Closed and High-Quality Bacterial Genome Sequences of the Oligo-Mouse-Microbiota Community

Quentin Lamy-Besnier,a,b,c Romain Koszul,b Laurent Debarbieux,a Martial Marboutyb

aBacteriophage, Bacterium, Host Laboratory, Department of Microbiology, Institut Pasteur, Paris, France
bUnité Régulation Spatiale des Génomes, Institut Pasteur, CNRS, UMR 3525, Paris, France
cUniversité Paris Descartes, Paris, France

ABSTRACT The Oligo-Mouse-Microbiota (OMM12) gnotobiotic murine model is an increasingly popular model in microbiota studies. However, following Illumina and PacBio sequencing, the genomes of the 12 strains could not be closed. Here, we used genomic chromosome conformation capture (Hi-C) data to reorganize, close, and improve the quality of these 12 genomes.

The Oligo-Mouse-Microbiota (OMM12) is a murine bacterial synthetic community (12 strains) introduced in axenic mice. The resulting gnotobiotic murine model has been increasingly used for diverse gut microbiota studies across several animal facilities (1–4). This bacterial consortium comprises members from the 5 major phyla naturally present in mice microbiota and reconstitutes its main functionalities, such as metabolism and colonization resistance against pathogens (5).

Despite two rounds of sequencing using PacBio and Illumina technologies (6, 7), one-half of the genomes currently accessible remain made of 2 to 20 contigs (Table 1). In particular, the genome of Bacteroides caecimuris, the most abundant bacterium of the OMM12 consortium, consists of 19 contigs (5). Improving scaffolding would facilitate a number of downstream analyses, such as prophage prediction.

We used genomic chromosome conformation capture (Hi-C), a technique that uses chemical fixation (formaldehyde) to in vivo cross-link nearby DNA sequences, digestion with restriction enzymes, DNA extremities refilling with biotinylated nucleotides, and finally, proximity ligation, DNA extraction, and deep sequencing. The relative ligation frequencies between nonadjacent DNA segments reflect the average three-dimensional (3D) organization of the genome of interest (8) and also, because of the polymer nature of DNA, the relative distance separating them along the chromosome. The latter property has therefore been exploited to bridge scaffolding gaps resulting, for instance, from the presence of repeated elements (9). Hi-C scaffolding is now commonly used in sequencing projects of large eukaryotic genomes but applies to bacteria as well (10, 11).

A Hi-C metagenomic protocol for mammalian gut samples developed in our laboratory (12) was applied to fecal pellets from the OMM12 mice bred at Institut Pasteur (protocol 20.173 approved by the veterinary staff and under authorization APAFIS 26874 by the national ethics committee). Libraries were prepared using streptavidin beads to capture biotinylated ligation junction as described previously (13) and sequenced with an Illumina NextSeq 550 system to generate a total of 101,182,905 reads of 35 bp. The quality of the reads was first verified with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Then, for each bacterium, the python library hicstuff (https://github.com/koszullab/hicstuff) was used to generate Hi-C contact maps using the most recent OMM12 genomes (6). Contigs were reorganized (i.e., scaffolded) based on their relative contact frequencies (14). When necessary, preexisting contigs were split and rearranged to better fit the expected 3D structure, based on the contact data. In order for users to detect junctions...
between contigs, contigs were separated by $10 \times N$ in the final fasta file. We successfully closed the genomes of the 12 strains except for *Flavonifractor plautii*, where the position of a 70,760-bp region could not be assigned into the main scaffold. The genome of *F. plautii* therefore remains split in 2 scaffolds. An automatic annotation was then performed using RAST (15). The access to a single scaffold of high quality will greatly improve the accuracy of genomic analyses performed by the community of users of the OMM12 gnotobiotic mice.

**Data availability.** The 12 reassembled and closed genome sequences of the OMM$^{12}$ bacteria as well as the FastQ reads have been deposited under the accession numbers provided in Table 1 under the BioProject number PRJNA680355.

**ACKNOWLEDGMENTS**

We thank Cyril Matthey Doret and Agnès Thierry for assistance during computational and experimental work, respectively.

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. This research was supported by funding to L.D. and M.M. from ANR-20-CE92-0048 and to R.K. from the European Research Council under the Horizon 2020 Program (ERC grant agreement 771813) and from JPI-EC-AMR STARCS ANR-16-JPEC-0003-05. Q.L.-B. is funded by École Doctorale FIRE-Program Bettencourt.

**REFERENCES**

1. Bolsge S, Basic M, Smoczek A, Buettner M, Eberl C, Ahrens D, Oudem KA, Stecher B, Bleich A. 2019. Composition of the intestinal microbiota determines the outcome of virus-triggered colitis in mice. Front Immunol 10:1708. https://doi.org/10.3389/fimmu.2019.01708.

2. Eberl C, Ring D, Münch PC, Beutler M, Basic M, Slack EC, Schwarzer M, Srutkova D, Lange A, Frick JS, Bleich A, Stecher B. 2019. Reproducible colonization of germ-free mice with the Oligo-Mouse-Microbiota in different animal facilities. Front Microbiol 10:2999. https://doi.org/10.3389/fmicb.2019.02999.

3. Lorentzo M, Chaffriongoen L, Lamy-Besnier Q, Lamy-Besnier M, Campagne P, Eberl C, Bérard M, Stecher B, Debarbieux L, De Sordi L. 2020. The spatial heterogeneity of the gut limits predation and fosters coexistence of bacteria and bacteriophages. Cell Host Microbe 28:390–401.e395. https://doi.org/10.1016/j.chom.2020.06.002.

