) show the different phylogenetic makeup of dominant bacterial communities. **b**, Taxa richness and relative abundance categories for Bacteroidetes and Firmicutes across facilities. Taxa richness is shown as boxplots (with median, interquartile range (IQR) and whiskers extending to the last data points within +/− 1.5× IQR). Individual values appear as grey dots. **c**, Coverage of the 772 molecular species by members of the mouse collection in terms of OTU number (histograms) and percentage of reads (pie chart). **d**, Phylogenetic tree of core OTUs and their match to miBC members. F, animal facility; Gen., genus level (95% sequence identity); OTU, operational taxonomic unit; Sp., species level (97% sequence identity).

miBC strains (at 97% identity) and probably do not correspond to representative ecosystems from the human gut, but rather to specific diet- or health-related conditions where a limited number of taxa dominate. Altogether, these data speak in favour of better coverage of 16S rRNA gene data sets from the mouse than human intestine using strains that were isolated from mice to query the database.
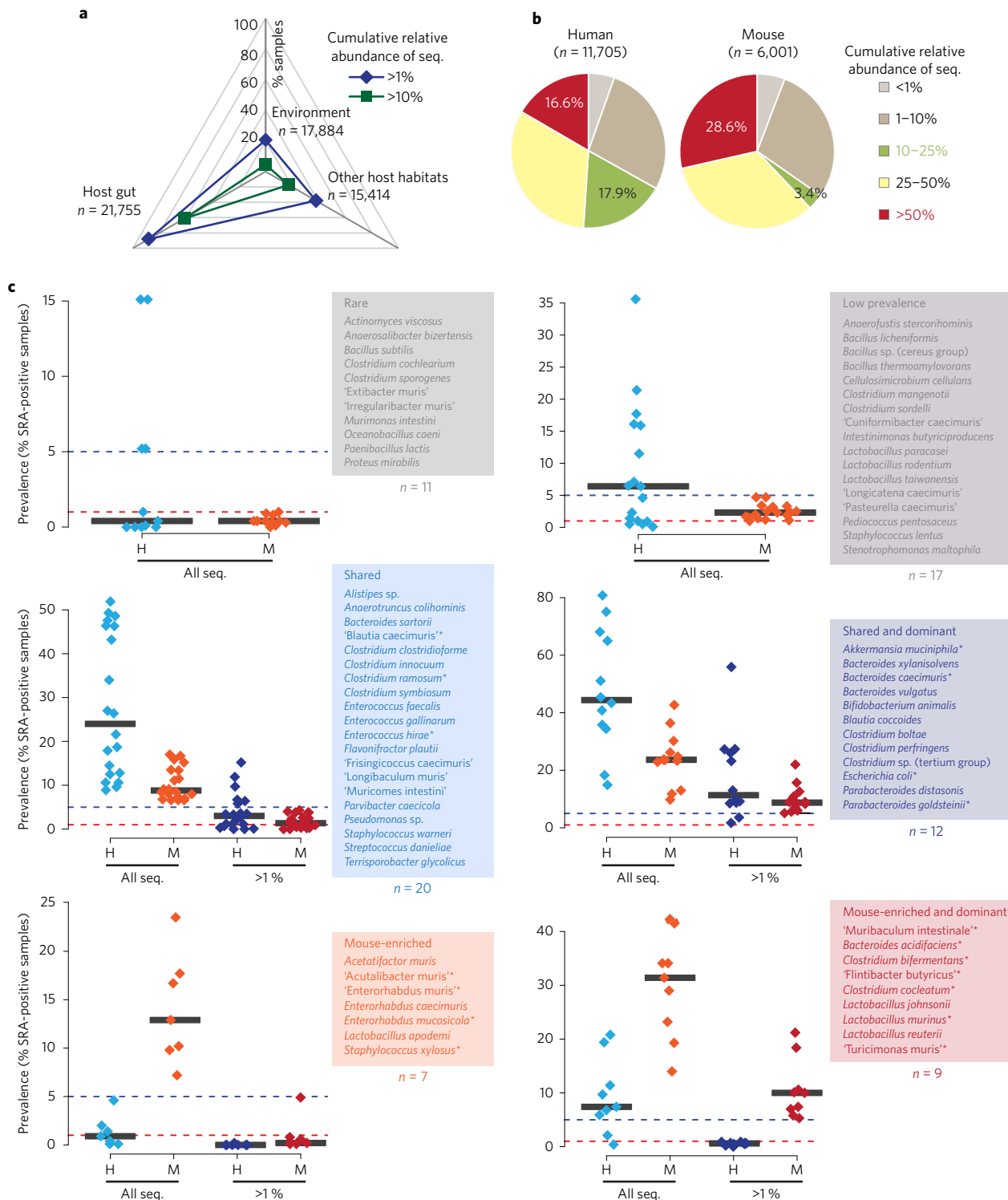
Figure 3 | Metasearch of 16S rRNA gene short reads for prevalence and abundance studies of miBC members. Species in the collection were used to query all 16S rRNA gene amplicon data sets stored in the SRA using IMNGS (www.imngs.org), as described in detail in the text. **a**, Percentage of samples from main ecological categories for which cumulative abundance of miBC species covered >1 or 10% total reads. **b**, As in **a**, including human and mouse gut samples only. Additional cumulative abundance categories are shown. **c**, Species-specific prevalence analysis in human and mouse gut samples. Black horizontal bars indicate median values. The 76 species in the collection were categorized as described in the Methods. The 18 bacteria selected for creation of the Minimal Bacteriome (MIBAC-1) are marked with asterisks. These species were selected based on their prevalence and abundance using both the MGS- and universal 16S rRNA gene amplicon-based surveys described in the text and also according to their phylogenetic diversity. Quotation marks indicate novel bacterial names proposed in the present study.

Next, to determine whether certain species may be host-specific, we categorized the 76 miBC species according to their prevalence in human and mouse gut samples (for details see Methods). We identified 32 shared species (present in both human and mouse) and 16 that were characteristic for the mouse intestine (Fig. 3c). Interestingly, three previously published mouse-derived species

(*Enterorhabdus* spp., *Acetatifactor muris*)[10,13,17] and five novel taxa isolated in the present study ('Acutalibacter muris', 'Enterorhabdus muris', 'Turicimonas muris', 'Muribaculum intestinale', 'Flintibacter butyricus') were indeed found to be enriched in the mouse intestine using IMNGS. The remaining 28 species were either rare or characterized by low prevalence. Among the 16 mouse-enriched species, 9 were also present when considering a relative abundance of >1% total reads, indicating that these species are dominant in the mouse gut microbiota.

These findings suggest that colonization by certain bacterial species, or at least their occurrence in dominant communities as detected by sequencing, is host-specific. A selection of 18 shared or mouse-enriched species for establishment of a minimal bacteriome (MIBAC-1) to model the mouse gut ecosystem are shown with asterisks in Fig. 3c. This bacterial consortium was further investigated using genomic approaches.

Genomic novelty and representativeness of cultured mouse gut bacteria. To evaluate the relevance of miBC and the minimal consortium of 18 species (MIBAC-1) at the metagenomic level, draft genomes of the 76 miBC species were compared to published shotgun metagenomes from the mouse ($n = 15$) and human ($n = 21$) gut (Supplementary Table 10) and to two metagenomes generated in-house (caecum and faeces of one specific-pathogen free (SPF) wild-type mouse). First, a local database of miBC genomes and the metagenomes in a presence/absence binary code (1/0) of protein families (PFAM) was created. Deterministic incremental selection of best-fitted genomes showed that 12 to 15 species sufficed to cover ~80% of known functional diversity in metagenomes (Fig. 4a). It also showed that cumulative information in miBC genomes plateaued at approximately 90% coverage and that coverage of mouse vs. human metagenomes was higher ($P = 1.66e-10$, Kolmogorov–Smirnov test).

Next, we investigated coverage of the mouse metagenomes ($n = 17$) by comparing miBC and the minimal bacteriome MIBAC-1 with consortia from the literature, that is, the ASF[26] and the Simplified Human Intestinal Microbiota (SIHUMI)[27]. Coverage analysis included not only the PFAM approach, but also protein sequence coverage by BLAST analysis (that is, independent of functional annotation) for determination of both homology ($\geq$80% length coverage, $E < 10^{-5}$) and close sequence match ($\geq$80% both coverage and sequence similarity). As expected, both PFAM and sequence coverages were highest for the entire collection, and coverage decreased with the number of species included in minimal consortia (18 in MIBAC-1 versus 8 in each of ASF and SIHUMI) (Fig. 4b). To test the relevance of MIBAC-1 (obtained by data-driven selection), we analysed 10 selections of 18 strains picked randomly from miBC (MIBAC-R1 to 10) (Supplementary Table 11). The designed minimal bacteriome MIBAC-1 was characterized by higher coverage than the mean of random sets for both sequence homologues and close matches but not for PFAM (Fig. 4c). This strengthens the rationale for having selected these strains and emphasizes the known shallow resolution of analysis when looking at broad functional categories and the more detailed specificity when looking, for instance, at genes[3].

Finally, PFAM diversity was investigated by heatmap and dendrogram analysis. The mouse collection and the metagenomes formed a cluster that was distant from minimal bacteriomes (Supplementary Fig. 6). Moreover, although ~10% of PFAM-derived information captured by shotgun sequencing was not contained in miBC (Fig. 4a), 437 PFAM detected in the collection and minimal bacteriomes (representing 2.7% of the total 16,231 PFAM in all data sets) were absent from metagenomes, 129 of which were unique to miBC (Fig. 4c). These PFAM (Supplementary Table 12) are thus not captured by metagenomics, probably because they originate from taxa that are subdominant or lost during sample preparation.

Discussion

The extensive use of mouse models is in stark contrast to the lack of reference bacterial strains from the mouse intestine. In the present work, we have created a publicly available collection of bacteria from the mouse gut and provide important insight into host-specific bacterial diversity and functions.

Besides delivering an array of mouse gut bacteria to the research community, miBC contains representative strains of novel taxa, which represent 15% of the collection (~1.0% of original isolates tested). Thus, a substantial number of novel bacteria can still be discovered even via classical bacteriological methods, some of these bacteria being apparently of particular functional relevance. For example, mice are colonized by a clade of bacteria from the order Bacteroidales, originally referred to as MIB (mouse intestinal bacteria)[28] and nowadays misleadingly classified as family S24-7 or Porphyromonadaceae spp. depending on the database. The present work is the first to report the isolation of a cultured member of this important family in the mouse gut ('Muribaculum intestinale'), opening avenues for functional studies. The collection also includes one strain of each species *Intestinimonas butyriciproducens* and 'Flintibacter butyricus' able to produce butyrate not only from sugars but also amino acids. A human isolate of *I. butyriciproducens* does so via a metabolic pathway so far unknown in gut environments[29], highlighting the value of isolating bacteria to unravel functional novelty.

Investigating the mechanisms that underlie the specificity of bacterial colonization in the gut is important to understand the population dynamics associated with resilience, resistance to invaders, or survival of exogenous strains for therapeutic purposes. Occurrence analysis of the 76 miBC species following a large-scale 16S rRNA gene amplicon approach revealed that 20% were enriched in mice, including members of all major phyla. These results add precision to recent surveys on taxonomic comparison between human and mouse metagenomes, proposing that lactobacilli and several genera within the Firmicutes (including *Coprobacillus* and *Turicibacter* spp. within Erysipelotrichaceae, *Anaerotruncus* within Ruminococcaceae, *Marvinbryantia* within Lachnospiraceae, and *Pseudoflavonifractor* within Clostridiales) are characteristic of mice[3,25].

We found that members of the proposed novel family 'Muribaculaceae' are specific to the mouse gut, even though their presence in humans has been reported by others using quantitative polymerase chain reaction (qPCR)[30]. This discrepancy may be linked to primer specificity, and additional work is needed to assess the occurrence of these bacteria in other mammalian species, including pigs[31]. Recent data comparing human and mouse microbiota reported mouse-specific colonization by this family[6]. Via colonization of germfree mice with microbial communities from mammals, insects and environmental samples, the authors of ref. 6 demonstrated that colonization success (maintained taxa richness and similarity to input samples) was highest for gut samples, indicating ecosystem-selective pressure on microbial communities[6]. Competitive experiments with human versus mouse microbiota revealed transient colonization by human indicator species, but the gut ecosystem was dominated by mouse species after 14 days (>99% of the community), especially members of family S24-7 (proposed to be renamed 'Muribaculaceae'). These observations are in accordance with our findings supporting the notion of host-specific gut colonization.

