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

Nature Microbiology **1,** 16131 (2016); published 8 August 2016; corrected 17 October 2016

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.

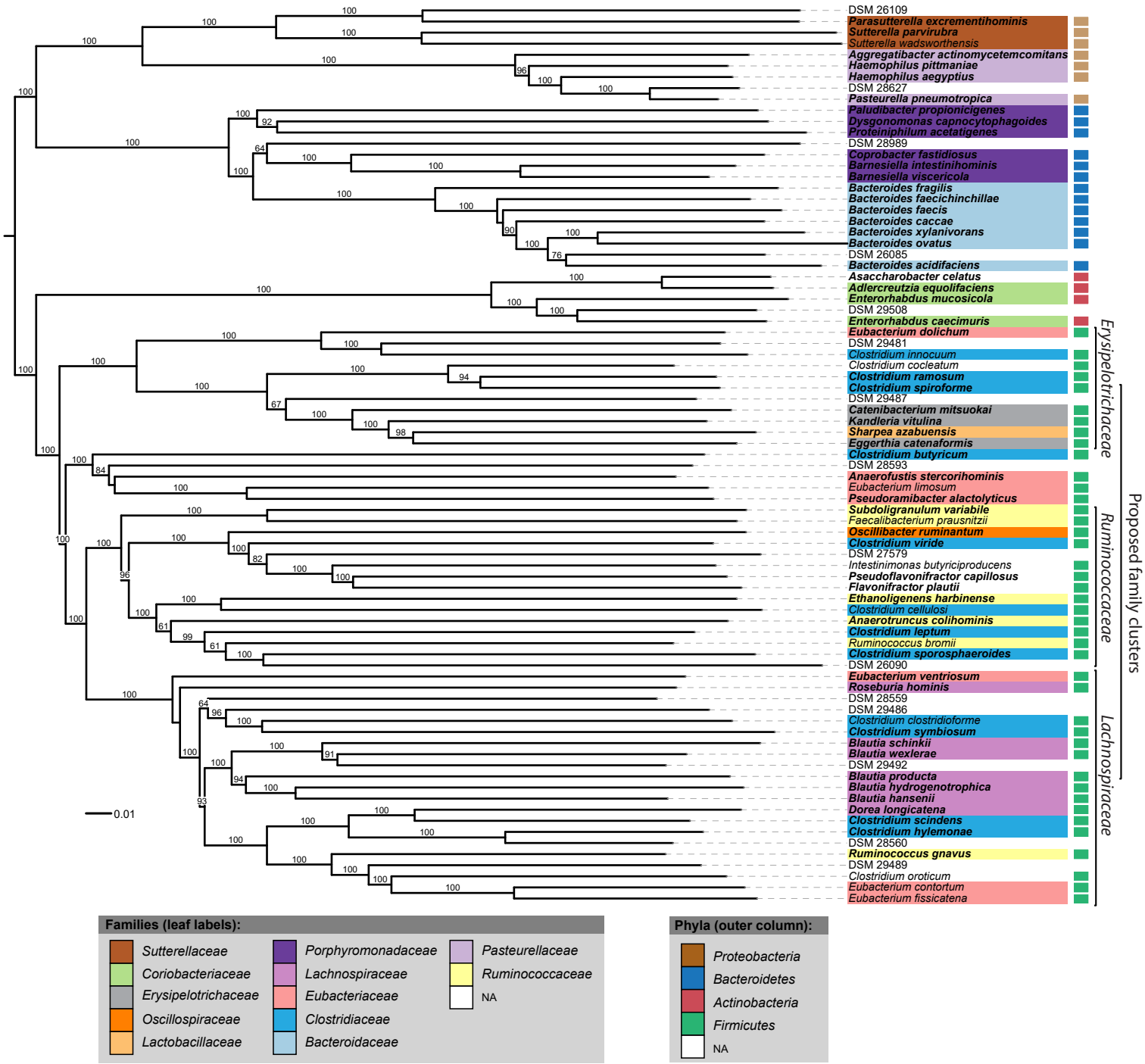
nature microbiology

SUPPLEMENTARY INFORMATION

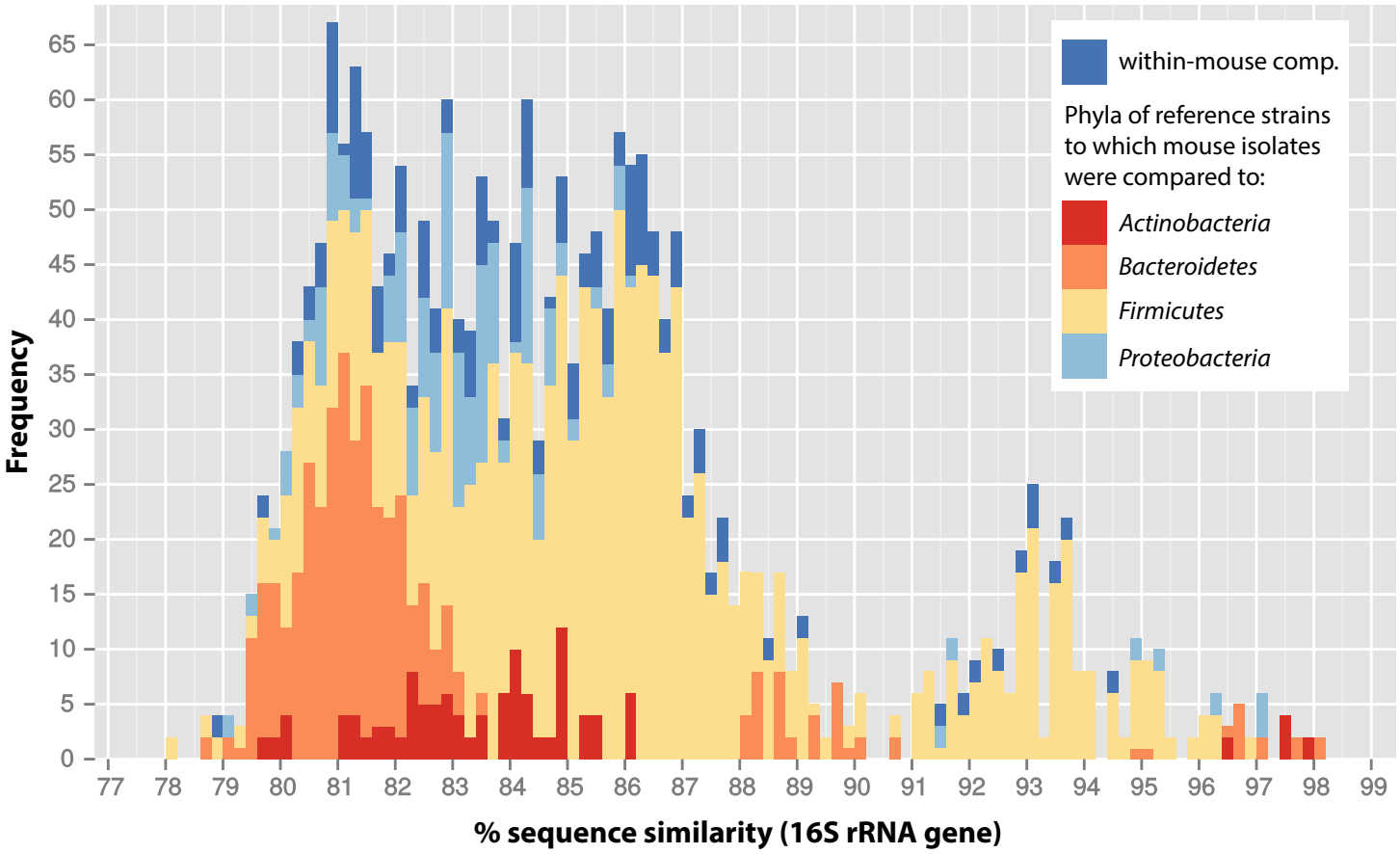
ARTICLE NUMBER: 16131 | DOI: 10.1038/NMICROBIOL.2016.131

