

UC San Diego

UC San Diego Previously Published Works

Title

Metagenome-Assembled Genomes from Murine Fecal Microbiomes Dominated by Uncharacterized Bacteria

Permalink

<https://escholarship.org/uc/item/9gw7h1wb>

Journal

Microbiology Resource Announcements, 12(3)

ISSN

2576-098X

Authors

Rastegar, Kiarash

Kelley, Scott T

Thackray, Varykina G

Publication Date

2023-03-16

DOI

10.1128/mra.01162-22

Peer reviewed



# Metagenome-Assembled Genomes from Murine Fecal Microbiomes Dominated by Uncharacterized Bacteria

Kiarash Rastegar,<sup>a</sup>  Scott T. Kelley,<sup>a,b</sup> Varykina G. Thackray<sup>c</sup>

<sup>a</sup>Bioinformatics and Medical Informatics Program, San Diego State University, San Diego, California, USA

<sup>b</sup>Department of Biology, San Diego State University, San Diego, California, USA

<sup>c</sup>Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Diego, La Jolla, California, USA

**ABSTRACT** The laboratory mouse gut microbiome has been extensively studied, but our understanding of its diversity remains incomplete. We report the assembly of 51 draft metagenome-assembled genomes (MAGs) from murine fecal samples dominated by uncharacterized bacteria. These MAGs add to our understanding of gut microbial diversity in this critical model organism.

Murine models are an important tool for studying the role of the microbiome in different aspects of mammalian physiology and disease (1–3). Our research has explored the relationship of the gut microbiome to sex steroid hormones, with a particular focus on the hyperandrogenism that occurs in polycystic ovary syndrome (4–6). Here, we report on metagenome-assembled genome (MAG) assemblies from two murine fecal metagenomes dominated by uncharacterized bacteria. Three-week-old C57BL/6N female mice were purchased from Envigo and housed in a vivarium at the University of California, San Diego, with an automatic 12-h light/12-h darkness cycle (light period from 6:00 a.m. to 6:00 p.m.). Mice were given *ad libitum* access to water and food (Teklad global 18% protein extruded diet; Envigo). Fecal samples were collected in the morning at 7 weeks of age. DNA was extracted from ~100-mg frozen murine fecal samples using the DNeasy PowerSoil kit (Qiagen) according to the manufacturer's instructions, and extracted DNA was stored at –80°C. Libraries were prepared and sequenced on an Illumina NovaSeq system at the University of California, San Diego Institute for Genomic Medicine Genomics Center. Specifically, 100 ng of genomic DNA was sonicated using an E220 focused ultrasonicator (Covaris) to produce 600-bp fragments, which were purified using Agencourt AMPure XP beads (Beckman Coulter). A KAPA HyperPrep kit (Kapa Biosystems) was used to prepare Illumina libraries following the manufacturer's instructions. Libraries were quality checked for their size and concentration with electrophoresis using a high-sensitivity D1000 kit on a 2200 TapeStation (Agilent). Sequencing generated a total of 130.58 million reads combined for the two samples (60.64 million reads and 69.94 million reads), and the resulting paired-end metagenomic fastq files were trimmed and quality controlled using fastp v0.12.4 (7). The fastp program trimmed adapters, filtered out low-quality reads, excised poor-quality bases from the 5' and 3' ends, and produced forward and reverse sequences 100 nucleotides in length for further analysis. Assembly of the metagenomes was performed with metaSPAdes v3.15.3 (8). The contigs were uploaded to KBase (9) and further analyzed with the MaxBin 2.0 v2.2.4 (10) binning algorithm. These bins were then further refined using DAS Tool v1.1.2 (11), which collapsed the number of bins from 116 to 51. The DAS Tool bins were then annotated using the microbial genome annotator tool RASTtk v1.0.73 (12). Once the annotations were complete, we used GTDB-Tk v1.7.0 (13) to taxonomically classify the bins. To check the composition of the genome, we used CheckM v1.0.18 (14), which shows the quality of the bins and how many single-copy genes they contain. To determine which bins were the most abundant in the metagenomes, we used Salmon v1.8.0 (15) to measure the abundance of the contigs in each bin by mapping the reads

**Editor** Vanja Klepac-Ceraj, Wellesley College

**Copyright** © 2023 Rastegar et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Scott T. Kelley, skelley@sdsu.edu.

The authors declare no conflict of interest.

