



# Metagenomic Hi-C of a Healthy Human Fecal Microbiome Transplant Donor

 Matthew Z. DeMaere,<sup>a</sup>  Michael Y. Z. Liu,<sup>a</sup>  Enmoore Lin,<sup>b</sup>  Steven P. Djordjevic,<sup>a</sup>  Ian G. Charles,<sup>c</sup>  Paul Worden,<sup>a</sup>  
 Catherine M. Burke,<sup>a</sup>  Leigh G. Monahan,<sup>a</sup>  Melissa Gardiner,<sup>a</sup>  Thomas J. Borody,<sup>b</sup>  Aaron E. Darling<sup>a</sup>

<sup>a</sup>The iThree Institute, University of Technology Sydney, Sydney, NSW, Australia

<sup>b</sup>The Centre for Digestive Diseases, Five Dock, NSW, Australia

<sup>c</sup>The Quadram Institute, Norwich, United Kingdom

**ABSTRACT** We report the availability of a high-quality metagenomic Hi-C data set generated from a fecal sample taken from a healthy fecal microbiome transplant donor subject. We report on basic features of the data to evaluate their quality.

Metagenomic Hi-C is a recently emerged technique that enables the physical proximity of DNA sequences in a sample to be assayed (1–3). This type of three-dimensional (3D) spatial information about sequences has historically been missing from metagenomic shotgun sequencing data sets (4) and has led to the development of extensive, elaborate, and often failure-prone computational methods that attempt to reconstruct genomic content using other signals in the data (5).

We generated metagenomic Hi-C data for a human fecal sample as part of a larger technology evaluation program that also evaluated metagenomic double-digest restriction-site-associated DNA (ddRADseq) (6) and low-cost, low-bias Illumina shotgun library preparation protocols (7). The sample was obtained from a member of the healthy fecal microbiome transplant (FMT) donor pool used at the Centre for Digestive Diseases (Five Dock, NSW, Australia) in 2014. Briefly, the sample was collected fresh, stored frozen at  $-80^{\circ}\text{C}$  for 1 year, and then thawed, cross-linked with 1% formalin for 1 hour, quenched with 125 mM glycine for 30 min, and stored frozen again prior to shipping to Phase Genomics LLC (Seattle, WA, USA) for Hi-C library preparation using an established protocol (8) and sequencing on an Illumina NextSeq 500 instrument. Ethical approval for this study was obtained from the University of Technology Sydney Human Research Ethics Committee (UTS HREC reference number 2014000448).

Sequencing produced 20.1 million 150-bp shotgun read pairs (totaling 5.8 Gbp) and 71.6 million 80-bp Hi-C read pairs (totaling 11.4 Gbp) composed of two technical replicates. The fraction of read pairs containing proximity ligation junctions (Hi-C read pairs) was estimated using the recently developed qc3C tool v0.2.6.6 (9) (default parameters used). qc3C has two methods for estimating the fraction of Hi-C read pairs in the data, (i) by mapping reads to a metagenome assembly and (ii) using an assembly-free technique based on k-mer counts. Using the mapping-based technique, the fraction of Hi-C read pairs was estimated to be within the range of 0.36 to 0.67%. To put this in context, the same estimate for another recently published metagenomic Hi-C data set (8) was 1.38 to 2.38%.

Cleanup of the shotgun and Hi-C read sets was performed with fastp v0.20.0 (10) using default options, with the exception that overlapping shotgun pairs were merged. A metagenomic assembly was generated from the cleaned shotgun reads using SPAdes v3.13.1 (11) (command-line “–meta”) and comprised 181,642 contigs and a total size of 196,582,935 bp ( $N_{50}$ , 2,965 bp). The scaffolds from this assembly were used in conjunction with the metagenomic Hi-C data from the same sample to reconstruct metagenome-assembled genomes (MAGs) using bin3C v0.3.3 (Fig. 1) (10)

**Citation** DeMaere MZ, Liu MYZ, Lin E, Djordjevic SP, Charles IG, Worden P, Burke CM, Monahan LG, Gardiner M, Borody TJ, Darling AE. 2020. Metagenomic Hi-C of a healthy human fecal microbiome transplant donor. *Microbiol Resour Announc* 9:e01523-19. <https://doi.org/10.1128/MRA.01523-19>.