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

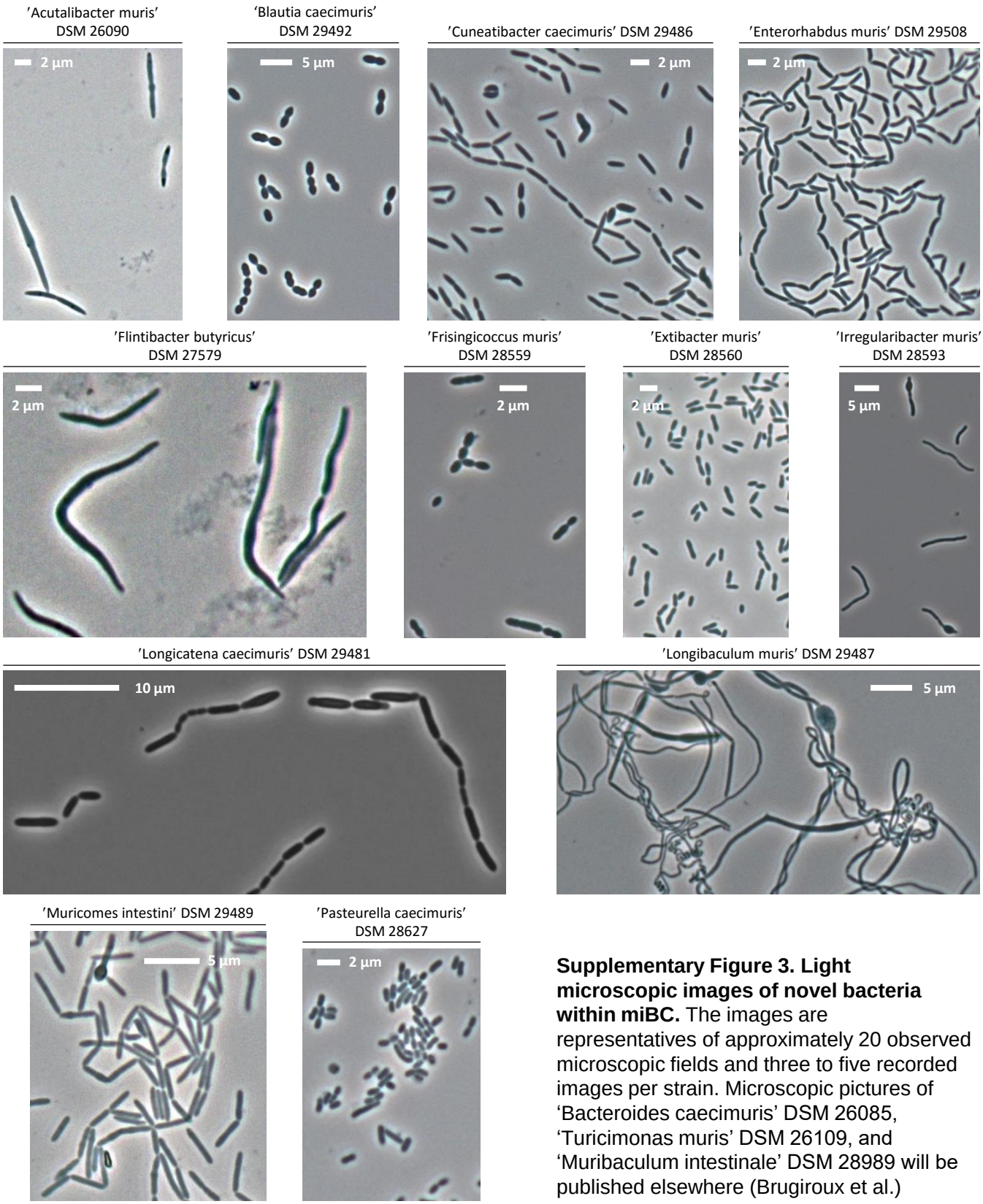
Ilias Lagkouvardos, Rüdiger Pukall, Birte Abt, Bärbel U. Foesel, Jan P. Meier-Kolthoff, Neeraj Kumar, Anne Bresciani, Inés Martínez, Sarah Just, Caroline Ziegler, Sandrine Brugiroux, Debora Garzetti, Mareike Wenning, Thi P. N. Bui, Jun Wang, Floor Hugenholtz, Caroline M. Plugge, Daniel A. Peterson, Mathias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