In contrast to common statements in the literature on the vast majority of gut bacteria not being cultured, work by Goodman and colleagues[32] reported that approximately 56% of 16S rRNA gene sequences detected in human faecal samples belong to readily cultured species. A recent review also reports that the majority of most dominant species detected by molecular analysis
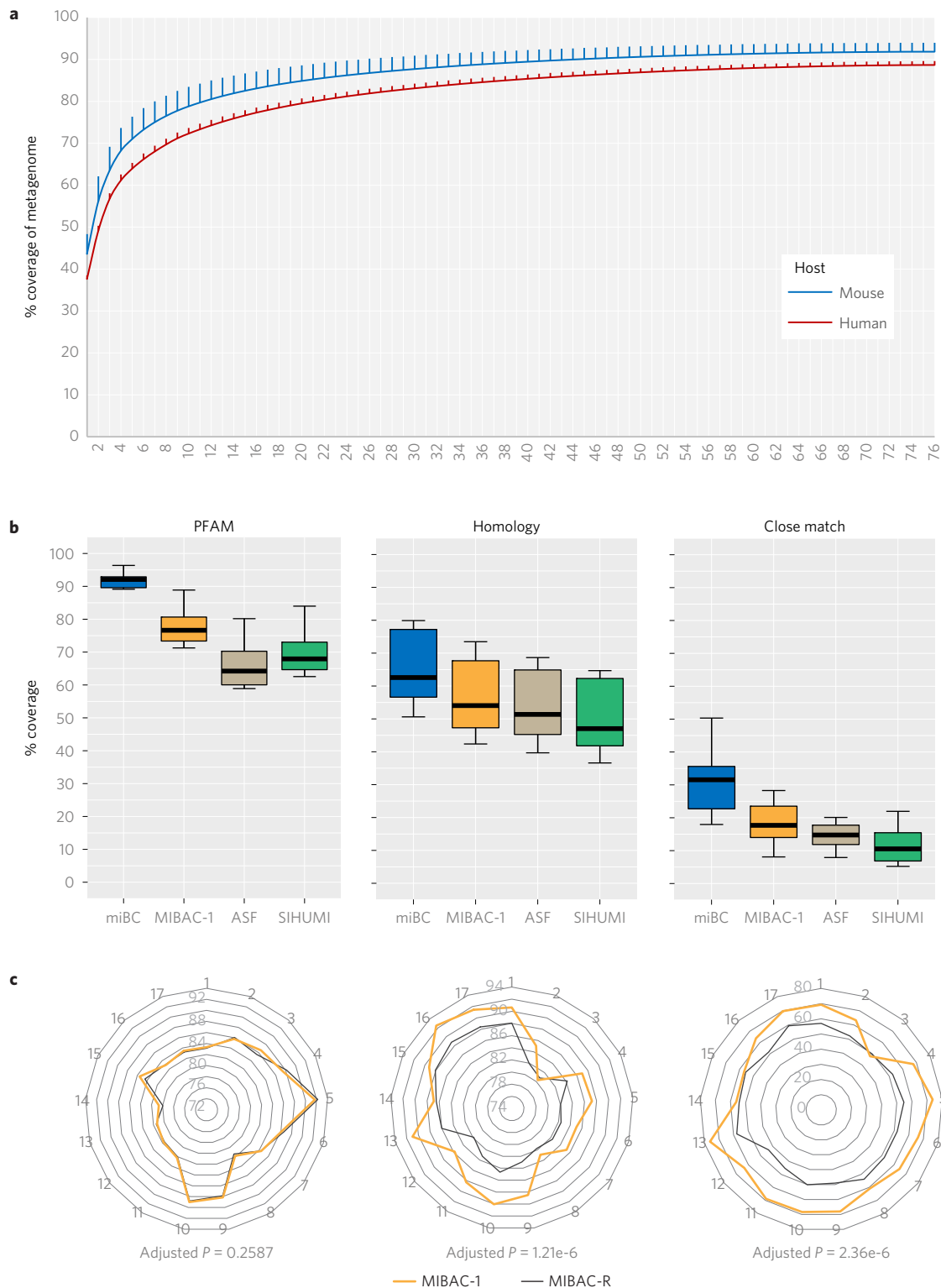
Figure 4 | Metagenomic coverage. Coverage was calculated following three complementary approaches: at the level of protein families (PFAM), homologous proteins (≥80% coverage, $E < 10^{-5}$ after BLASTp search) and close match (≥80% sequence similarity). **a**, PFAM-based coverage of faecal metagenomes from mouse (blue; $n = 17$) or human (red; $n = 21$). Data are means ± s.d. Genomes from the mouse collection (x axis) were ranked according to decreasing contribution to the pool of PFAM in each metagenome. **b**, Coverages of mouse metagenomes ($n = 17$) obtained by each of the three approaches already mentioned for the entire collection (miBC), the minimal consortium described in the present study (MIBAC-1), the Altered Schaedler Flora (ASF) and the Simplified Human Intestinal Microbiota (SIHUMI). Boxplots show the median, IQR and whiskers extending to the last data points within +/− 1.5× IQR. Values for a given analysis type were significantly different between all categories (t-test with Benjamini–Hochberg correction). **c**, Radar plots showing the coverage of 17 mouse metagenomes (outer ring) by MIBAC-1 and by 10 random sets of 18 strains from miBC (MIBAC-R) at each analysis level (as in **b**). Random sets are shown as mean values. Coverages are shown as percentages (grey numbers in the grid) of PFAM, sequence homologues or close matches covered by the entire collection (miBC values in **b**). P values were obtained by t-test.

have cultured representatives[33]. Analysis of high-throughput 16S rRNA gene amplicons from a mouse cohort covering different facilities revealed a set of 42 core bacterial OTUs in the mouse gut, which was covered up to 21% (OTU diversity; 9 of 42) and 30% (number of reads) by miBC at the genus level. This is the best available to date, but still lower than the reference values for humans[32]. A very recent study reported that the majority of human faecal bacteria can actually be cultured[1]. This highlights the need for further work to isolate and characterize novel mouse gut bacteria, the 33 core OTUs without match to miBC representing 'most wanted taxa'. Moreover, these data support the existence of substantial differences in gut bacterial diversity and composition between mouse facilities, which can have dramatic consequences on host phenotypes[34]. This emphasizes the need for better characterization and standardization of complex gut microbial communities based on the knowledge and bacterial strains gained in the present study and by others since pioneering work in the 1960s[35,36].

The representation of cultured bacteria within miBC was also assessed at the metagenomic level. The entire collection covered 60–90% of the functional potential in mouse faecal metagenomes. This high coverage reflects the limitation of the analysis to known functions or homologous protein sequences. Narrowing measurements to the 18-species consortium MIBAC-1, which can be used as a proxy for native mouse gut ecosystems, revealed coverages of 55–75% (PFAM and sequence homologues) and 20% (close sequence matches). Minimal consortia of bacteria already exist in the literature, but strains were selected based on educated guesses or for specific purposes and either originated from the human intestinal tract or are not easily available[26,27,37]. In contrast, the minimal bacteriome presented here contains strains that originate from the mouse intestine, is publicly available, was selected on the basis of comprehensive sequence-based approaches, and was overall characterized by a higher coverage of mouse faecal metagenomes. Hence, it represents important progress towards the standardization of mouse models, as emphasized by others[35,38]. However, additional work is required for establishment of the consortium *in vivo*. Demonstrating stable colonization, which can be influenced by many ecological factors, is indeed very important and challenging, but is beyond the scope of the present study.

In conclusion, the significance of our work to the field and the broader community is manifold: (1) it provides the basis for genetic studies that will eventually improve the resolution of meta-omics analyses of the mouse gut microbiome; (2) unrestricted access to miBC strains is a major advance for the research community, especially for functional studies on cause-and-effect relationships and for standardizing experiments among laboratories; and (3) the collection includes mouse-enriched taxa and strains with specific metabolic functions such as butyrate production, allowing ecological studies and assessment of the impact of specific strains on host physiology and disease onset. We acknowledge the fact that miBC is not an exhaustive collection; important bacterial taxa are still missing (for example, segmented filamentous bacteria or species of the phyla TM7 and Deferribacteres) and other microorganisms such as fungi and archaea, as well as viruses, are also important to investigate[39,40]. Moreover, key metabolic functions such as the conversion of bile acids require more detailed investigation. Also, strain diversity is most probably very important in gut environments[41], but the collection offers resolution at the strain level for only a few species (especially lactobacilli), as the aim was to cover as much diversity as possible at the species level. Nevertheless, recent metagenomic work[41] also emphasizes the necessity to obtain reference genomes and thus the benefit of isolating and describing bacteria, as done extensively in the present work. Hence, knowledge of the mouse gut bacterial diversity made available via miBC can be viewed as a foundation that now requires effort by the entire community of gut microbiome researchers for further development.

Methods

Mouse samples for cultivation. The use of mice was approved by the local authorities in charge (animal welfare authorization 32–568, Freising District Office; Regierungspräsidium Freiburg T05-28). Laboratory mice were housed in conventional or specific pathogen-free facilities at the WZW School of Life Sciences (TU München), the Institute for Medical Microbiology and Hygiene (Universitätsklinikum Freiburg) or the Rodent Center (ETH Zurich). Samples from mice captured in the wild were obtained as described previously[42]. Starting materials for bacteria isolation included fresh faeces collected from living mice, mucosal samples, as well as small intestinal, caecal or colonic content collected from mice that had been euthanized by CO_2 inhalation or neck dislocation. The origins of bacteria (those eventually included in the collection) in terms of mouse genotype and gut location are provided in Supplementary Table 1. The working area and dissection set were cleaned and mice were copiously sprayed with 80% (vol/vol) ethanol before dissection in order to avoid contamination. In some instances, mice were dissected inside the anaerobic workstation to prevent any contact of gut samples with oxygen.

Culture media. All quantities are per litre of medium.

A II: Brain–heart–infusion (BHI), 18.5 g; yeast extract, 5 g; trypticase soy broth, 15 g; K_2HPO_4, 2.5 g; hemin, 10 µg; glucose, 0.5 g; palladium chloride, 0.33 g; agar, 15 g. After autoclaving, add Na_2CO_3, 42 mg; cysteine, 50 mg; menadione, 5 µg; fetal calf serum (complement-inactivated), 3% (vol/vol). Adapted from Aranki and Freter[43].

BAC: WCA (Oxoid; CM0643; see manufacturer's instructions); Agar, 1.5 g; PdCl, 1 mg, dissolve first in HCl; phenosafranine, 2.5 mg; set pH to 7.0. After autoclaving (121 °C, 15 min), add kanamycin, 50 µg ml^{-1}; cysteine (0.05% wt/vol), dithiothreitol (DTT) (0.02%), sheep blood (5% vol/vol), 1 vial of G-N Anaerobe Supplement (Oxoid SR0108) per 475 ml medium.

BARN: 1× Sigma salt (M6030); yeast extract, 0.5 g; maltose, 1 g; agar, 15 g; $PdCl_2$ 2 mg; phenosafranin; set pH to 7.6. After autoclaving (121 °C, 15 min), add penicillin (25 µg ml^{-1}), vancomycin (20 µg ml^{-1}), kanamycin (50 µg ml^{-1}), cysteine (0.05% wt/vol), trace element and vitamin solutions (see DSMZ medium 104 PYG).

BHI: DSMZ medium 215c (supplemented or not with 15 g agar).

BIL: 1× Sigma salt (M6030); peptone, 2 g; yeast extract, 200 mg; agar, 1.5 g; PdCl (1 mg; dissolve first in HCl); phenosafranine (2.5 mg). After autoclaving (121 °C, 15 min), add Oxgall (1.5% vol/vol; Fluka ref. B3883), taurine (200 µM), cysteine (0.05% wt/vol), TUDCA (100 µg l^{-1}; Santa Cruz Biotechnol, ref. sc296449).

BLS: Blood agar with 10 mg l^{-1} colistin sulfate and 5 mg l^{-1} oxolinic acid.

COR: 1× Sigma salt (M6030); Tween 80 (1 ml); pig mucin type III (Sigma) (250 mg; Sigma M1778); agar (1.5% wt/vol); PdCl (1 mg; dissolve first in HCl); phenosafranine (2.5 mg). After autoclaving (121 °C, 15 min), add arginine, glycine, histidine and lysine (each 2.5 g), bile salts (0.5% wt/vol; Fluka ref. 48305), cysteine (0.05%), DTT (0.02%), sheep blood (5% vol/vol). For COR2 agar, add ampicillin and neomycin (50 µg ml^{-1} each) after autoclaving (121 °C, 15 min).

ECSA: 17.0 g pancreatic digested casein; 3.0 g peptic digested casein; 5.0 g yeast extract; 10.0 g ox gall; 5.0 g sodium chloride; 1.0 g aesculin; 0.5 g ammonium iron (III) citrate; 0.25 g sozium azide; 1.0 g sodium citrate; 13.5 g agar.

mOs-SRB: 11.28 g M9 minimal salt medium (Sigma 6030); 20.0 g agar; 2.0 g Na_2SO_4 anhydr.; 75 mg yeast extract; 75 mg Li-lactate. After autoclaving, add cysteine (0.05% wt/vol), DTT (0.02%).

MT10: 1× Sigma salt (M6030); pig mucin type III (Sigma) (250 mg; Sigma M1778); $PdCl_2$ 2 mg; thioglycolic acid, 0.1 g; phenosafranin; 50 ml rumen fluid; agar, 15 g; set pH to 7. After autoclaving (121 °C, 15 min), add rifampicin (20 µg ml^{-1}), cysteine (0.05% wt/vol), 2-mercaptoethanol (50 µM), trace element and vitamin solutions (see DSMZ medium 104 PYG).

Postgate (after Postgate, Appl Microbiol 1963, 11:265): KH_2PO_4, 0.5 g; NH_4Cl, 1 g; Na_2SO_4, 2 g; $CaCl_2.6H_2O$, 0.1 g; $MgSO_4.7H_2O$, 1 g; Na-lactate, 3.5 g; yeast extract, 1 g; thioglycolic acid, 0.1 g; $FeSO_4.7H_2O$, 0.5 g; agar, 15 g; $NaHCO_3$, 2 g; $PdCl_2$ 2 mg; phenosafranin; set pH to 7.6. After autoclaving (121 °C, 15 min), add penicillin (25 µg ml^{-1}), trace element and vitamin solutions (see DSMZ medium 104 PYG).

PYF: Peptone-yeast extract broth with 4% (vol/vol) Fildes' digest (peptic digestive of horse blood)[19].

PYG: DSMZ medium 104; 15 g agar.

TA-BHI: Tannic acid-treated BHI agar, prepared as described in ref. 44.

TGB: Thioglycollate medium brewer modified (Becton, Dickinson and Company, cat. no. 5211716); 15 g agar; 0.5 ml l^{-1} thioglycolic acid (AppliChem; cat. no. A3533).

WCA: Wilkins–Chalgren Anaerobe (Oxoid, ref. CM0643).

WHC: WCA (oxoid) + 25 µg ml^{-1} penicillin (sodium salt).

WSB: WCA (Oxoid), 33 g; glucose, 4 g; hemin, 10 µg; L-cystine, 0.4 g; agar, 15 g. After autoclaving, add defibrinated sheep blood (Oxoid) 5% (vol/vol).