**Editor** Frank J. Stewart, Georgia Institute of Technology

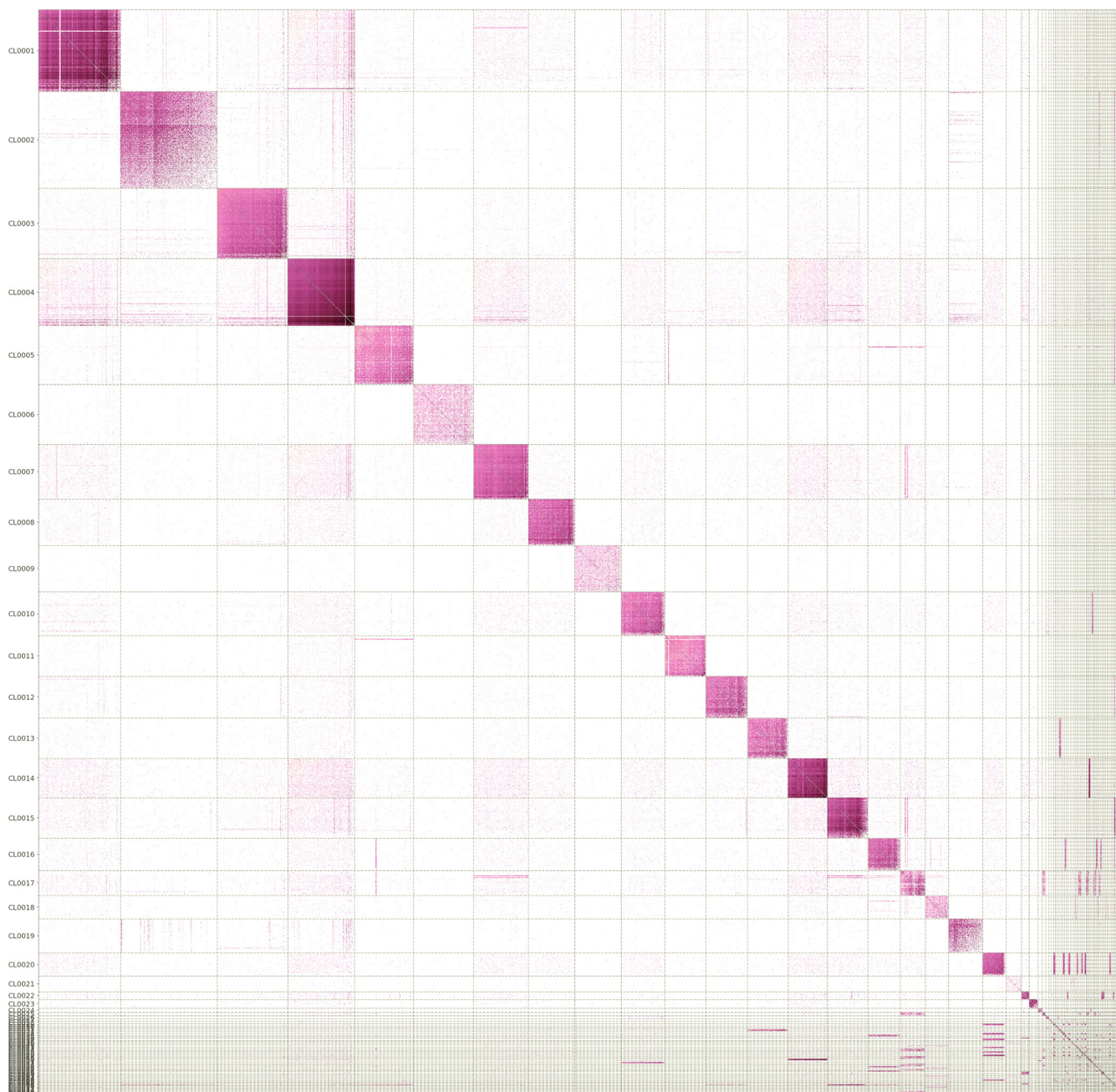
**Copyright** © 2020 DeMaere et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Aaron E. Darling, [aaron.darling@uts.edu.au](mailto:aaron.darling@uts.edu.au).

**Received** 12 December 2019

**Accepted** 6 January 2020

**Published** 6 February 2020



**FIG 1** Hi-C contact map generated using bin3C from the metagenome assembly, ordered by decreasing cluster extent. Rows and columns correspond to contigs binned in windows of no more than 5 kbp. The log-scaled intensity of each cell represents the normalized interaction strength derived from the observed number of Hi-C read pairs that link the pair of loci. Blocks of color along the diagonal line correspond to groups of contigs that are in physical contact in the sample, typically because they are in the same chromosome or cell. Light dashed lines indicate the cluster boundaries determined with bin3C; the large bins correspond to MAGs.