**Received** 30 November 2022

**Accepted** 20 January 2023

**Published** 13 February 2023

**TABLE 1** Accession numbers, assembly statistics, and taxonomic assignments for MAGs

GenBank accession no. <sup>a</sup>	GenBank assembly accession no. for FastANI reference <sup>b</sup>	FastANI ANI (%) <sup>c</sup>	Genome size (bp)	GC content (%)	Avg coverage (%)	Completeness (%) <sup>d</sup>	Contamination (%) <sup>e</sup>	No. of contigs <sup>f</sup>	N <sub>50</sub> (bp) <sup>g</sup>	NCBI organism name and bin number <sup>h</sup>
JAPKV010000000			1,103,159	49.41	0.24	60.34	2.93	96	67,037	<i>Clostridia</i> bacterium, bin 001
JAPKV001000000	GCF_000012845.1	97.32	4,373,110	54.31	0.54	94.10	1.67	131	62,302	<i>Parabacteroides distans</i> , bin 002
JAPKV000000000	GCF_003762875.1	96.16	3,022,371	45.04	0.67	95.74	27.8	122	51,464	<i>Duncaniella</i> sp., bin 003
JAPKV000000000			4,108,057	55.39	1.03	75.35	10.3	219	43,157	<i>Dyosmobacter</i> sp., bin 004
JAPKV000000000			1,786,207	43.74	1.68	95.16	<0.01	91	41,188	<i>Clostridia</i> bacterium, bin 005
JAPKV000000000			2,212,258	30.52	1.20	91.57	8.43	158	21,184	<i>Bacillus</i> bacterium, bin 006
JAPKV010000000	GCF_000487995.1	99.94	2,331,977	31.48	5.77	98.28	2.59	21	127,566	<i>Mucispirillum schaedleri</i> , bin 007
JAPKV001000000	GCA_009774395.1	98.61	5,946,436	39.21	2.75	95.15	8.35	81	103,071	<i>Lachnospiraceae</i> bacterium, bin 008
JAPKV010000000			2,068,202	51.65	0.42	99.04	0.51	126	94,090	<i>Rikenellaceae</i> bacterium, bin 009
JAPKV010000000			3,247,007	48.90	1.03	97.65	1.34	33	85,270	<i>Oscillospiraceae</i> bacterium, bin 010
JAPKV010000000			2,298,632	41.29	0.32	88.97	12.3	56	77,641	<i>Clostridia</i> bacterium, bin 011
JAPKV010000000			1,825,488	43.67	0.99	95.70	2.15	62	197,687	<i>Alphaproteobacteria</i> bacterium, bin 012
JAPKV010000000			2,898,828	47.09	0.49	77.29	1.68	102	45,504	<i>Lachnospiraceae</i> bacterium, bin 013
JAPKV010000000			3,584,479	45.43	1.51	96.26	4.70	103	59,311	<i>Kineothrix</i> sp., bin 014
JAPKV010000000	GCF_000403495.2	98.42	4,637,001	40.24	2.89	97.13	3.55	142	25,930	<i>Lachnospiraceae</i> bacterium, bin 015
JAPKV010000000			3,070,752	48.32	6.20	96.43	0.48	74	101,673	<i>Bacteroidales</i> bacterium, bin 016
JAPKV010000000	GCA_009774765.1	99.96	2,706,227	46.62	2.02	98.17	1.19	194	43,433	<i>Bacteroidales</i> bacterium, bin 017
JAPKV010000000	GCA_001689535.1	98.61	1,836,513	54.39	1.25	65.28	3.84	432	48,531	<i>Paramuribaculum</i> sp., bin 018
JAPKV010000000			4,495,251	48.39	4.54	97.27	3.62	96	52,116	<i>Lachnospiraceae</i> bacterium, bin 019
JAPKV010000000			1,581,772	56.66	2.56	74.53	0.75	98	48,943	<i>Muribaculaceae</i> bacterium, bin 020
JAPKV010000000	GCA_002490725.1	99.39	1,954,743	55.72	1.50	87.17	0.38	580	27,976	<i>Paramuribaculum intestinale</i> , bin 021
JAPKV010000000	GCF_003024925.1	99.23	3,068,355	44.52	0.42	88.96	3.09	29	90,657	<i>Lachnospiraceae</i> bacterium, bin 022
JAPKV010000000	GCF_003612565.1	99.25	1,524,160	50.48	0.68	69.25	0.38	108	44,576	<i>Muribaculaceae</i> bacterium, bin 023
JAPKV010000000	GCA_009775535.1		3,481,771	37.50	4.36	88.31	1.54	135	50,027	<i>Lachnospiraceae</i> bacterium, bin 024
JAPKV010000000			1,861,201	27.30	0.67	86.02	1.29	272	59,996	<i>Clostridia</i> bacterium, bin 025