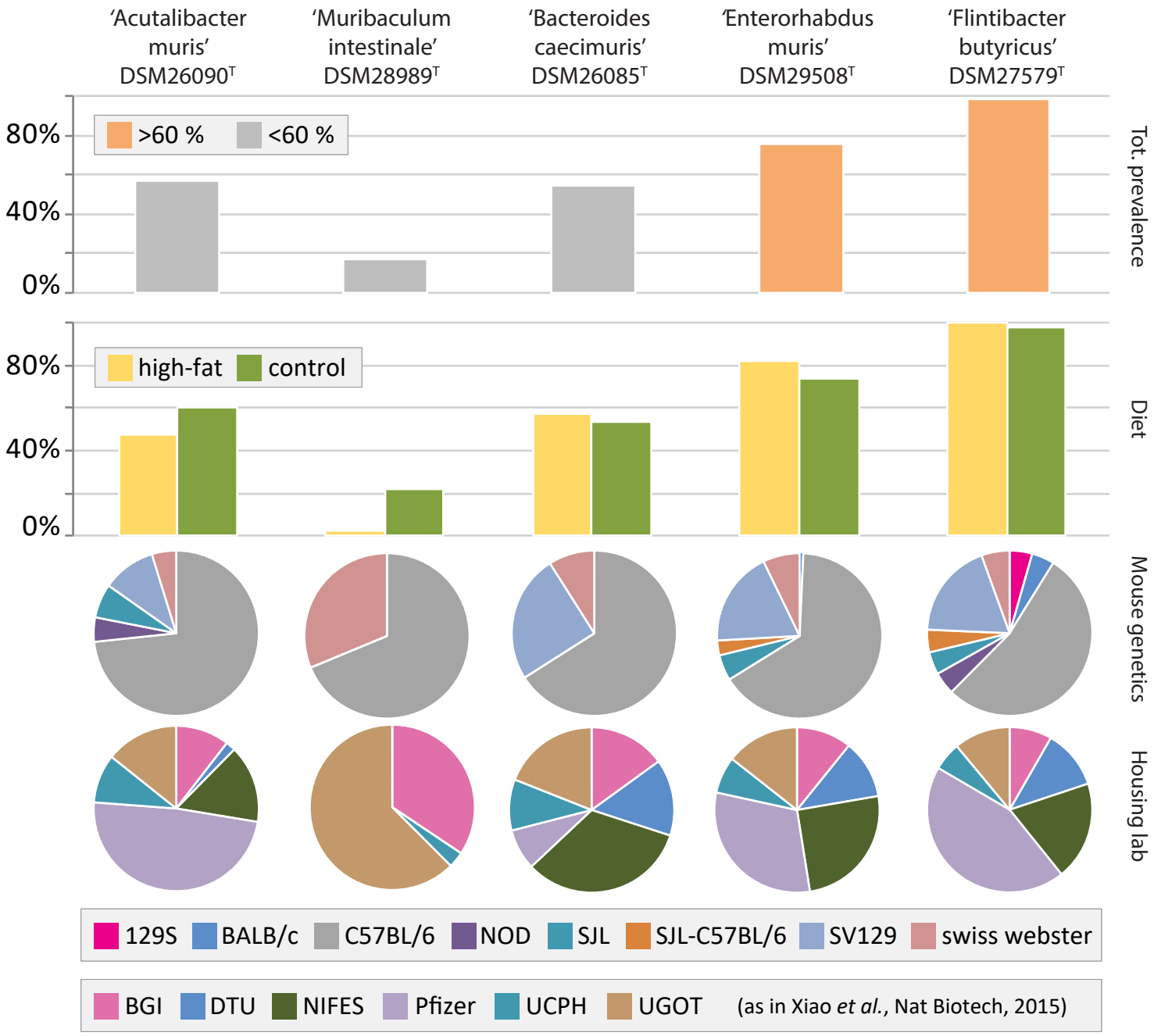
Supplementary Figure 1. Whole proteome-based phylogenomic tree of novel taxa within miBC and closely related species. Affiliation to phyla and families is as in the List of Prokaryotic names with Standing in Nomenclature[1]. Analysis was based on the latest GBDP algorithm[2] and inferred as described previously[3] and visualized using iTOL[4]. Details are given in the Methods section. Type strains are printed in bold face. Label colors represent family affiliation, whereas the outer column denotes the respective phylum.