Bacterial strains isolation. Gut samples were re-suspended (1:10 wt/vol) in reduced buffered solutions or broth media (see composition above in the section 'Culture media'). Mucosal samples were prepared as described previously[10]. Tenfold dilution series of gut suspensions were plated on agar media (see composition above) and bacteria were allowed to grow for 2–30 days under aerobic conditions, or in an anaerobic chamber containing a mixture of hydrogen, nitrogen and carbon dioxide

(5:10:85) or hydrogen and nitrogen only (10:90). Single colonies were streaked at least three times onto fresh agar plates before transfer into broth media. Bacterial smears formed after growth of low dilutions were also re-streaked to obtain single colonies of low-abundant bacteria. Culture purity was ensured by observing colony morphology as well as cell morphology by light microscopy after Gram-staining. For identification and phylogenetic analysis of isolates, DNA was extracted from pure cultures and 16S rRNA genes were amplified and sequenced as described previously[15]. Electropherogram quality was checked and contigs were built using Bioedit[45]. Sequences were identified using EzTaxon[46]. The value of 97% 16S rRNA gene sequence identity, which is a widely used and generally accepted though rather conservative threshold, was chosen to delineate novel taxa in a consistent manner across bacterial phyla[23]. Further characterization of the strains is detailed in the next section. Routine media used for subculturing strains included reduced Wilkins Chalgren anaerobic (WCA) broth (Oxoid) supplemented with cysteine (0.05% wt/vol) and DTT (0.02%) as reducing agents. Cryo-stocks were prepared from freshly grown cultures by mixing bacterial suspensions 1:1 with filter-sterilized glycerol in culture medium (40% vol/vol) before freezing at −80 °C. Live and cryo-cultures of isolates were shipped to the German Collection of Microorganisms and Cell Cultures (DSMZ), where stocks for long-term storage were prepared and detailed taxonomic characterization was carried out. Strain information (including culture conditions) is available online at www.dsmz.de/miBC.

Strain characterization. Analyses included cell morphology by microscopy, enzymatic testing, cellular fatty acids, diaminopimelic acid, automated ribotyping and mass spectrometry analysis using a MALDI-biotyper (Bruker). Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) sample preparation was performed according to protocol 3 as described by Schumann and Maier[47]. Ribotyping was carried out using the automated Riboprinter microbial characterization system (Dupont, Qualicon). Sample preparation and analysis were performed according to the manufacturer's instructions and EcoRI restriction enzyme was used to generate the DNA fragments. Protocol 1, published by Schumann[48] was applied to screen Gram-positive bacterial cells for the presence of diaminopimelic acid. Two loops of wet biomass were added to 200 μl of 4 N HCl in 2 ml glass ampoules, which were heat-sealed and kept for 16 h at 100 °C. After cooling, hydrolysates were pressed through a charcoal bed using a pipette bulb and dried in a gentle air stream at 35 °C until all traces of acid were eliminated. Hydrolysates were dissolved in 100 μl distilled water and a volume of 2 μl was spotted on the baseline of cellulose thin-layer chromatography (TLC) plates (Merck, P/N 1.05577) using a glass capillary. For separation of the diamino acid, a solvent system was used according to ref. 49 (32 ml methanol, 4 ml pyridine, 7 ml distilled water and 1 ml 12 N HCl; all from Sigma-Aldrich). TLC plates were sprayed with ninhydrin reagent after development. Spots became visible after heating (100 °C, 5 min). Cellular fatty acids were determined from cells grown on blood agar plates or collected from pre-reduced liquid media. Fatty acids were analysed as methyl ester derivatives obtained from 10 mg dry cell material by saponification, methylation and extraction using the modifications by Kuykendall et al.[50] of the method by Miller[51]. The fatty acid methyl esters were separated and identified using the Microbial Identification System (MIDI) Sherlock version 6.1 (TSBA40 database). Enzymatic testing was done according to the manufacturer's instructions (Biomérieux). Cell morphology was examined using phase-contrast microscopy (Zeiss Axioscope A.1, ×100 Plan-Neofluar oil-immersion objective, Ph3; Zeiss Axiocam MRc; Axiovision software). Slides were coated with a layer of highly purified agar (2%).

High-throughput 16S ribosomal RNA gene analysis. For analysis of mouse gut microbiota based on 16S rRNA gene amplicon sequencing with ensuing determination of core mouse gut bacteria, caecal samples were obtained from a total of 93 mice housed in eight different facilities in Europe and America (Supplementary Table 6). The aim was to obtain a data set that was not specific to one facility, and thus was as representative as possible of bacterial communities in the distal mouse intestine.

Mice were killed and caeca were immediately removed, placed in cryo-vials and snap-frozen in liquid nitrogen. Samples were stored at −80 °C until shipped on dry ice to the University of Nebraska at Lincoln and stored again at −80 °C until further processed. Once all samples were collected, they were thawed under anaerobic conditions and diluted 1:10 in sterile phosphate-buffered saline with 10% glycerol. Aliquots (200 μl) of each sample were used for DNA extraction following procedures described elsewhere[52]. 16S rRNA gene tag sequencing (MiSeq, Illumina) of V5–V6 regions was performed at the University of Minnesota Genomics Center using primers 784F (5′-RGGATTAGATACCC′) and 1064R (5′-CGACRRCCATGCANCACCT)[53].

Raw sequence reads were processed using IMNGS (www.imngs.org) with a pipeline developed in-house based on UPARSE[54]. Details about the analysis have already been described elsewhere[55]. OTUs clustered at 97% sequence identity and occurring at a relative abundance of ≥0.1% total reads in at least one sample were analysed further. The analysis delivered 14,906,121 quality- and chimera-checked reads (160,281 ± 60,686 per sample) that clustered in 772 OTUs (344 ± 45 OTUs per sample) (Supplementary Table 7 and Supplementary Sequence File).

For phylogenetic analysis, the evolutionary history was inferred by using the maximum likelihood method based on the 'General Time Reversible model' and

using 200 bootstrap replications in MEGA6 (ref. 56). The tree with the highest log-likelihood (representing the most probable tree topology) was selected. The percentage of trees in which the associated taxa clustered together was shown next to the branches (the tree was condensed to values ≥50%). All positions with <70% site coverage were eliminated. That is, fewer than 30% alignment gaps, missing data and ambiguous bases were allowed at any position. There were a total of 261 positions in the final data set.

Integrated 16S rRNA gene amplicon studies. To evaluate the prevalence and abundance of miBC members in gut microbiota, other host-derived ecosystems and environmental samples, we used an integrated metagenomic platform developed in-house and used in previous work (www.imngs.org)[16,57]. All 16S rRNA gene amplicon studies available in the SRA[58] were extracted and organized in sample-specific databases. The IMNGS build 1510 used in the present study contained a total of 55,073 samples, including 6,001 and 11,705 from the mouse and human gut, respectively. The corresponding databases were searched using UBLAST[59] based on nearly full-length 16S rRNA gene sequences of the 76 miBC species as queries to ensure good coverage across all studies, independent of 16S rRNA gene regions. Results were filtered on length (>200 nt), coverage (>70% read length) and similarity (>97 or 99%).

To determine whether certain species were host-specific, we looked at the number of SRA-derived samples positive for each species included in the collection, that is, those mouse or human samples in which sequences matched the corresponding 16S rRNA gene sequence at the level of 97% similarity. Of note, each OTU with a miBC match collected via IMNGS was subsequently assigned to one single closest sequence in the miBC reference set in order to avoid redundancy issues in the case of genera with closely related species (for example, for lactobacilli, enterococci, staphylococci). Prevalence had to be at least 5% in mouse gut samples so that species were considered for discriminative analysis between human and mice. The parameters used to categorize species were as follows: (1) rare: the species had a prevalence of <1% of mouse samples; (2) low prevalence: <5% mouse samples; (3) shared: the percentage of positive samples from human gut was at least as high as those from mice; (4) shared and dominant: definition in category 3 held true also when considering only SRA-derived OTUs occurring at a relative sequence abundance of >1% total reads (dominant OTUs); (5) mouse-enriched: the percentage of positive gut samples from mice was at least twice that from human subjects; (6) mouse-enriched and dominant: the given species was also enriched in mice when considering only dominant SRA-derived OTUs (as defined in category (4)).

For the interpretation of results, it is important to remember that the analysis described above is based on short 16S rRNA gene amplicons, which can affect similarity-based results when compared to full-length gene analysis. Moreover, although it is reasonable to assume that the large-scale analysis (including thousands of samples from many different studies) guarantees a certain degree of representativeness, sample descriptions in SRA can be imprecise and the analysis certainly includes outlier samples for a given ecosystem (for example, from hosts with peculiar pathophysiological conditions).

Genome sequencing and processing. A total of 53 draft genomes were obtained in the present study (EBI project accession no. PRJEB10572). The 53 strains were selected following three criteria: (1) novelty; (2) relevance according to the IMNGS and MGS analysis (mouse-enriched, prevalent or dominant); and (3) need to fill diversity gaps (species with no genome yet available). The miBC collection also includes 12 strains, for which novel genomes have already been deposited and will be published in a separate study (SRA accessions are provided in Supplementary Table 1). Genomes from 11 additional species in the collection were retrieved from NCBI (Supplementary Table 1). The corresponding raw sequence files were processed as for the genome sequences generated in-house.

Genomic DNA was obtained from pure cultures by precipitation after mechanical lysis following a protocol published previously[60]. DNA libraries were prepared using the TruSeq DNA PCR-Free Sample Preparation Kit (Illumina). The protocol was optimized (DNA shearing and fragment size selection) to improve assembly quality[61]. Libraries were sequenced using the Illumina MiSeq system according to the manufacturer's instructions.

Reads were assembled using Spades v3.6.1 with activated BayesHammer tool for error correction and MismatchCorrector module for post-assembly mismatch and indel corrections[62]. Assemblies were evaluated using Quast v3.1 (ref. 63). Prediction and annotation of open reading frames (ORFs) on contigs >1,000 nt were performed with prokka v1.11 (ref. 64). Genome sequences were submitted to the European Nucleotide Archive and are available under the accession numbers provided in Supplementary Table 1.

Genome-based analyses. 16S rRNA gene sequences were extracted from both the mouse isolate genomes and reference genomes using RNAmmer (ref. 65). All pairwise similarities among these sequences were calculated from exact pairwise sequence alignments using recommended settings[23].

For digital DNA:DNA hybridization (dDDH), the Genome-to-Genome Distance Calculator 2.0 (GGDC), a web service freely available at http://ggdc.dsmz.de, provided a genome sequence-based delineation of (sub-)species by reporting dDDH

estimates as well as their confidence intervals[66]. This approach was shown to have several advantages over alternative species delineation approaches[66,67], without mimicking the pitfalls of conventional DDH.

A whole-genome phylogeny (based on the proteome data) was inferred using the latest version of the Genome-BLAST Distance Phylogeny (GBDP) method[66,68]. Here, pairwise proteome comparisons (including pseudo-bootstrap replicates) were performed under the greedy-with-trimming algorithm and further recommended settings[69]. The final tree was inferred using FastME v2.07 with TBR post-processing[70]. GBDP settings were trimming algorithm, formula d5 and an e-value threshold of 10e-8. The tree was rooted at the mid-point[71] and numbers above branches are greedy-with-trimming pseudo-bootstrap support values from 100 replicates. Only support above 60% is shown.

The difference in genomic G+C content was used to delineate species. When computed from genome sequences, the G+C content differences vary no more than 1% within species[67].

Similarity search against the mouse gene catalogue. The mouse gene catalogue[3] was searched against each of the 15 genome databases separately, using BLASTn (ref. 72). The gene matched criteria were percent identity ≥95% over 100 bp or more and e-value ≤10×10^{-5} (very similar results were obtained using 80% coverage threshold with ≥95% identity). The number of gene matches assigned to metagenomic species (MGS), as defined previously[3,24], was counted (see Supplementary Table 5 for details). The occurrences of the five MGS through the mouse cohort sampled in ref. 3, with 85% of the genes matching the novel isolates, were used to generate Supplementary Fig. 1b.

Metagenome analysis. Metagenomic DNA was extracted from mouse intestinal contents as described above for pure cultures[60]. Libraries of 300 bp insert sizes were prepared from 500 ng fragmented DNA (Covaris S220 AFA System; duty factor 10%, peak incident power 175 W, 200 cycles per burst, 40 s) for each sample using the NEBNext Ultra DNA Library Prep Kit for Illumina (E7370S, New England Biolabs) according to the manufacturer's instructions. Samples were barcoded and sequenced on a TruSeq Rapid PE flowcell (PE-402-4001; Illumina) using an Illumina HiSeq 2500 sequencer with TruSeq Rapid SBS chemistry (FC-402-4001) and the 2 × 100 bp paired-end read module. Real-time analysis (RTA) software (1.17.20) was used for image analysis and base calling. Sequence files (.fastq) were generated with the CASAVA BCL2FASTQ Conversion Software (1.8.3). Paired-end sequencing reads were assembled into metagenomics contigs using Megahit[73]. Contigs with a length of >500 nt were classified with Kraken[74] and those classified as bacteria were kept for further analysis.

Three complementary approaches were followed to determine metagenomic coverage by cultivated bacteria. First, a local database was created of all miBC genomes and the metagenomes in a presence/absence binary code (1/0) of protein families (PFAM). Deterministic incremental selection of best fitted genomes was performed to assess coverage of the metagenomes by all miBC species ($n = 76$), or by species included in the minimal bacteriome MIBAC-1 ($n = 18$; this study), SIHUMI ($n = 8$)[27] and ASF ($n = 8$)[26]. Second, ORFs in metagenomic contigs were selected and initial annotation was carried out with prokka[64]. The predicted metaproteome of miBC and the three minimal bacteriomes were used for creation of protein BLAST databases. Sequence coverage of metagenomes was calculated based on a BLASTp search. ORFs that significantly (expected value $E < 10^{-5}$) aligned with a database entry protein at minimum length coverage ≥80% were used for the calculation of functional coverage (that is, at the level of homologous proteins). Third, from the ORFs retained in step (2), only those with ≥80% sequence similarity were used for calculation of close match sequence coverage. Of note, all newly generated or database-derived metagenomes used in the present analysis were completely independent of the samples used for strain isolation.