4. Studer N, Deshamais L, Beutler M, Brugiroux S, Terrazos MA, Menin L, Schürch CM, McCoy KD, Kuehne SA, Minton NP, Stecher B, Bernier-Latmani R, Hapflmeier S. 2016. Functional intestinal bile acid 7α-dehydroxylation by Clostridium scindens associated with protection from Clostridium difficile infection in a gnotobiotic mouse model. Front Cell Infect Microbiol 6:191. https://doi.org/10.3389/fcimb.2016.00191.

5. Brugiroux S, Beutler M, Pfann C, Garzetti D, Ruscheweyh HJ, Ring D, Diehl M, Herp S, Lötscher Y, Hussain S, Bunk B, Pukall R, Huson DH, Münch PC, McHardy AC, McCoy KD, Macpherson AJ, Loy A, Clavel T, Berry D, Stecher B. 2016. Genome-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium. Nat Microbiol 2:16215. https://doi.org/10.1038/nmicrobiol.2016.215.

6. Garzetti D, Brugiroux S, Bunk B, Pukall R, McCoy KD, Macpherson AJ, Stecher B. 2017. High-quality whole-genome sequences of the Oligo-Mouse-Microbiota bacterial community. Genome Announc 5:e00758-17. https://doi.org/10.1128/genomeA.00758-17.

7. Uchimura Y, Wyss M, Brugiroux S, Limenitakis JP, Stecher B, McCoy KD, Macpherson AJ. 2016. Complete genome sequences of 12 species of stable defined moderately diverse mouse microbiota 2. Genome Announc 4:e00951-16. https://doi.org/10.1128/genomeA.00951-16.

8. Dekker J, Rippe K, Dekker M, Klekner N. 2002. Capturing chromosome conformation. Science 295:1306–1311. https://doi.org/10.1126/science.1067799.

9. Plot JF, Marie-Nelly H, Koszul R. 2015. Contact genomics: scaffolding and phasing (meta)genomes using chromosome 3D physical signatures. FEBS Lett 589:2966–2974. https://doi.org/10.1016/j.febslet.2015.04.034.

10. Baudry L, Guiglielmoni N, Marie-Nelly H, Cormier A, Marbouty M, Avia K, Mie YL, Godfroy O, Sterck L, Cock JM, Zimmer C, Coelho SM, Koszul R. 2020. instaGRAAL: chromosome-level quality scaffolding of genomes

---

**TABLE 1** Description and accession numbers of the OMM$^{12}$ genomes$^a$

| Strain                              | Genome size (bp) | Previous no. of contigs$^b$ | No. of genes | DSM no. | Accession no. | Reads (%)$^c$ |
|------------------------------------|------------------|-----------------------------|-------------|---------|---------------|---------------|
| Akkermansia muciniphila YL44        | 2,737,357        | 1                           | 2,717       | 26127   | CP065322      | 2.8           |
| Acutalbacter muris KB18             | 3,802,913        | 1                           | 4,271       | 26090   | CP065321      | 0.002         |
| Bifdobacterium animalis YL2        | 2,021,936        | 2                           | 1,740       | 26074   | CP065311      | 0.0007        |
| Bacteroides cacicimuris I48        | 4,800,606        | 19                          | 4,322       | 26085   | CP065319      | 55.8          |
| Blautia cocoides YLS8               | 5,128,582        | 1                           | 5,351       | 26115   | CP065312      | 20.5          |
| Enteroclostridium formi YLS32       | 7,157,610        | 16                          | 7,955       | 26114   | CP065314      | 4.9           |
| Clostridium innocuum I46           | 4,452,146        | 1                           | 4,966       | 26113   | CP065320      | 5.6           |
| Enterococcus faecalis KB1           | 3,025,655        | 1                           | 2,919       | 32036   | CP065317      | 0.1           |
| Flavonifractor plautii YL31         | 3,813,715        | 5                           | 4,049       | 26117   | CP065315      | 0.38          |
| Limosilactobacillus reuteri I49     | 2,063,624        | 3                           | 1,958       | 32035   | CP065318      | 0.99          |
| Muribaculum intestinale YL27        | 3,307,069        | 20                          | 2,957       | 28989   | CP065316      | 4.5           |
| Turicimonas muri YLS45              | 2,887,949        | 20                          | 2,753       | 26109   | CP065313      | 0.089         |

$^a$Raw data can be accessed at SRX9907181.

$^b$Garzetti et al. (6).

$^c$Unmapped reads = 4.34 %.
11. Marbouty M, Baudry L, Cournac A, Koszul R. 2017. Scaffolding bacterial genomes and probing host-virus interactions in gut microbiome by proximity ligation (chromosome capture) assay. Sci Adv 3:e1602105. https://doi.org/10.1126/sciadv.1602105.

12. Marbouty M, Thierry A, Millot GA, Koszul R. 2021. MetaHiC phage-bacteria infection network reveals active cycling phages of the healthy human gut. Elife 10:e60608. https://doi.org/10.7554/eLife.60608.

13. Moreau P, Cournac A, Palumbo GA, Marbouty M, Mortaza S, Thierry A, Cairo S, Lavigne M, Koszul R, Neuveut C. 2018. Tridimensional infiltration of DNA viruses into the host genome shows preferential contact with active chromatin. Nat Commun 9:4268. https://doi.org/10.1038/s41467-018-06739-4.

14. Marie-Nelly H, Marbouty M, Cournac A, Flot JF, Liti G, Parodi DP, Syan S, Guillén N, Margeot A, Zimmer C, Koszul R. 2014. High-quality genome (re)assembly using chromosomal contact data. Nat Commun 5:5695. https://doi.org/10.1038/ncomms6695.

15. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil K, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.