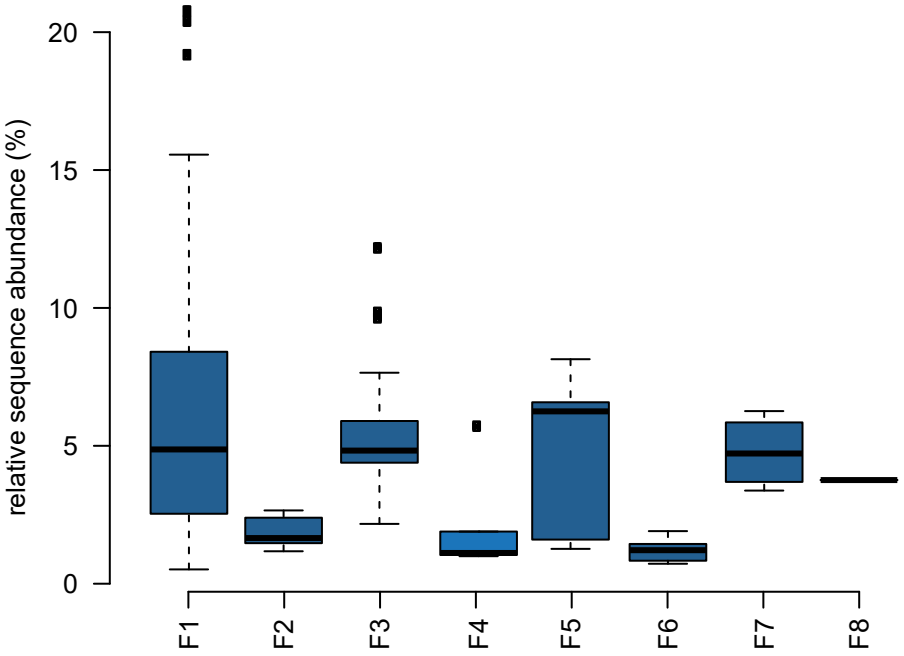
Supplementary Figure 2. Distribution of all pairwise 16S rRNA gene sequence similarities. The histograms show the number of comparisons that matched the given percentage identity displayed on the x-axis. Analysis included the 15 novel taxa isolated from the mouse gut and 62 reference strains. Bars are colored according to the phyla involved in each respective comparison. The figure was visualized using the ggplot package[6] for the R statistical framework..



Supplementary Figure 3. Light microscopic images of novel bacteria within miBC. The images are representatives of approximately 20 observed microscopic fields and three to five recorded images per strain. Microscopic pictures of 'Bacteroides caecimuris' DSM 26085, 'Turicimonas muris' DSM 26109, and 'Muribaculum intestinale' DSM 28989 will be published elsewhere (Brugiroux et al.)