Description of novel bacteria. The descriptions are based on genome sequence analysis, enzymatic testing and chemotaxonomy (cellular fatty acids and detection of *meso*-diaminopimelic acid). Genome-based analyses included whole proteome-based phylogenomic GBDP analysis, dDDH, 16S rRNA gene sequence analysis and differences in G+C content of DNA. First, the relative subtree height (RSH) of any given putative novel bacteria within the phylogenomic tree relative to the RSH of related species allowed reliable conclusions on the taxonomic properties of novel strains. Second, a dDDH value of <70% indicated affiliation of an isolate to a novel species. Third, a 16S rRNA gene sequence identity of ≤94.5% was considered strong evidence for different genera and ≤86.5% for distinct families[75]. Fourth, because within-species differences in the genome-based G+C content of DNA are almost exclusively <1%, larger differences strongly supported the status of distinct species. Finally, percentage of conserved proteins (POCP) analysis was done using the IMG software tool Genus definition[76,77], considering genus delineation at a POCP value of <50%.

Novel bacterial taxa included (1) four species: 'Bacteroides caecimuris' strain I48^T (=DSM 26085^T = KCTC 15547^T), 'Blautia caecimuris' strain SJ18^T (=DSM 29492^T = KCTC 15541^T), 'Enterorhabdus muris' strain WAC-131-CoC-2^T (=DSM 29508^T = KCTC 15543^T) and 'Pasteurella caecimuris' strain AA-424-CC-1^T (=DSM 28627^T = KCTC 52216^T); (2) ten genera: 'Longicatena caecimuris' strain PG-426-CC-2^T (=DSM 29481^T = KCTC 15535^T), 'Longibaculum muris' strain MT10-315-

CC-1.2-2^T (=DSM 29487^T = KCTC 15536^T), 'Extibacter muris' strain JM-40^T (=DSM 28560^T = KCTC 15546^T), 'Muricomes intestini' strain 2-PG-424-CC-1^T (=DSM 29489^T = KCTC 15545^T), 'Turicimonas muris' strain YL45^T (=DSM 26109^T = KCTC 15542^T), 'Frisingicoccus caecimuris' strain PG-426-CC-1^T (=DSM 28559^T = KCTC 15538^T), 'Flintibacter butyricus' strain BLS21^T (=DSM 27579^T = KCTC 15544^T), 'Cuneatibacter caecimuris' strain BARN-424-CC-10^T (=DSM 29486^T = KCTC 15539^T), 'Acutalibacter muris' strain KB18^T (=DSM 26090^T = KCTC 15540^T), 'Irregularibacter muris' 2PG-426-CC-4.2^T (=DSM 28593^T = KCTC 15548^T); and (3) one member of a novel family: 'Muribaculum intestinale' strain YL27^T (=DSM 28989^T = KCTC 15537^T) within the 'Muribaculaceae'.

Description of Acutalibacter gen. nov. *Acutalibacter* (A.cu.ta.li.bac'ter. L. adj. *acutalis* tapered, pointed; N.L. n. *bacter* rod; N.L. masc. n. *Acutalibacter* a rod-shaped bacterium with tapered ends, pertaining to the cell morphology of the type strain of the type species).

The closest phylogenomic neighbour is *Clostridium sporosphaeroides*. Both genomes cluster distantly from *Clostridium leptum*. They do not represent members within the *Clostridium* sensu stricto cluster. The dDDH is 23% and POCP is 41%, indicating the status of a separate genus, for which the name *Acutalibacter* is proposed. The dDDH value between the genome of strain KB18^T and *Ruminococcus bromii* L2-63 is 25.6%. The G+C content of genomic DNA of the type strain is 54.6 mol%. The type species is *Acutalibacter muris*.

Description of Acutalibacter muris sp. nov. *Acutalibacter muris* (mu'ris L. gen. n. *muris* of a mouse). Its cells are long rods with tapered ends. It is strictly anaerobic. It is positive for α-galactosidase, β-galactosidase, β-galactosidase-6-phosphate, β-glucosidase, α-arabinosidase, β-glucuronidase and α-fucosidase (Rapid ID32A) and esculin hydrolysis (Api 20A).

D-xylopyranose can be formed from xylose, which is converted to D-xylulose by xylose isomerase activity. D-xylulose is subsequently phosphorylated to D-xylose 5-phosphate (E.C. 2.7.1.17, xylokinase, 01104), an intermediate of the pentose phosphate pathway. Hemicellulose consists of xylan, which is linked with sugar side chains (arabinose, galactose). Arabinoxylan can be converted to xylan by an arabinofructofuranose (E.C. 3.2.1.55, 00023). Xylan can be hydrolysed by E.C. 3.2.1.37, xylosidase (00237, 02775) or E.C. 3.2.1.54, dextrinase (03478). Several esterases are detectable within the genome.

The main cellular fatty acids comprise $C_{16:0}$ (22.1%), $C_{16:0\ DMA}$ (21.7%), iso-$C_{15:0}$ (16.8%), anteiso-$C_{15:0}$ (9.8%) and $C_{18:1\ \omega9c}$ (12.2%). The diagnostic amino acid of the peptidoglycan type is *meso*-diaminopimelic acid. The type strain is KB18^T (=DSM 26090^T = KCTC 15540^T).

Description of Bacteroides caecimuris sp. nov. *Bacteroides caecimuris* (cae.ci. mu'ris. L. n. *caecum* caecum; L. gen. n. *muris* of a mouse; N. L. gen. n. *caecimuris* from the caecum of a mouse). The closest phylogenomic neighbour is *Bacteroides acidifaciens* and the dDDH value between both genomes is 40.2%.

It is strictly anaerobic, Gram-negative, and in the form of short rods. It is positive for α- and β-galactosidase, β-galactosidase 6-phosphate, α- and β-glucosidase, α-arabinosidase, β-glucuronidase, α-fucosidase, alkaline phosphatase, leucine arylamidase and alanine arylamidase (Rapid ID 32A). It is positive for esculin hydrolysis and for acidification from D-glucose, D-lactose, saccharose, D-maltose, salicin, D-xylose, L-arabinose, D-cellobiose, D-mannose, D-raffinose, L-rhamnose and D-trehalose (Api 20A).

The strain is able to degrade pectin consisting of polygalacturonan and rhamnogalacturonan. The end products are D-galactopyranuronate and 4-deoxy-L-threo-hex-4-enopyranuronate. Pectin can be depolymerized by esterase activity. Enzymes include E.C. 3.1.1.11, Prokka_03269; E.C. 3.2.1.82, galacturonisidase, Prokka_03292; E.C. 4.2.2.23, rhamnogalacturonan endolyase, Prokka_03307; E.C. 4.2.2.24, rhamnogalacturonan exolyase, Prokka_03311; E.C. 3.2.1.172, rhamnogalacturonyl hydrolase, Prokka_03310.

β-L-rhamnopyranose can be converted to L-lactaldehyde and dihydroxyacetone phosphate, which can enter glycolysis (E.C. 5.3.1.14, L-rhamnose isomerase, Prokka_01322; E.C. 2.7.1.5, L-rhamnulokinase, Prokka_01323; E.C. 4.1.2.19, phosphate-aldolase, Prokka_01320). Xylose can be converted to D-xylulose 5-phosphate, which can enter the pentose phosphate pathway (E.C. 5.3.1.5, D-xylose epimerase, Prokka_03917; E.C. 2.7.1.17, kinase, Prokka_03916). The pentose phosphate pathway includes E.C. 5.3.1.6, Prokka_01016; E.C. 5.1.3.1, Prokka_00323; E.C. 2.2.1.1, Prokka_01617; E.C. 2.2.1.2, Prokka_03873; E.C. 2.2.1.1, Prokka_01617). The end product is D-glyceraldehyde 3-phosphate.

Trehalose can be converted to α-D-glucose and β-D-glucopyranose. The latter can be used to form β-D-glucose 6-phosphate, which can enter glycolysis (E.C. 3.2.1.28, Prokka_00582; E.C. 5.1.3.3, Prokka_01268; E.C. 2.7.1.2, Prokka_01994). Melibiose can be converted to α-D-galactose and α-D-glucose. Starch can be degraded to form β-D-glucose 6-phosphate and β-D-glucopyranose, which may enter glycolysis (E.C. 3.2.1, α-amylase, Prokka_03907; E.C. 2.4.1.1, Prokka_00503; E.C. 5.4.2.2, Prokka_01849; E.C. 2.4.1.25, Prokka_00656; E.C. 5.1.3.3, Prokka_01994).

L-glutamine can be converted to L-glutamate (E.C. 3.5.1.2, Prokka_01943), which can be dehydrogenized to 2-oxoglutatarate (E.C. 1.4.1.2, Prokka_02968) or converted to ammonium and fumarate (E.C. 2.6.1.1, Prokka_01814; E.C. 4.3.1.1, Prokka_02293).

Major cellular fatty acids are anteiso-$C_{15:0}$ (33.7%), iso-$C_{17:0\ 3-OH}$ (15.9%), iso-$C_{15:0}$ (13.3%), $C_{16:0\ 3-OH}$ (7.5%) and $C_{16:0}$ (4.8%). The G+C content of genomic DNA is 42.6%. The type strain is I48^T (=DSM 26085^T = KCTC 15547^T).

Description of *Blautia caecimuris* sp. nov. *Blautia caecimuris* (cae.ci.mu'ris. L. n. *caecum* caecum; L. gen. n. *muris* of a mouse; N. L. gen. n. *caecimuris* from the caecum of a mouse). The closest phylogenomic neighbour is *Blautia wexlerae* and the dDDH value between both genomes is 25.6%. The G+C content difference between strain SJ18^T and *Blautia wexlerae* is 1.6%.

The cells are coccoid and oval-shaped. It is a Gram-positive, strictly anaerobic bacterium. It is positive for α- and β-galactosidase, α-glucosidase, α-fucosidase, arginine arylamidase and leucine arylamidase (Rapid ID32A). It produces acid from D-glucose, D-lactose, D-saccharose, D-maltose, salicin, D-xylose, L-arabinose, D-raffinose, D-trehalose and L-rhamnose (Api 20A).

Nine alpha-amylases, which can be assigned to glycoside hydrolase family GH13 and catalyse the glycosidic linkage of glycosides, are detectable within the genome. All enzymes responsible for degradation of glycogen/starch are present: GlgX, a glycogen branching enzyme (E.C. 2.4.1.18; several copies), glycogenphosphorylase (E.C. 2.4.1.1, Prokka_03262), maltodextrinphosphorylase (E.C. 2.4.1.1, Prokka_02645) and phospoglucomutase (E.C. 5.4.2.2, Prokka_01686). A second pathway for glycogen degradation to D-glycopyranose 6-phosphate is also present: the first two steps are identical, but then maltotetraose is converted to maltotriose (E.C. 2.4.1.1, Prokka_02645), which is degraded to D-glycopyranose 6-phosphate: amylomaltase (E.C. 2.1.4.25, Prokka_02646); glucokinase (E.C. 2.7.1.2, Prokka_03476). The glucokinase is also involved in degradation of trehalose. Trehalose uptake is carried out by a trehalose-specific phosphotransferase system (PTS; Prokka_03240). The strain also possesses trehalose import proteins (Prokka_02649, 00179), transport system permease (Prokka_0062, 0067) and the operon repressor TreR (03239).

Chitobiose can also be transported and is subsequently phosphorylated to *N,N′*-diacetylchitobiose 6′-phosphate (E.C. 2.4.1.280; several copies), which is deacetylated (Prokka_03287) and hydrolysed by a glycosyl hydrolase to D-glucosamine and *N*-acetyl-D-glucosamine 6-phosphate (E.C. 3.2.1.86, Prokka_00121). *N*-acetyl-D-glucosamine 6-phosphate can enter the *N*-acetylglucosamine degradation I pathway for conversion to β-D-fructofuranose 6-phosphate, which may enter glycolysis (E.C. 3.5.1.25, *N*-acetylglucosamine 6-phosphate deacetylase, Prokka_00134; glucosamine 6-phosphate deaminase, Prokka_03782).

Xylose can be bound and imported by XylF and XylG proteins (Prokka_01553, 01554) in combination with the transport permease XylH (Prokka_01554). The xylose operon is regulated by XylR (Prokka_03489). D-xylopyranose is converted to D-xylulose by an isomerase (Prokka_02659) and phosphorylated to D-xylulose-5-phosphate by a kinase (Prokka_01743). The latter compound may enter the pentose phosphate pathway to form D-glyceraldehyde-3-phosphate (epimerase E.C. 5.1.3.1, Prokka_00665; isomerase E.C. 5.3.1.6, Prokka_02662; transketolase E.C. 2.2.1.1, several copies; transaldolase E.C. 2.2.1.2, Prokka_01937).

Lactose can be hydrolysed by a β-D-galactosidase (E.C. 3.2.1.23, Prokka_03707). Several amino acids can be degraded: arginine decarboxylase (E.C. 4.1.1.19, Prokka_03783), L-serine ammonia lyase (E.C. 4.3.1.17, Prokka_02259), threonine ammonia lyase (E.C. 4.3.1.19, Prokka_03091), methionine γ-lyase (E.C. 4.4.1.11, Prokka_00824) and D-cysteine (E.C. 4.4.1.15, Prokka_02595).

Major cellular fatty acids comprise $C_{18:0}$ (26.9%), $C_{12:0}$ (22.5%), iso I-$C_{19:1}$ (15.5%), $C_{16:0}$ (13.1%), iso -$C_{17:1}$/anteiso-$C_{17:1}$ (8.8%) and $C_{14:0}$ (4.8%). The diagnostic amino acid of the peptidoglycan type is *meso*-diaminopimelic acid. The G+C content of genomic DNA is 43.0 mol%. The type strain is SJ18^T (=DSM 29492^T = KCTC 15541^T).

Description of *Cuneatibacter* gen. nov. Cuneatibacter (Cu.ne.a.ti.bac'ter. L adj *cuneatus* wedge-shaped; L. masc. n. *bacter* rod; N.L. masc. n. *Cuneatibacter* a rod-shaped bacterium with wedge-shaped ends). The nearest phylogenomic neighbours are *Clostridium clostridioforme* and *Clostridium symbiosum*. It is placed into the Lachnospiraceae cluster, apart from the type species of the genus *Clostridium*, *Clostridium butyricum*. POCP between *C. butyricum* and *C. clostridioforme* or *C. symbiosum* is 39.9% and 39.7%, respectively. These low values confirm that both are not members of the *Clostridium* sensu stricto cluster. POCP between the genome of *C. clostridioforme* and *C. symbiosum* is 36.5% and dDDH is 22.2%. The dDDH values between strain BARN-424-CC-10^T and *C. clostridioforme* or *C. symbiosum* are 21.6% or 22.3%, respectively. The difference in G+C content of DNA is 1.4% for both comparisons. All these values confirm the separate genus status of the isolate. The G+C content of genomic DNA of the type strain is 49.1 mol%. The type species is *Cuneatibacter caecimuris*.