(command-line “cluster –min-signal 4 –n-iter 20 –seed 12345 –assembler spades”). The analysis yielded 15 genomes that were estimated to be  $\geq 50\%$  complete with  $\leq 5\%$  contamination as determined via CheckM v1.0.18 (12) (default parameters used).

The data we have released may be useful for a range of analyses, including the study of host-virus and host-plasmid associations, as well as the study of the 3D chromosome structure of dominant members of the human gut microbiome.

**Data availability.** Metagenomic Hi-C data are available under the Sequence Read Archive accession numbers [SRR7427737](https://www.ncbi.nlm.nih.gov/sra/SRR7427737) and [SRR10566997](https://www.ncbi.nlm.nih.gov/sra/SRR10566997). The corresponding shotgun

library is available under accession number [SRR5298275](https://zenodo.org/record/3598124). The metagenomic assembly and derived metagenome-assembled genomes produced using bin3C are available from zenodo (<https://zenodo.org/record/3598124>).

## ACKNOWLEDGMENTS

This work was funded in part by the Australian Research Council's Discovery scheme under ARC Linkage project LP150100912 and ARC Discovery project DP180101506. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Thomas J. Borody owns and operates the Centre for Digestive Diseases, a privately held medical clinic.

## REFERENCES

1. Beitel CW, Froenicke L, Lang JM, Korf IF, Michelmore RW, Eisen JA, Darling AE. 2014. Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. *PeerJ* 2:e415. <https://doi.org/10.7717/peerj.415>.
2. Burton JN, Liachko I, Dunham MJ, Shendure J. 2014. Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. *G3 (Bethesda)* 4:1339–1346. <https://doi.org/10.1534/g3.114.011825>.
3. Marbouty M, Cournac A, Flot J-F, Marie-Nelly H, Mozziconacci J, Koszul R. 2014. Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *Elife* 3:e03318. <https://doi.org/10.7554/eLife.03318>.
4. Liu M, Darling A. 2015. Metagenomic chromosome conformation capture (3C): techniques, applications, and challenges. *F1000Res* 4:1377. <https://doi.org/10.12688/f1000research.7281.1>.
5. Szczyrba A, Hofmann P, Belmann P, Koslicki D, Janssen S, Dröge J, Gregor I, Majda S, Fiedler J, Dahms E, Bremges A, Fritz A, Garrido-Oter R, Jørgensen TS, Shapiro N, Blood PD, Gurevich A, Bai Y, Turaev D, DeMaere MZ, Chikhi R, Nagarajan N, Quince C, Meyer F, Balvočiūtė M, Hansen LH, Sørensen SJ, Chia BKH, Denis B, Froula JL, Wang Z, Egan R, Don Kang D, Cook JJ, Deltel C, Beckstette M, Lemaître C, Peterlongo P, Rizk G, Lavenier D, Wu Y-W, Singer SW, Jain C, Strous M, Klingenberg H, Meinicke P, Barton MD, Lingner T, Lin H-H, Liao Y-C, Silva GGZ, Cuevas DA, Edwards RA, Saha S, Piro VC, Renard BY, Pop M, Klenk H-P, Göker M, Kyrpides NC, Woyke T, Vorholt JA, Schulze-Lefert P, Rubin EM, Darling AE, Rattei T, McHardy AC. 2017. Critical assessment of metagenome interpretation—a benchmark of metagenomics software. *Nat Methods* 14:1063–1071. <https://doi.org/10.1038/nmeth.4458>.
6. Liu MY, Worden P, Monahan LG, DeMaere MZ, Burke CM, Djordjevic SP, Charles IG, Darling AE. 2017. Evaluation of ddRADseq for reduced representation metagenome sequencing. *PeerJ* 5:e3837. <https://doi.org/10.7717/peerj.3837>.
7. Gaio D, To J, Liu M, Monahan L, Anantanawat K, Darling AE. 2019. Hackflex: low cost Illumina sequencing library construction for high sample counts. *bioRxiv*. <https://www.biorxiv.org/content/10.1101/779215v1>.
8. Press MO, Wiser AH, Kronenberg ZN, Langford KW, Shakya M, Lo C-C, Mueller KA, Sullivan ST, Chain PSG, Liachko I. 2017. Hi-C deconvolution of a human gut microbiome yields high-quality draft genomes and reveals plasmid-genome interactions. *bioRxiv*. <https://www.biorxiv.org/content/10.1101/198713v1>.
9. DeMaere MZ, Darling AE. 2019. qc3C. Github. <https://github.com/cerebis/qc3C>.
10. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
11. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27:824–834. <https://doi.org/10.1101/gr.213959.116>.
12. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.