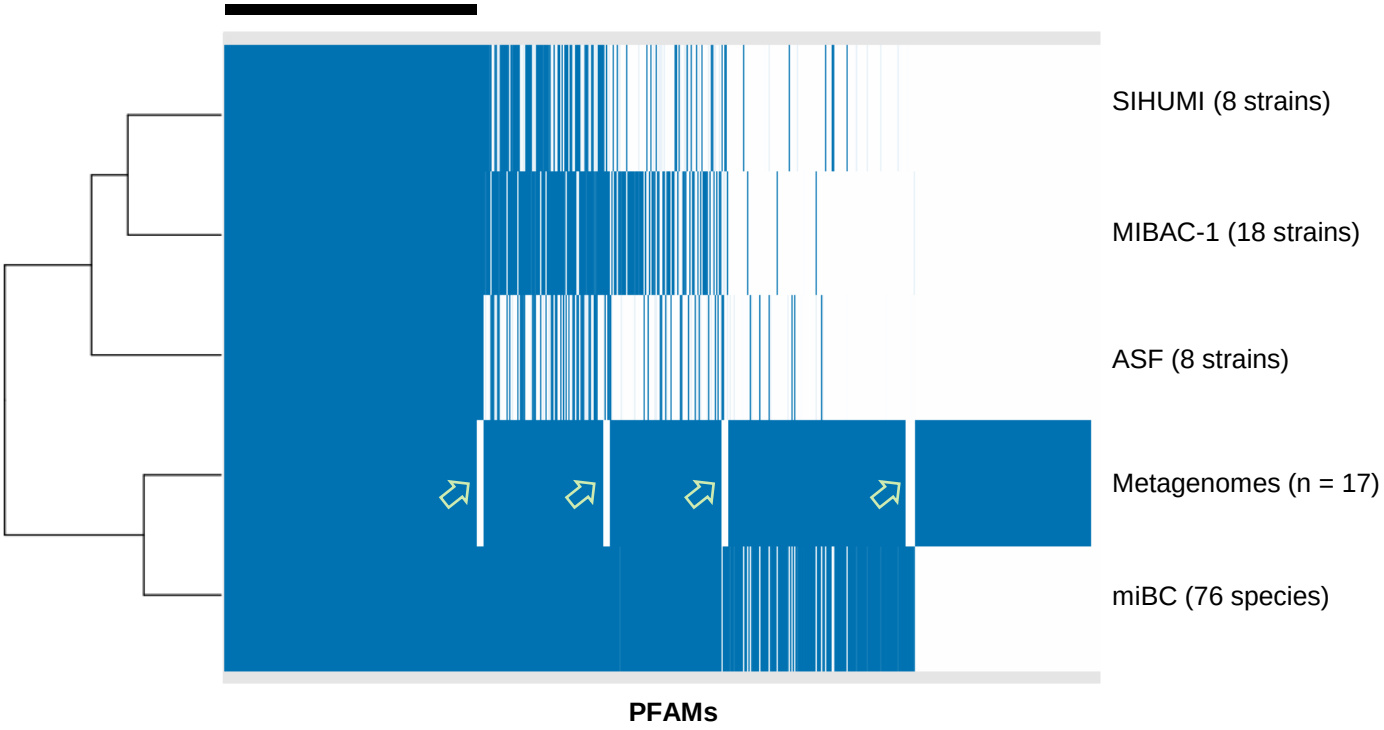


Supplementary Figure 4. Occurrence of the 5 novel bacteria in miBC that matched metagenomic species. Technical details are given in the text. The metagenomic dataset was obtained from 184 mouse fecal samples originating from 6 housing laboratories and 5 different providers[5]. Original data are provided Supplementary Table 5.



	Number of positive samples	Total
F1	37	38
F2	5	5
F3	23	23
F4	5	5
F5	5	5
F6	6	8
F7	4	4
F8	1	5

Supplementary Figure 5. Relative sequence abundance of family S24-7. Caecal samples collected from 93 mice housed in eight animal facilities (F1-8) were analyzed by high-throughput 16S rRNA gene sequencing. The total number of mice per facility and those which were positive for family S24-7 (number of positive samples) are indicated below the boxplots, which show median values and inter-quartile ranges. The p-value was obtained using ANOVA and corrected following the Benjamini-Hochberg method.