Description of *Cuneatibacter caecimuris* sp. nov. *Cuneatibacter caecimuris* (cae.ci. mu'ris. L. n. *caecum* caecum; L. gen. n. *muris* of a mouse; N. L. gen. n. *caecimuris* from the caecum of a mouse). Its cells are small rods with tapered ends, 2–3 μm in length. It is Gram-positive and strictly anaerobic. It is positive for α- and β-galactosidase, β-glucoronidase, α- and β-glucosidase and α-arabinosidase (Rapid ID32A), esculin hydrolysis and acidification from L-rhamnose (Api 20A).

β-D-galactose can be degraded to form UDP-glucose. In a first step, the aldose-1-epimerase converts β-D-galactose to α-D-galactose (E.C. 5.1.3.3, Prokka_02857), which is subsequently phosphorylated by a galactokinase (E.C. 2.7.1.6, Prokka_00204). As a result of galactose 1-phosphate uridinyltransferase activity

(E.C. 2.7.7.12, Prokka_00203), UDP-α-galactose is formed and rearranged to UDP-glucose (E.C. 00996; Prokka_0096).

N-acetyl-D-glucosamine can be degraded to form D-fructose 6-phosphate: *N*-acetyl-glucosamine 6-phosphate deacetylase (E.C. 3.5.1.25, Prokka_01387); glucosamine 6-phosphate isomerase (E.C. 3.5.99.6, Prokka_01388).

L-arabinose can be degraded to D-xylulose 5-phosphate due to L-arabinose isomerase activity (E.C. 5.3.1.4, Prokka_00574) and L-ribulose-5-phosphate 4-epimerase (E.C. 5.1.3.4; Prokka_01532). D-xylulose 5-phosphate may also be formed via degradation of D-xylose (E.C. 5.1.3.1.5, Prokka_01741; E.C. 2.7.1.17, Prokka_01860), which may enter the pentose phosphate pathway to generate D-fructose 6-phosphate (E.C. 2.2.1.1, transketolase, Prokka_00573; E.C. 2.2.1.2 transaldolase, Prokka_00768) and D-erythrose 4-phosphate. The latter can be converted to D-glyceraldehyde 3-phosphate by a transketolase activity, which may be converted to phosphoenolpyruvate (E.C. 1.2.1.12, Prokka_02031; E.C. 4.4.1.12, Prokka_02034; E.C. 2.7.2.3, Prokka_02032; E.C.5.4.2.12, Prokka_02345; E.C. 4.2.1.11, Prokka_02345). Phosphoenolpyruvate can be transformed to L-lactate (E.C. 2.7.1.40, pyruvate kinase, Prokka_01764; E.C. 1.1.1.27, dehydrogenase, Prokka_00578). D-erythrose 4-phosphate can also enter the chorismate biosynthesis pathway (all enzymes of this pathway are present in the genome of strain BARN-424-CC-10): E.C. 2.5.1.54, Prokka_00306; E.C. 4.2.3.4, Prokka_0226; E.C. 4.2.1.10, Prokka_00678; E.C. 1.1.24, Prokka_00680; E.C. 2.7.1.71, Prokka_03116; E.C. 2.5.1.19, Prokka_00308; E.C. 4.2.3.5, Prokka_00530.

The glycosyl hydrolases of family GH 28 (E.C. 3.2.1.15, polygalacturoronase), GH 78 (α-L-rhamnosidase) and GH105 (E.C. 4.2.2.23, rhamnogalacturonyl hydrolase) are involved in the degradation of pectin at the beginning of the pathway followed by activity of YesY (E.C. 3.1.1, Prokka_02705) and YesR (E.C. 3.2.1.172, Prokka_00382). The resulting 4-deoxy-L-threo-hex-4-enopyranuronate can be degraded by action of E.C. 5.3.1.17, isomerase, Prokka_01284; E.C. 1.1.1.67, dehydrogenase, Prokka_00822; E.C. 2.7.1.45, kinase, Prokka_00327 and aldolase (Prokka_00328). The end products are D-glyceraldehyde 3-phosphate and pyruvate.

A pathway for degradation of glycogen is detectable and includes the following reactions: phosphorylase (E.C. 2.4.1.1, Prokka_02446); 4-α-glucanotransferase (E.C. 2.4.1.25, Prokka_03092); α-glucosidase (E.C. 3.2.1.33; Prokka_02446); phosphorylase (E.C. 2.4.1.1, Prokka_02446) and phosphoglucomutase (E.C. 5.4.2.2, Prokka_02715). The end product is D-glucose 6-phosphate, which may enter glycolysis.

The predominant cellular fatty acid is $C_{16:0}$ (43.4%). Other fatty acids detectable with the MIDI system are iso-$C_{17:1}$/anteiso-$C_{17:1}$ (10.1%), iso-$C_{15:1}$/$C_{13:0\ OH}$ (9.0%), $C_{17:1\ ω8c}$ (8.3%) and $C_{14:0}$ (6.2%). *Meso*-diaminopimelic acid is absent. The type strain is BARN-424-CC-10^T (=DSM29486^T = KCTC 15539^T).

Description of *Enterorhabdus muris* sp. nov. *Enterorhabdus muris* (mu'ris L. gen. n. *muris* of a mouse). Phylogenomic analysis places strain WCA-131-CoC-2^T into the *Enterorhabdus* cluster. The dDDH value between the genome of *E. mucosicola* and *E. caecimuris* is 35.3% and between strain WCA-131-CoC-2^T, *E. mucosicola* and *E. caecimuris* is 38.3 and 43.1%, respectively. These values clearly confirm the separate species status of strain WCA-131-CoC-2^T.

The cells are rods, up to 5 μm in length, and are strictly anaerobic. The substrate profile in Api 20A kit is negative. Arginine hydrolase activity is detected with Rapid ID 32A. Arginine can be degraded to CO_2 by activity of the following enzymes: E.C. 3.5.3.6, deaminase, Prokka_01275; E.C. 2.1.3.3, carbamoyltransferase, Prokka_01150; E.C. 2.7.2.2, carbamate kinase, Prokka_00063. L-serine can be degraded to pyruvate: E.C. 4.3.1.17, Prokka_02086. Formate can be converted to 5,10-methylenetetrahydrofolate or to CO_2 (E.C. 6.3.4.3, ligase, Prokka_01739; E.C. 3.5.4.9, cyclohydrolase, Prokka_01741; E.C. 1.5.1.5, reductase, E.C. 1.2.1.2, Prokka_00906). Alkylnitronate can be metabolized to nitrite (E.C. 1.13.12.16, Prokka_00873). L-rhamnose can be degraded to form L-lactaldehyde dihydroxyacetone phosphate (E.C. 5.3.1.14, isomerase, Prokka_00371; E.C. 2.7.1.5, rhamnulokinase, Prokka_00372; E.C. 4.1.2.19, aldolase, Prokka_00370). All enzymes of the pentose phosphate pathway are present: E.C. 5.3.1.6, Prokka_01108; E.C. 5.3.1.1, Prokka_01187; E.C. 2.2.1.1, Prokka_02151; E.C. 2.2.1.2, Prokka_00123; E.C. 2.1.1, Prokka_02151.

The major cellular fatty acids are $C_{18:1\ ω9c}$ (14.3%), iso-$C_{17:1}$/anteiso-$C_{17:1}$ (12.1%), anteiso-$C_{15:0}$ (12.3%), iso-$C_{15:0}$ (7.9%), $C_{16:0}$ (13.8%), $C_{18:0}$ (7.5%), iso-$C_{14:0}$ (5.1%) and $C_{14:0}$ (3.7%). The diagnostic amino acid of the peptidoglycan type is *meso*-diaminopimelic acid. The G+C content of genomic DNA is 65.1 mol%. The type strain is strain WCA-131-CoC-2^T (=DSM 29508^T = KCTC 15543^T).

Description of *Extibacter* gen. nov. *Extibacter* (Ex.ti.bac'ter. L. neut. n. *extum* bowels or entrails of an animal; N.L. masc. n. *bacter* rod; N.L. masc. n. *Extibacter* a rod isolated from the intestine). The nearest phylogenetic neighbour is *Clostridium hylemonae*. POCP between the genome of *C. hylemonae* and *Clostridium butyricum*, the type species of the genus *Clostridium*, is only 14.3%, indicating the separate genus status of *C. hylemonae*. The dDDH value between both *Clostridium* genomes is 28.4% and between *C. hylemonae* and strain 40cc-B5824-ARE^T 23.8%. The G+C content of genomic DNA of the type strain is 47.9 mol%. Type species is *Extibacter muris*.

Description of *Extibacter muris* sp. nov. *Extibacter muris* (mu'ris, L. gen. n. *muris* of a mouse). Cells stain Gram-positive, and are strictly anaerobic rods up to 3 μm in

11

length. It is positive for β-galactosidase and proline arylamidase (Rapid ID32A) and acidification from D-glucose, D-lactose, L-rhamnose and xylose (Api20A).

Fructose 6-phosphate can be converted to acetyl-CoA via phosphoenolpyruvate. ABC transporters for Na^+, phosphate, cobalt, biotin, raffinose and melibiose are present. Trehalose can be converted via phosphoenolpyruvate to acetyl-CoA (E.C. 2.4.1.64, Prokka_03664). A β-galactosidase (E.C. 3.2.1.23, Prokka_01478), which may hydrolyse lactose, and all components of the xylose operon are present: xylose isomerase (Prokka_02910); 12 copies of *xylB* (kinase); the transport binding proteins XylF (Prokka_02319) and XylG (Prokka_02978); and the membrane transporter XylH (Prokka_02977). The xylFGH operon is positively regulated by XylR (Prokka_03195).

Degradation of glycogen/starch starts with a glycogenphosphorylase (E.C. 2.4.1.1; Prokka_00499, 02735), which releases α-D-glucopyranose 1-phosphate. The 1,6-glucohydrolase (E.C. 3.2.1, Prokka_02220) produces a debranched α-limit dextrin and shorter maltotetraose units first. α-D-glucopyranose 1-phosphate is then converted to D-glucopryanose 6-phosphate due to the activity of a phosphoglucomutase (E.C. 5.4.2.2, Prokka_00906). A maltodextrin phosphorylase cleaves maltodextrin to α-D-glucose 1-phosphate after several steps (E.C. 2.4.1.25, amylomaltase; Prokka_03040) and glucokinase activity (E.C. 2.7.1.2, Prokka_01911).

Several amino acids can be degraded and the enzymes involved are highly specific to their substrates: arginine decarboxylase (E.C. 4.1.1.19, Prokka_00457), L-serine ammonia lyase (E.C. 4.3.1.17, Prokka_00547), D-serine ammonia lyase (E.C. 4.3.1.18, Prokka_01317), threonine ammonia lyase (E.C. 4.3.1.19, Prokka_03487), methionine γ-lyase (E.C. 4.4.1.11, Prokka_01328).

Predominant fatty acids are $C_{16:0}$ (25.3%), $C_{14:0}$ (23.7%), iso-$C_{17:1}$/anteiso-$C_{17:1}$ (14.0%) and iso-$C_{15:1}$/$C_{13:0\ OH}$ (10.0%). *Meso*-diaminopimelic acid is absent. The type strain is 40ccB-5824-ARE^T (=DSM 28560^T = KCTC 15546^T).

Description of *Flintibacter* gen. nov.

Flintibacter (Flin.ti.bac'ter. N.L. n. *Flint* name Flint; L. masc. n. *bacter* rod; N.L. masc. n. *Flintibacter* rod-shaped bacterium named after Professor Harry J. Flint, a microbiologist from Aberdeen, UK, who dedicated his career to the investigation of gut bacteria, especially butyrate producers and their role in health and disease). The nearest phylogenomic neighbours are *Intestimonas butyriciproducens*, *Pseudoflavonifractor capillosus* and *Flavonifractor plautii*. Corresponding dDDH values are 24.4, 24.7 and 24.7%, respectively. The respective differences in G+C content of DNA are 1.4, 1.1 and 3.0%. The dDDH value between *P. capillosus* and *F. plautii* is 26.6% (and the POCP is 42.7%), and the dDDH between the genomes of *I. butyriciproducens* and *F. plautii* is 23.3%. The G+C content of genomic DNA of the type strain is 58 mol%. The type species is *Flintibacter butyricus*.

Description of *Flintibacter butyricus* sp. nov.

Flintibacter butyricus (bu.ty'ri.cus. N. L. masc. adj. *butyricus* related to butter, butyric, referring to the production of the short-chain fatty acid butyrate). It is strictly anaerobic, non-motile, non-spore forming, and catalase and oxidase negative, and stains Gram-negative. It grows at temperatures between 30 and 40 °C and in RCM broth containing up to 1% NaCl. It forms long filaments that separate into small irregular needle-like rods. It is negative for all reactions in Api 20A, and is positive for α-galactosidase in Rapid ID32A. It produces butyrate and acetate when grown in the presence of amino acids glutamine and glutamate, but there is no growth on leucine, lysine, arginine, glutamic and aspartic acid.

L-glutamate can be fermented by the hydroxyglutarate pathway to crotonyl-CoA, and further to butanoate (butyrate) and acetate. The following reactions are included in this pathway: E.C. 1.4.1.2, dehydrogenase, Prokka_00142; E.C. 2.8.3.12, transferase, Prokka_01566, 01567; E.C. 4.1.1.70, decarboxylase, Prokka_02670; E.C.1.1.1.157, 3-hydroxybutyryl-CoA dehydrogenase, Prokka_00188, 01591, 01835. Pyruvate can be converted to lactate (E.C. 1.1.1.27, Prokka_00634) or dehydrogenized (Prokka_00041) to form acetyl-CoA, which can be used to form acetate (E.C. 2.3.1.8, acetyl transferase, Prokka_00599).