Supplementary Figure 6. Dendrogram analysis of PFAM profiles. The dendrogram was calulcated in the R programming environment using the Jaccard method. The heatmap shows PFAM in a binary presence/absence fashion (blue/white) ranked according to decreasing PFAM prevalence (from left to right) across all samples (metagenomes and cultivable genome collections). Core (present in all samples) and isolate-specific PFAM are indicated by the black line and green arrows, respectively. Corresponding PFAM lists are provided in Supplementary Table 12.

Supplementary Table 1. List of strains included in the mouse intestinal bacterial collection. This file contains (A) the list of 100 strains in miBC with corresponding information, including: species name, original strain designation, DSMZ number, Genbank accession number of 16S rRNA gene sequences, most closely related species, taxonomy, genome accession number, atmosphere requirement, and origin; (B) information on the minimal bacterial consortia analyzed in the present study.

Supplementary Table 2. Genome-derived pairwise similarities between all 16S rRNA gene sequences.

Supplementary Table 3. All pairwise digital DNA:DNA hybridization (dDDH) estimates with confidence intervals. See the Methods section for details on the calculation procedure.

Supplementary Table 4. Enzymatic profiles of novel bacterial taxa as assessed with the suitable API test strips.

Supplementary Table 5. Metagenomic species analysis. The 15 novel bacterial taxa obtained in the present study were analyzed against the mouse gene catalog[5] and searched for similarity with metagenomic species (MGS). This file contains information on (A) the number of hits to the gene catalog for each new species; (B) MGS

characteristics for the 5 new taxa with significant hits; and (C) number of fecal samples positive for each of these 5 taxa across the different mouse categories

Supplementary Table 6. High-throughput 16S rRNA gene sequence analysis: Metadata and OTU table with normalized relative sequence abundances. A total of 93 mouse caecal samples were analyzed by sequencing V5/6 regions of the 16S rRNA gene as described in detail in the text. Data were analyzed using IMNGS (www.imngs.org) with a pipeline developed in-house based on UPARSE[7]. OTUs occurring at a relative abundance >0.1 % total reads in at least one sample were kept for analysis.

Supplementary Table 7. High-throughput 16S rRNA gene sequence analysis: raw sequence counts.

Supplementary Table 8. Animal facility-specific indicator species as derived by high-throughput 16S rRNA gene sequence analysis.

Supplementary Table 9. Raw output of the integrated 16S rRNA-based studies based on IMNGS. The tool is available at www.imngs.org. The nearly full-length 16S rRNA gene sequences of all 76 species in miBC were used as queries to assess their occurrence and prevalence in the pool of 51,073 high-throughput 16S rRNA datasets

stored in the Sequence Read Archive (SRA). The present file contains the raw output of the analysis as delivered by IMNGS, including sequence counts for each miBC species and SRA samples at the level of ≥97% sequence identity.

Supplementary Table 10. List of mouse and human fecal metagenomes used in the present study to assess coverage by the collection of cultured strains and by minimal bacteriomes.

Supplementary Table 11. Composition of the random sets of miBC-derived strains used to test the relevance of the data-driven selection MIBAC-1.

Supplementary Table 12. Core and miBC-specific PFAM features.

Supplementary Sequence File. This file contains the partial nucleotide sequences of all operational taxonomic units from the high-throughput 16S rRNA gene dataset generated in the present study.

Supplementary references

1. Euzeby JP. List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. *International journal of systematic bacteriology* **47**, 590-592 (1997).

2. Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC bioinformatics* **14**, 60 (2013).

3. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Highly parallelized inference of large genome-based phylogenies. In: *Concurrency Computation Practice and Experience* (ed^(eds) (2014).

4. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic acids research* **39**, W475-478 (2011).

5. Xiao L*, et al.* A catalog of the mouse gut metagenome. *Nature biotechnology* **33**, 1103-1108 (2015).

6. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer (2009).

7. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods* **10**, 996-998 (2013).