Raffinose may be hydrolysed (E.C. 3.2.1.22, α-galactosidase) to form α-D-galactose and β-D-glucose. Melibiose can also be hydrolysed to α-D-galactose to form β-D-glucose 6-phosphate: E.C. 5.1.3.3, aldose epimerase, Prokka_02811; E.C. 2.7.1.6, galactokinase, Prokka_0072; E.C. 2.7.7.12, uridyltransferase, Prokka_0773; E.C. 1.3.2, UDP-glucose-4-epimerase, Prokka_01687; E.C. 5.4.2.2, phosphoglucomutase, Prokka_00685, 00705.

N-acetyl-glucosamine can be deacetylated by E.C. 3.5.1.25 (Prokka_01697) and converted to β-D-fructofuranose 6-phosphate by E.C. 3.5.99.6 (Prokka_01696). Fructofuranose 6-phosphate can also be formed from D-mannose, which is picked up by a PTS system. D-mannose 6-phosphate is converted to fructofuranose 6-phosphate via isomerase E.C. 5.3.1.8 (Prokka_01350).

A cellulase that forms cellobiose is detectable in the genome (E.C. 3.2.1.4, Prokka_02188), and cellobiose can be converted to a glucopyranose by glucosidase activity (E.C. 3.2.1.21, Prokka_02301).

Glycogen/starch can be degraded to form α-D-glucose 6-phosphate from maltotetraose or β-D-glucose 6-phosphate. The first reactions include E.C. 2.4.1.1 (Prokka_00610, 02033) and E.C. 3.2.1 (Prokka_03259), followed by E.C. 2.4.1.1 again and phosphoglucomutase activity (E.C. 5.4.2.2, Prokka_00685). Maltotriose can be converted to β-D-glucose 6-phosphate by glucosidase activity (E.C. 3.2.1.20, Prokka_03810), 4-α-glucotransferase (E.C. 2.4.1.25, Prokka _02266, 00609) and glucokinase K activity (E.C. 2.7.1.2, Prokka_01053, 01439).

Main cellular fatty acids are iso $C_{17:1}$/anteiso $C_{17:1}$ (16.1%), iso-$C_{19:1}$ (14.8%), $C_{18:1\ \omega9c}$ (13.5%), $C_{12:0}$ (12.2%), $C_{14:0}$ (11.4%), $C_{16:0}$ (11.2%) and $C_{18:0}$ (8.8%). The G+C content of the genomic DNA is 58 mol%. The type strain is BLS21^T (=DSM 27579^T = KCTC 15544^T).

Description of *Frisingicoccus* gen. nov.

Frisingicoccus (Fri.sing.i.coc'cus. L. fem. n. *Frisinga* the Roman name for the city of Freising; N.L. masc. n. *coccus* coccus; N.L. masc. n. *Frisingicoccus*; the type strain of the type species grows as coccobacilli and was originally isolated in the German city Freising). In the phylogenomic tree, strain PG-426-CC-1^T branches before the group composed of *Cuneatibacter muris* DSM 29486^T, *Clostridium clostridioforme* and *Clostridium symbosium*. The dDDH between both *Clostridium* genomes is 22.2%. dDDH of strain PG-426-CC-1^T against *Cun. muris* is 21.9% and against *C. clostridioforme* and *C. symbosium* 23.3 and 22.5%, respectively. The respective differences in genomic G+C content are 5.4, 5.4 and 4.0%, respectively. The G+C content of genomic DNA of the type strain is 43.7 mol%. The type species is *Frisingicoccus caecimuris*.

Description of *Frisingicoccus caecimuris* sp. nov.

Frisingicoccus caecimuris (cae.ci.mu'ris. L. n. *caecum* caecum; L. gen. n. *muris* of a mouse; N. L. gen. n. *caecimuris* from the caecum of a mouse). The cells are rods that separate into coccoid forms, and are strictly anaerobic. All Api 20A reactions are negative. It is positive for arginine and leucine arylamidase in Rapid ID 32A. It is able to utilize various amino acids by decarboxylation or non-oxidative deamination: arginine (E.C. 4.1.1.19, Prokka_01494); aspartate (E.C. 4.1.1.11, Prokka_00650); serine (E.C. 4.3.1.17, Prokka_00515); and threonine (E.C. 4.1.3.19, Prokka_02055). L-glutamate can be degraded to L-aspartate by transaminase activity (E. 2.6.1.1, Prokka-02564), which is metabolized to fumarate and NH_3 (E.C. 4.3.1.1, aspartate ammonia lyase, Prokka_00708).

A total of 50 ABC transporters can be detected in the genome of strain PG-426-CC-1, including those for phosphate, iron, branched chain amino acids, oligopeptides and D-methionine. Allose can be converted to β-D-fructofuranose 6-phosphate, which enters glycolysis (E.C. 2.7.1.55, D-allose kinase, Prokka_00524; E.C. 5.3.1.–, isomerase Prokka_00367). *N*-acetylglucosamine 6-phosphate may enter the cell by a PTS system and can be converted to D-glucosamine-6-phosphate by deacetylation and to β-D-fructofuranose 6-phosphate by deamination (E.C. 3.1.5.25, Prokka_00143; PTS phosphor carrier protein, Prokka_00706; PTS phosphorylation component, Prokka_00564; E.C. 3.5.99.6, Prokka_00155). β-D-fructofuranose 6-phosphate can then enter glycolysis.

The cellular fatty acids are $C_{18:1\ \omega9c}$ (22.3%), $C_{16:0}$ (22.1%), iso-$C_{17:1}$/anteiso-$C_{17:1}$ (10.3%) and $C_{18:0}$ (10.0%), with low amounts of $C_{14:0}$ (7.5%), iso-$C_{15:1}$/$C_{13:0\ 3-OH}$ (4.6%) and an unknown component (4.7%; RT 14.949). *Meso*-diaminopimelic acid is absent. The G+C content of genomic DNA is 43.7 mol%. The type strain is PG-426-CC-1^T (=DSM 28559^T = KCTC 15538^T).

Description of *Irregularibacter* gen. nov.

Irregularibacter (Ir.re.gu.la.ri.bac'ter. L. adj. *irrgeularis* irregular; N.L. masc. n. *bacter* rod; N.L. masc. n. *Irregularibacter* an irregular rod-shaped bacterium). Phylogenomic analysis places strain 2PG-426-CC-4.2^T into a separate lineage within the Firmicutes. dDDH values in comparison to the three closest phylogenomic neighbours (*Anaerofustis stercorihominis*, *Eubacterium limosum* and *Pseudoramibacter alactolyticus*) are 25.5, 28.5 and 31.5%, respectively. The respective differences in G+C content of DNA are 2.4, 11.7 and 15.9%. The RSH is larger than within the neighboured group that already contains three distinct genera. Hence, we propose to place the strain into a novel genus. The G+C content of genomic DNA of the type strain is 35.8 mol%. The type species is *Irregularibacter muris*.

Description of *Irregularibacter muris* sp. nov.

Irregularibacter muris (mu'ris L. gen. n. *muris* of a mouse). Cells are rods that can form long filaments of up to 10 μm. It is positive for indole production (Rapid ID 32A) and esculin hydrolysis (Api 20A). All other reactions are negative.

Glutamate can be degraded to n-butanoate including the following reactions: E.C. 1.4.1.2, dehydrogenase, Prokka_00372; E.C.2.8.3.12, CoA-transferase, Prokka_02420; CoA-dehydratase, Prokka_0065; E.C. 1.1.1.157, butyryl-CoA-dehydrogenase, Prokka_02002, 02766; E.C. 2.8.3.8, acetate CoA-transferase, Prokka_02840, 00642. Glycerol can be converted to glycerol-3-phosphate (E.C. 2.7.1.30, kinase, Prokka_02919, 00840) and further metabolized to form glycerone phosphate, which may enter glycolysis (E.C. 1.1.1.94, glycerol 3-phosphate dehydrogenase, Prokka_00228).

Glycogen biosynthesis from ADP-D-glucose is possible, including the following enzyme activities: E.C. 2.4.1.21, Prokka_02946; E.C. 2.4.1.18, Prokka_02943; E.C. 2.7.7.27, Prokka_02944; E.C. 5.4.2.2, phosphoglucomutase, Prokka_02192.

It is able to ferment amino acids: arginine (E.C. 4.1.1.19, Prokka_00095); threonine (E.C. 4.3.1.19, Prokka_00048); L-serine (E.C. 4.3.1.17, Prokka_00048). L-cysteine may be converted to pyruvate, NH_3 and H_2S (E.C. 4.4.1.1, Prokka_01832). Tryptophan may also be metabolized to indole, pyruvate and ammonium (E.C. 4.1.99.1; Prokka_00116).

The cellular fatty acids profile comprises $C_{18:1\ \omega9c}$ (23.0%), $C_{16:0}$ (19.6%), $C_{14:0}$ (18.5%) and $C_{18:0}$ (12.8%) as major components; $C_{12:0}$ (8.7%), $C_{16:1\ w7c}$/iso-$C_{15:0\ 2-OH}$ (5.3%) and $C_{10:0}$ (4.9%) also occur. *Meso*-diaminopimelic acid is absent. The type strain is 2PG-426-CC-4.2^T (=DSM 28593^T = KCTC 15548^T).

Description of *Longicatena* gen. nov. *Longicatena* (Lon.gi.ca.te'na. L. adj. *longus* long; L. fem. n. *catena* chain, N.L. fem. n. *Longicatena* the type strain of the type species grows as long chains). The closest phylogenomic neighbour of strain PG-426-CC-2^T is *Clostridium innocuum* and the corresponding dDDH is 21%. The dDDH between the isolate and the type species of the genus *Clostridium*, *Clostridium butyricum*, is 27.8%. The respective difference in genomic G+C content is 9.3%. *Eubacterium dolichum* is distantly related to strain PG-426-CC-2^T, as is the type species of the genus *Eubacterium*, *Eubacterium limosum*, with which strain PG-426-CC-2^T shares a dDDH of only 30.7%. The corresponding difference in genomic G+C content is 9.7%. This value supports the creation of a novel genus to accommodate strain PG-426-CC-2^T. The G+C content of genomic DNA of the type strain is 37.8 mol%. The type species is *Longicatena muris*.

Description of *Longicatena caecimuris* sp. nov. *Longicatena caecimuris* (cae.ci. mu'ris. L. n. *caecum* caecum; L. gen. n. *muris* of a mouse; N. L. gen. n. *caecimuris* from the caecum of a mouse). The cells are rods, forming long filaments. All reactions in Rapid ID 32A are negative, but it is positive for esculin hydrolysis in Api 20A.

The complete pathway for formaldehyde oxidation is present, starting with D-ribulose 5-phosphate and formaldehyde to form hexulose 6-phosphate: E.C. 4.1.2.43, hexulose 6-phosphate synthase, Prokka_00565; E.C. 5.3.1.27, hexuloisomerase, Prokka_00566; E.C. 5.3.1.9, glucose 6-phosphate isomerase, Prokka_01514; E.C. 1.1.1.49, glucose 6-phosphate dehydrogenase, Prokka_02255; E.C. 1.1.31 phosphoglucanolactonase, Prokka_02253; E.C. 1.1.1.343, phosphogluconate dehydrogenase, Prokka_02254).

Pyruvate can be converted to lactate (E.C. 1.1.1.27, Prokka 01579). L-serine and L-threonine can be fermented (E.C. 43, 1.17, Prokka_02192; E.C. 4.3.1.19, Prokka_01200). Several enzymes involved in the degradation of glycogen are available: E.C. 2.4.1, phosphorylase, Prokka_01627, 02537; E.C. 2.4.1.25, transferase, Prokka_01628; E.C. 5.4.2.2, phosphoglucomutase, Prokka_01780.

The predominant fatty acid is $C_{18:1\ \omega9c}$ (77.4%), and minor fatty acids are $C_{18:1\ \omega7c}$ (5.7%), $C_{16:0}$ (4.6%), iso-$C_{19:0}$ (2.9%), $C_{18:1-2OH}$ (2.0%) and $C_{18:0}$ (1.9%). *Meso-* diaminopimelic acid is absent. The type strain is PG-426-CC-2^T (=DSM 29481^T = KCTC 15535^T).

Description of *Longibaculum* gen. nov. *Longibaculum* (Lon.gi.ba'cu.lum. L. adj. *longus* long; L. neut. n. *baculum* rod, stick; N.L. neut. n. *Longibaculum* a long rod, pertaining to the ability to grow as long cells, filaments). Phylogenomic analysis places strain MT10-315-CC-1.2-2^T into the Erysipelotrichaceae cluster, distantly related to all other members of this cluster. The RSH of strain MT10-315-CC-1.2-2^T is larger than within the neighboured group, which includes four distinct genera (*Catenibacterium mitsuokai*, *Kandleria vitulina*, *Sharpea azabuensis* and *Eggerthia catenaformis*). The nearest neighbour is 'Stoquefichus massiliensis', which currently has no standing in nomenclature and was therefore not included in the analysis. We thus propose that strain MT10-315-CC-1.2-2^T represents a novel genus. The G+C content of genomic DNA is 30.8 mol%. The type species is *Longibaculum muris*.

Description of *Longibaculum muris* sp. nov. *Longibaculum muris* (mu'ris L. gen. n. *muris* of a mouse). It is able to form long filaments >20 μm in length, and is strictly anaerobic. It is positive for β-galactosidase, β-glucoronidase, *N*-acetyl glucosamine and pyroglutamic acid arylamidase (Rapid ID 32A). All reactions for acidification of carbohydrates are negative (Api 20A).

The genome contains 30 glucosidases of family GH1. Arbutin and salicin can be converted to β-D-glucose-6-phosphate (E.C. 2.7.1.69, phosphotransferase, Prokka_00070; 3.2.1.86, glucosidase, Prokka_00223). Sucrose can be degraded to form β-D-fructofuranose-6-phosphate (E.C. 3.2.1.26, hydrolase, Prokka_01036; E.C. 2.7.1.2, glucokinase, Prokka_00474; E.C. 5.3.1.9, Prokka_00116; or directly by fructokinase activity (E.C. 2.7.1.4; Prokka_00908). Rhamnose can be converted to L-lactaldehyde dihydroacetone, which may enter glycolysis (E.C. 5.3.1.14, isomerase, Prokka_00519; E.C. 2.7.1.5, kinase, Prokka_00518; E.C. 4.1.2.19, aldolase, Prokka_00520). β-D-galactose can be converted to glucose 6-phophaste, which can also enter glycolysis (E.C. 5.1.3.3, epimerase, Prokka_02823; E.C. 2.7.1.6, galactokinase, Prokka_02824; E.C. 2.7.7.12, transferase, Prokka_02822). Lactose can be used to form β-D-galactose and β-D-glucose (E.C. 3.2.1.23, Prokka_00851).

A type II secretion pathway including eps E, F, H is present. The following enzymes for degradation of glycogen are also present: E.C. 2.4.1.1, phosphorylase, Prokka_02190; E.C. 2.4.1.25, glucotransferase, Prokka_01371; E.C. 3.2.1.10, 1,6-glucosidase Prokka_02259, 03135; E.C.5.4.2.2, phosphoglucomutase, Prokka_02540.

The main fatty acids are $C_{16:0}$ (30.1%), $C_{18:1\ \omega9c}$ (15.4%), $C_{18:1\ \omega7c}$ (9.9%), $C_{16:1\ \omega7c}$/iso-$C_{15:0-OH}$ (8.9%), $C_{18:0}$ (6.6%) and $C_{16:1\ \omega11c}$ (5.4%). The diagnostic amino acid of the peptidoglycan type is *meso*-diaminopimelic acid. The type strain is MT10-315-CC-1.2-2^T (=DSM 29487 T = KCTC 15536^T).

Description of *Muribaculum* gen. nov. *Muribaculum* (L. n. *mus*, *muris* a mouse; L. n. *baculum* rod, stick; *Muribaculum* rod-shaped bacterium from mice).

Phylogenomic analysis places strain YL27^T into a separate lineage nearby the genera *Coprobacter* and *Barnesiella*. The nearest phylogenetic neighbour as determined by 16S rRNA gene sequence analysis is *Barnesiella intestinihomis*, showing only 86.2% sequence similarity, which is below the proposed threshold of

86.5% for families[75]. dDDH values between strain YL27^T and the genomes of *Coprobacter fastidiosus*, *B. intestinihominis* and *Barnesiella viscericola* are 23.8, 23.4 and 23.4%, respectively. Corresponding differences in genomic G+C content are 11.6, 6.0 and 1.7%. The G+C content of genomic DNA of the type strain is 49.9 mol%. The type species is *Muribaculum intestinale*.

Description of *Muribaculum intestinale* sp. nov. *Muribaculum intestinale* (in.tes.ti. na'le. N.L. neut. adj. *intestinale* pertaining to the intestines). It is strictly anaerobic. It is positive for α-galactosidase, β-galactosidase, *N*-acetyl-β-glucosamidase, alkaline phosphatase, leucyl glycine arylamidase and alanine arylamidase (Rapid ID 32A), and is negative for all reactions in Api 20A.

Galactose can be degraded to form β-D-glucose 6-phosphate (E.C. 5.1.3.3, epimerase, Prokka_01653; E.C. 2.7.1.6, kinase, Prokka_01654; E.C. 5.1.3.2, epimerase, Prokka_02048; E.C. 5.4.2.2, phosphoglucomutase, Prokka_00248).

Pectin can be depolymerized to D-galactopyranuronate (E.C. 3.1.1.11, pectin esterase, Prokka_02543; E.C. 3.2.1.82, galacturonidase, Prokka_02522; E.C. 4.2.2.2, pectic acid lyase PL10, Prokka_00090).

The main cellular fatty acid is anteiso-$C_{15:0}$ (59.9%); $C_{16:0-30H}$ (8.3%), iso-$C_{15:0}$ (5.6%), $C_{14:0}$ (3.9%) and $C_{15:0\ 3-OH}$ (3.3%) are detected in minor amounts. *Meso*-diaminopimelic acid is absent. The type strain is YL27^T (=DSM 28989^T = KCTC 15537^T).

Description of *Muricomes* gen. nov. *Muricomes* (Mu.ri.co'mes, L.n. *mus*, *muris* mouse; L.n. *comes* a companion; N.L. masc. n., *Muricomes* a companion of a mouse). Phylogenomic analysis places strain 2-PG-424-CC-1^T into a separate lineage with a RSH clearly larger than within the neighboured group holding species from two distinct genera: *Clostridium oroticum*, *Eubacterium contortum* and *Eubacterium fissicatena*. These species share dDDH values of 20.1, 20.0 and 20.1% with strain 2-PG-424-CC-1^T, respectively. These data indicate that strain 2-PG-424-CC-1^T represents a novel genus. *C. oroticum* is characterized by dDDH values of 21.0 and 21.2% with *E. contortum* and *E. fissicatena*, respectively. The two *Eubacterium* species show a dDDH of 27.6%. They are distantly related to the type species of the genus *Eubacterium* (*Eubacterium limosum*) and must be reclassified in future. The G+C content of genomic DNA of the type strain is 43.1 mol%. The type species is *Muricomes intestini*.

Description of *Muricomes intestini* sp. nov. *Muricomes intestini* (in.tes'ti.ni. L. gen. n. *intestini* of the gut, referring to the ecosystem of origin of the type strain).

Its cells are Gram-positive rods able to grow under anaerobic conditions, and it is positive for α-glucosidase and proline arylamidase (Rapid ID32A). It is able to utilize L-arabinose, D-glucose, D-mannitol, D-maltose, D-mannose, L-rhamnose, D-sorbitol and D-xylose (Api 20A).

Carbohydrate binding modules (CBM) with carbohydrate-binding activities are detectable within the genome for xylan/mannan (CBM 35 family; Prokka_00901) and chitin (CBM 50 family, Prokka_01736). Chitin can be degraded to *N*-acetylglucosamine by a diacetylchitobiase (E.C. 2.4.1.280, Prokka_01771). Subsequently, *N*-acetyl-glucosamine 6-phosphate is formed by the activity of a phosphotransferase system and can be transformed to D-glucosamine 6-phosphate due to deacetylase activity (E.C. 3.5.1.25, Prokka_02773). Deamination (E.C. 3.5.99.6, Prokka_01084) leads to formation of D-fructofuranose 6-phosphate. Cellulose can be degraded (E.C. 3.2.1.4, Prokka_00599) to form more soluble components. An arabinofuranidase (E.C. 3.2.1.55, Prokka_02409) can extract arabinose and xylan to produce xylodextrin and subsequently xylotriose and xylobiose (E.C. 3.2.1.54, dextrinase, Prokka_02684).

All enzymes for degradation of glycogen to α-glucose 6-phosphate are detectable within the genome: glycogenphosphorylase (E.C. 2.4.1.1, Prokka_01745); α-1,6 glucohydrolase (E.C. 3.2.1.111, Prokka_00827); maltodextrinphosphorylase (E.C. 2.4.1.1, Prokka_02991); and a phosphoglucomutase (E.C. 5.4.2.2, Prokka_01606). A second pathway for glycogen degradation to D-glycopyranose 6-phosphate is also present: the first two steps are as above, but maltotetraose is then converted to maltotriose (E.C. 2.4.1.1, phosphorylase; Prokka_02991), which is degraded to D-glycopyranose 6-phosphate (E.C. 2.1.4.25, amylomaltase, Prokka_02987; E.C. 2.7.1.1, glucokinase, Prokka_0362).

β-D-glucose 6-phosphate and β-D-fructofuranose 6-phosphate can be converted to phosphoenolpyruvate and pyruvate via glycolysis. Phosphoenolpyruvate and pyruvate can be utilized within a mixed acid fermentation with acetate, CO_2 and H_2 being the end products: pyruvate kinase (E.C. 2.7.1.40, Prokka_03100); pyruvate formate-lyase (E.C. 2.3.1.54, Prokka_00010); formate hydrogenlyase E.C. 1.2.1.2, Prokka_03349); phosphate acetyltransferase (E.C. 2.3.1.8, Prokka_00287); acetate kinase (E.C. 2.7.2.1, Prokka_00288).

The fatty acid profile comprises $C_{16:0}$ (40.7%), $C_{14:0}$ (15.7%), iso-$C_{15:1}$/$C_{13:0\ 3-OH}$ (12.2%), $C_{19:0\ 10-methyl}$ (7.6%), iso-$C_{17:1}$ (5.2%) and $C_{18:1\ \omega7c}$ (5.2%). The diagnostic amino acid of the peptidoglycan type is *meso*-diaminopimelic acid. The type strain is 2-PG-424-CC-1^T (=DSM 29489^T = KCTC 15545^T).

Description of *Pasteurella caecimuris* sp. nov. *Pasteurella caecimuris* (cae.ci.mu'ris. L. n. *caecum* caecum; L. gen. n. *muris* of a mouse; N. L. gen. n. *caecimuris* from the caecum of a mouse). The nearest phylogenomic neighbour is *Pasteurella pneumotropica*. dDDH between the genome of strain AA-424-CC-1^T and

P. pneumotropica is 30.8%, confirming the separate species status of strain AA-424-CC-1^T. It is Gram-negative. It is positive for resorufin-β-D-galactopyranoside, β-galactosidase, α-galactosidase, alkaline phosphatase, trehalose, maltose, urease, potassium nitrate, glucose and cytochrome oxidase (Rapid ID32 Strep and Api 20NE). It has weak reactions for D-ribose, D-lactose, D-raffinose, sodium pyruvate, melibiose and sucrose.

N-acetyl D-glucosamine 6-phosphate can be converted to β-D-fructofuranose 6-phosphate (E.C. 3.5.1.25, deacetylase, Prokka_02486; E.C. 3.5.99.6, deaminase, Prokka_02485). Glycogen debranching enzymes are present: E.C. 3.2.1.33, α-glucosidase, Prokka_02499; E.C. 2.4.1.1, glycogen phosphorylase, Prokka_02298; E.C. 5.4.2.8, mannomutase, Prokka_02405; E.C. 2.4.1.25, glucanotransferase, Prokka_02299; E.C. 2.4.1.60, kinase, Prokka_02482. Sucrose can be converted to β-D-fructofuranose, including the following reactions: E.C. 2.7.7.9, transferase, Prokka_00902; E.C. 2.7.1.4, fructokinase, Prokka_01004; E.C. 5.4.2.8, mannomutase, Prokka_02405; E.C. 5.3.1.9, isomerase, Prokka_01476. Xylose can be converted to D-xylulose 5-phosphate, which may enter the pentose phosphate pathway: E.C. 5.3.1.5, xylulose isomerase, Prokka_01918; E.C. 2.7.1.17, xylulose kinase, Prokka_01919.

Enzymes for a mixed acid fermentation are detectable. Phosphoenolpyruvate can be metabolized to form succinate, but also to form lactate from pyruvate. Pyruvate can also be converted to formate and CO_2 or to acetate via acetyl-CoA: E.C. 2.7.1.40, kinase, Prokka_01479; E.C. 1.1.1.28, dehydrogenase, Prokka_00386; E.C. 2.3.1.54, transferase, Prokka_02383; E.C. 2.3.1.8 phosphotransferase, Prokka_02157; E.C. 2.7.2.1, acetate kinase, Prokka_02158; E.C. 4.1.1.31, PEP-carboxylase, Prokka_01077; E.C. 1.1.1.37, dehydrogenase, Prokka_01374; E.C. 4.2.1.2, fumarate dehydrogenase, Prokka_02131; E.C. 1.3.5.4, fumarate reductase, Prokka_01183.

The fatty acid profile comprises $C_{16:0}$ (31.0%), $C_{14:0}$ (29.4%), $C_{16:1\ \omega7c}$/iso-$C_{15:0\ 2-OH}$ (24.4%) and iso-$C_{16:1}$/$C_{14:0\ 3-OH}$ (8.2%). The G+C content of genomic DNA is 40.2 mol%. The type strain is AA-424-CC-1^T (=DSM 28627^T = KCTC 52216^T).

Description of *Turicimonas* gen. nov. *Turicimonas* (Tu.ri.ci.mo'nas. L. neut. n *Turicum* the Roman name for the city of Zurich; Gr. fem. n. *monas* a unit, a monad; N.L. fem. n. *Turicimonas* a unit of Turicum/Zurich; the type strain of the type species was isolated in the Swiss city Zurich). The closest phylogenomic neighbour is *Parasutterella excrementihominis*. The dDDH value and 16S rRNA gene sequence identity between the two bacteria are 21.4 and 93.9%, respectively. The amino acid identity (AAI) value between the two bacteria is 66.1% and the difference in genomic G+C content is 4.1%. The G+C content of genomic DNA of the type strain is 44.0 mol%. The type species is *Turicimonas muris*.

Description of *Turicimonas muris* sp. nov. *Turicimonas muris* (mu'.ris. L. gen. n. *muris* of a mouse). It grows as single, straight, rod-shaped cells, 1–2 μm in length. All reactions are negative in Api 20A and Rapid ID 32A. It is able to degrade L-asparagine (E.C.5.1.1, Prokka_02303) and L-serine (E.C. 4.3.1.17, Prokka_01696). The type strain is YL45^T (= DSM 26109^T = KCTC 15542^T).

References

1. Browne, H. P. *et al.* Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533,** 543–546 (2016).
2. Hugon, P. *et al.* A comprehensive repertoire of prokaryotic species identified in human beings. *Lancet Infect. Dis.* **15,** 1211–1219 (2015).
3. Xiao, L. *et al.* A catalog of the mouse gut metagenome. *Nature Biotechnol.* **33,** 1103–1108 (2015).
4. Chung, H. *et al.* Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* **149,** 1578–1593 (2012).
5. Frese, S. A. *et al.* The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. *PLoS Genet.* **7,** e1001314 (2011).
6. Seedorf, H. *et al.* Bacteria from diverse habitats colonize and compete in the mouse gut. *Cell* **159,** 253–266 (2014).
7. Clavel, T., Lagkouvardos, I., Blaut, M. & Stecher, B. The mouse gut microbiome revisited: From complex diversity to model ecosystems. *Int. J. Med. Microbiol.* http://dx.doi.org/10.1016/j.ijmm.2016.03.002 (2016).
8. Kyrpides, N. C. *et al.* Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. *PLoS Biol.* **12,** e1001920 (2014).
9. Gordon, J. H. & Dubos, R. The anaerobic bacterial flora of the mouse cecum. *J. Exp. Med.* **132,** 251–260 (1970).
10. Clavel, T. *et al.* Isolation of bacteria from the ileal mucosa of TNFdeltaARE mice and description of *Enterorhabdus mucosicola* gen. nov., sp. nov. *Int. J. Syst. Evol. Microbiol.* **59,** 1805–1812 (2009).
11. Clavel, T., Charrier, C. & Haller, D. *Streptococcus danieliae* sp. nov., a novel bacterium isolated from the caecum of a mouse. *Arch. Microbiol.* **195,** 43–49 (2013).
12. Clavel, T., Charrier, C., Wenning, M. & Haller, D. *Parvibacter caecicola* gen. nov., sp. nov., a bacterium of the family Coriobacteriaceae isolated from the caecum of a mouse. *Int. J. Syst. Evol. Microbiol.* **63,** 2642–2648 (2013).
13. Clavel, T. *et al.* *Enterorhabdus caecimuris* sp. nov., a member of the family *Coriobacteriaceae* isolated from a mouse model of spontaneous colitis, and emended description of the genus *Enterorhabdus*. *Int. J. Syst. Evol. Microbiol.* **60,** 1527–1531 (2010).
14. Clavel, T., Saalfrank, A., Charrier, C. & Haller, D. Isolation of bacteria from mouse caecal samples and description of *Bacteroides sartorii* sp. nov. *Arch. Microbiol.* **192,** 427–435 (2010).
15. Kläring, K. *et al.* *Intestinimonas butyriciproducens* gen. nov., sp. nov., a butyrate-producing bacterium from the mouse intestine. *Int. J. Syst. Evol. Microbiol.* **63,** 4606–4612 (2013).
16. Kläring, K. *et al.* *Murimonas intestini* gen. nov., sp. nov., an acetate-producing bacterium of the family Lachnospiraceae isolated from the mouse gut. *Int. J. Syst. Evol. Microbiol.* **65,** 870–878 (2015).
17. Pfeiffer, N. *et al.* *Acetatifactor muris* gen. nov., sp. nov., a novel bacterium isolated from the intestine of an obese mouse. *Arch. Microbiol.* **194,** 901–907 (2012).
18. Killer, J. *et al.* *Lactobacillus rodentium* sp. nov., from the digestive tract of wild rodents. *Int. J. Syst. Evol. Microbiol.* **64,** 1526–1533 (2014).
19. Miyamoto, Y. & Itoh, K. *Bacteroides acidifaciens* sp. nov., isolated from the caecum of mice. *Int. J. Syst. Evol. Microbiol.* **50** (Pt 1), 145–148 (2000).
20. Osawa, R., Fujisawa, T. & Pukall, R. *Lactobacillus apodemi* sp. nov., a tannase-producing species isolated from wild mouse faeces. *Int. J. Syst. Evol. Microbiol.* **56,** 1693–1696 (2006).
21. Kaneuchi, C., Miyazato, T., Shinjo, T. & Mitsuoka, T. Taxonomic study of helically coiled, sporeforming anaerobes isolated from the intestines of humans and other animals: *Clostridium cocleatum* sp. nov. and *Clostridium spiroforme* sp. nov. *Int. J. Syst. Bacteriol.* **29,** 1–12 (1979).
22. Söhngen, C. *et al.* BacDive—the bacterial diversity metadatabase in 2016. *Nucleic Acids Res.* **44,** D581–D585 (2015).
23. Meier-Kolthoff, J. P., Goker, M., Sproer, C. & Klenk, H. P. When should a DDH experiment be mandatory in microbial taxonomy? *Arch. Microbiol.* **195,** 413–418 (2013).
24. Nielsen, H. B. *et al.* Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nature Biotechnol.* **32,** 822–828 (2014).
25. Nguyen, T. L., Vieira-Silva, S., Liston, A. & Raes, J. How informative is the mouse for human gut microbiota research? *Dis. Model Mech.* **8,** 1–16 (2015).
26. Wannemuehler, M. J., Overstreet, A. M., Ward, D. V. & Phillips, G. J. Draft genome sequences of the altered schaedler flora, a defined bacterial community from gnotobiotic mice. *Genome Announc.* **2,** e00287–e00214 (2014).
27. Becker, N., Kunath, J., Loh, G. & Blaut, M. Human intestinal microbiota: characterization of a simplified and stable gnotobiotic rat model. *Gut Microbes* **2,** 25–33 (2011).
28. Salzman, N. H. *et al.* Analysis of 16S libraries of mouse gastrointestinal microflora reveals a large new group of mouse intestinal bacteria. *Microbiology* **148,** 3651–3660 (2002).
29. Bui, T. P. *et al.* Production of butyrate from lysine and the Amadori product fructoselysine by a human gut commensal. *Nature Commun.* **6,** 10062 (2015).
30. Kibe, R., Sakamoto, M., Yokota, H. & Benno, Y. Characterization of the inhabitancy of mouse intestinal bacteria (MIB) in rodents and humans by real-time PCR with group-specific primers. *Microbiol. Immunol.* **51,** 349–357 (2007).
31. Sun, Y., Zhou, L., Fang, L., Su, Y. & Zhu, W. Responses in colonic microbial community and gene expression of pigs to a long-term high resistant starch diet. *Front. Microbiol.* **6,** 877 (2015).
32. Goodman, A. L. *et al.* Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc. Natl Acad. Sci. USA* **108,** 6252–6257 (2011).
33. Walker, A. W., Duncan, S. H., Louis, P. & Flint, H. J. Phylogeny, culturing, and metagenomics of the human gut microbiota. *Trends Microbiol.* **22,** 267–274 (2014).
34. Carvalho, F. A. *et al.* Interleukin-1β (IL-1β) promotes susceptibility of Toll-like receptor 5 (TLR5) deficient mice to colitis. *Gut* **61,** 373–384 (2012).
35. Macpherson, A. J. & McCoy, K. D. Standardised animal models of host microbial mutualism. *Mucosal Immunol.* **8,** 476–486 (2015).
36. Schaedler, R. W., Dubs, R. & Costello, R. Association of germfree mice with bacteria isolated from normal mice. *J. Exp. Med.* **122,** 77–82 (1965).
37. McNulty, N. P. *et al.* Effects of diet on resource utilization by a model human gut microbiota containing *Bacteroides cellulosilyticus* WH2, a symbiont with an extensive glycobiome. *PLoS Biol.* **11,** e1001637 (2013).
38. Laukens, D., Brinkman, B. M., Raes, J., De Vos, M. & Vandenabeele, P. Heterogeneity of the gut microbiome in mice: guidelines for optimizing experimental design. *FEMS Microbiol. Rev.* **40,** 117–132 (2015).

39. Reyes, A., Wu, M., McNulty, N. P., Rohwer, F. L. & Gordon, J. I. Gnotobiotic mouse model of phage–bacterial host dynamics in the human gut. *Proc. Natl Acad. Sci. USA* **110,** 20236–20241 (2013).

40. Scupham, A. J. *et al.* Abundant and diverse fungal microbiota in the murine intestine. *Appl. Environ. Microbiol.* **72,** 793–801 (2006).

41. Greenblum, S., Carr, R. & Borenstein, E. Extensive strain-level copy-number variation across human gut microbiome species. *Cell* **160,** 583–594 (2015).

42. Wang, J. *et al.* Dietary history contributes to enterotype-like clustering and functional metagenomic content in the intestinal microbiome of wild mice. *Proc. Natl Acad. Sci. USA* **111,** E2703–E2710 (2014).

43. Aranki, A. & Freter, R. Use of anaerobic glove boxes for the cultivation of strictly anaerobic bacteria. *Am. J. Clin. Nutr.* **25,** 1329–1334 (1972).

44. Osawa, R. Formation of a clear zone on tannin-treated brain heart infusion agar by a Streptococcus sp. isolated from feces of koalas. *Appl. Environ. Microbiol.* **56,** 829–831 (1990).

45. Hall, T. A. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41,** 95–98 (1999).

46. Kim, O. S. *et al.* Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int. J. Syst. Evol. Microbiol.* **62,** 716–721 (2012).

47. Schumann, P. & Maier, T. MALDI-TOF mass spectrometry applied to classification and identification of bacteria. *Methods Microbiol.* **41,** 275–306 (2014).

48. Schumann, P. Peptidoglycan structure. *Methods Microbiol.* **38,** 101–129 (2011).

49. Rhuland, L. E., Work, E., Denman, R. F. & Hoare, D. S. The behavior of the isomers of α,ε-diaminopimelic acid on paper chromatograms. *J. Am. Chem. Soc.* **77,** 4844–4846 (1955).

50. Kuykendall, L. D., Roy, M. A., O'Neill, J. J. & Devine, T. E. Fatty acids, antibiotic resistance, and deoxyribonucleic acid homology groups of *Bradyrhizobium japonicum. Int. J. Syst. Bacteriol.* **38,** 358–361 (1988).

51. Miller, L. T. Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxyl acids. *J. Clin. Microbiol.* **16,** 584–586 (1982).

52. Martinez, I. *et al.* Diet-induced metabolic improvements in a hamster model of hypercholesterolemia are strongly linked to alterations of the gut microbiota. *Appl. Environ. Microbiol.* **75,** 4175–4184 (2009).

53. Krumbeck, J. A. *et al. In vivo* selection to identify bacterial strains with enhanced ecological performance in synbiotic applications. *Appl. Environ. Microbiol.* **81,** 2455–2465 (2015).

54. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* **10,** 996–998 (2013).

55. Lagkouvardos, I. *et al.* Gut metabolites and bacterial community networks during a pilot intervention study with flaxseeds in healthy adult men. *Mol. Nutr. Food Res.* **59,** 1614–1628 (2015).

56. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30,** 2725–2729 (2013).

57. Lagkouvardos, I. *et al.* Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the Chlamydiae. *ISME J.* **8,** 115–125 (2014).

58. Kodama, Y., Shumway, M. & Leinonen, R. The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res.* **40,** D54–D56 (2012).

59. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).

60. Godon, J. J., Zumstein, E., Dabert, P., Habouzit, F. & Moletta, R. Molecular microbial diversity of an anaerobic digestor as determined by small-subunit rDNA sequence analysis. *Appl. Environ. Microbiol.* **63,** 2802–2813 (1997).

61. Huptas, C., Scherer, S. & Wenning, M. Optimized Illumina PCR-free library preparation for bacterial whole genome sequencing and analysis of factors influencing de novo assembly. *BMC Res. Notes* **9,** 269 (2016).

62. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).

63. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29,** 1072–1075 (2013).

64. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30,** 2068–2069 (2014).

65. Lagesen, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35,** 3100–3108 (2007).

66. Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P. & Goker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14,** 60 (2013).

67. Meier-Kolthoff, J. P., Klenk, H. P. & Goker, M. Taxonomic use of DNA G+C content and DNA–DNA hybridization in the genomic age. *Int. J. Syst. Evol. Microbiol.* **64,** 352–356 (2014).

68. Henz, S. R., Huson, D. H., Auch, A. F., Nieselt-Struwe, K. & Schuster, S. C. Whole-genome prokaryotic phylogeny. *Bioinformatics* **21,** 2329–2335 (2005).

69. Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P. & Göker, M. Highly parallelized inference of large genome-based phylogenies. *Concur. Comput. Pract. Exp.* **26,** 1715–1729 (2014).

70. Desper, R. & Gascuel, O. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J. Comput. Biol.* **9,** 687–705 (2002).

71. Farris, J. S. Estimating phylogenetic trees from distance matrices. *Am. Nat.* **106,** 645–667 (1972).

72. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421 (2009).

73. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31,** 1674–1676 (2015).

74. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15,** R46 (2014).

75. Yarza, P. *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Rev. Microbiol.* **12,** 635–645 (2014).

76. Markowitz, V. M. *et al.* IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25,** 2271–2278 (2009).

77. Qin, Q. L. *et al.* A proposed genus boundary for the prokaryotes based on genomic insights. *J. Bacteriol.* **196,** 2210–2215 (2014).

Author contributions

I.L., R.P., B.A., J.W., H.B.N., B.S. and T.C. conceived the study. I.L., R.P., B.A., A.B., I.M., D.G., M.W., J.W., H.B.N., B.S. and T.C. developed methods. I.L., R.P., B.A., B.U.F., J.P.M.-K., N.K., A.B., I.M., S.J., C.Z., S.B., D.G., T.P.N.B., J.W., F.H. and T.C. performed analysis. C.M.P., D.A.P., M.W.H., J.F.B., H.S., J.W., K.K., D.H., J.O., B.S. and T.C. made resources or materials available. I.L., R.P., B.A., B.U.F., A.B., K.K., H.B.N. and T.C. curated data. R.P., B.A., B.U.F., C.M.P., J.F.B., H.S., J.W., K.K., H.B.N., D.H., J.O., B.S. and T.C. supervised the work. I.L., R.P., J.P.M.-K., I.M., A.B. and T.C. wrote the original draft. All authors reviewed and edited the draft manuscript. T.C. coordinated the project. D.H., J.O. and T.C. were responsible for funding acquisition.

Additional information

Supplementary information is available for this paper. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.C.

Competing interests

The authors declare no competing financial interests.

Corrigendum: The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota

Ilias Lagkouvardos, Rüdiger Pukall, Birte Abt, Bärbel U. Foesel, Jan P. Meier-Kolthoff, Neeraj Kumar, Anne Bresciani, Inés Martínez, Sarah Just, Caroline Ziegler, Sandrine Brugiroux, Debora Garzetti, Mareike Wenning, Thi P. N. Bui, Jun Wang, Floor Hugenholtz, Caroline M. Plugge, Daniel A. Peterson, Mathias W. Hornef, John F. Baines, Hauke Smidt, Jens Walter, Karsten Kristiansen, Henrik B. Nielsen, Dirk Haller, Jörg Overmann, Bärbel Stecher and Thomas Clavel

In the original version of this paper, several of the bacterial genus and species names were incorrect or incompatible with formal taxonomic validation and have had to be modified. The relevant names and descriptions have been amended in all versions of the Article. In addition, Supplementary Figs 3,4 and Supplementary Tables 1,4,5,11 have been replaced.