JAPKV010000000			4,400,590	51.07	2.34	90.73	0.92	41	126,016	<i>Lachnospiraceae</i> bacterium, bin 026
JAPKV010000000	GCA_009773975.1	99.81	2,383,586	52.87	0.41	87.49	0.59	113	83,350	<i>Desulfovibrio</i> sp., bin 027
JAPKV010000000	GCA_003979135.1	98.75	2,060,160	54.96	0.83	99.90	2.04	64	99,201	<i>Alistipes</i> sp., bin 028
JAPKV010000000			2,714,132	52.97	0.50	97.18	25.8	59	61,622	<i>Clostridia</i> bacterium, bin 029
JAPKV010000000			5,217,971	51.80	2.18	98.25	34.2	610	24,850	<i>Acetatifactor</i> sp., bin 030
JAPKV010000000			2,981,170	46.03	0.37	87.36	4.08	132	67,810	<i>Lachnospiraceae</i> bacterium, bin 031
JAPKV010000000			1,729,729	42.39	0.17	71.37	13.5	254	30,173	<i>Clostridia</i> bacterium, bin 032
JAPKV010000000	GCF_002933775.1	99.07	3,419,211	46.74	2.56	78.21	8.68	72	38,003	<i>Prevotella</i> sp., bin 033
JAPKV010000000			2,679,807	44.95	0.57	95.34	2.42	339	14,766	<i>Odoribacter</i> sp., bin 034
JAPKV010000000	GCA_002428825.1	98.53	1,914,688	59.72	0.95	98.32	1.04	224	87,352	<i>Alistipes</i> sp., bin 035
JAPKV010000000	GCA_002494015.1	99.94	2,917,273	52.01	4.10	90.75	6.60	52	82,727	<i>Duncaniella</i> sp., bin 036
JAPKV010000000	GCF_004803915.1	99.41	3,178,942	49.17	1.44	90.75	6.60	214	40,682	<i>Duncaniella dubosii</i> , bin 037
JAPKV010000000			2,142,509	36.95	0.39	93.16	5.84	245	6,302	" <i>Candidatus</i> Gastranaerophilales" bacterium, bin 038
JAPKV010000000	GCF_003833075.1	96.53	3,771,089	49.04	10.7	99.43	2.70	60	53,253	<i>Muribaculum</i> sp., bin 039
JAPKV010000000	GCA_003513705.1	99.50	2,077,509	46.87	0.51	66.96	1.95	481	5,638	<i>Muribaculaceae</i> bacterium, bin 040
JAPKV010000000			5,554,479	38.32	5.66	98.05	3.74	751	15,829	<i>Lachnospiraceae</i> bacterium, bin 041
JAPKV010000000	GCA_003979155.1	98.85	1,845,454	48.04	0.36	82.98	0.13	120	43,933	<i>Muribaculaceae</i> bacterium, bin 042
JAPKV010000000			2,587,822	51.28	4.12	98.29	6.24	256	33,926	<i>Muribaculaceae</i> bacterium, bin 043
JAPKV010000000	GCF_011959405.1	99.28	1,975,854	53.21	2.14	87.92	0.75	163	71,475	<i>Muribaculaceae</i> bacterium, bin 044
JAPKV010000000	GCF_011959105.1	99.40	3,970,688	51.83	5.04	94.25	1.26	390	20,216	<i>Acetatifactor</i> sp., bin 045
JAPKV010000000			1,552,516	26.73	0.37	93.26	2.81	492	87,467	<i>Bacillus</i> bacterium, bin 046
JAPKV010000000			1,133,810	48.93	0.09	42.36	2.67	47	58,474	<i>Firmicutes</i> bacterium, bin 047
JAPKV010000000	GCA_002362235.1	99.38	4,221,123	38.63	2.42	98.10	1.62	572	2,159	<i>Lachnospiraceae</i> bacterium, bin 048
JAPKV010000000			2,145,455	57.14	4.56	98.40	2.56	251	29,355	<i>Alistipes</i> sp., bin 049
JAPKV010000000			1,606,242	28.19	0.47	91.01	1.12	630	19,231	<i>Bacillus</i> bacterium, bin 050
JAPKV010000000			2,828,579	46.24	1.01	96.64	<0.01	20	185,047	<i>Oscillospiraceae</i> bacterium, bin 051

<sup>a</sup> All accession numbers are associated with BioProject accession number PRJNA876044.

<sup>b</sup> The closest genome as determined by average nucleotide identity (ANI) analysis. This and the closest reference genome are used to determine the species assignment.

<sup>c</sup> GTDB metric for fast alignment-free computation of the whole-genome ANI. This is calculated only when the query genome can be placed within a defined genus.

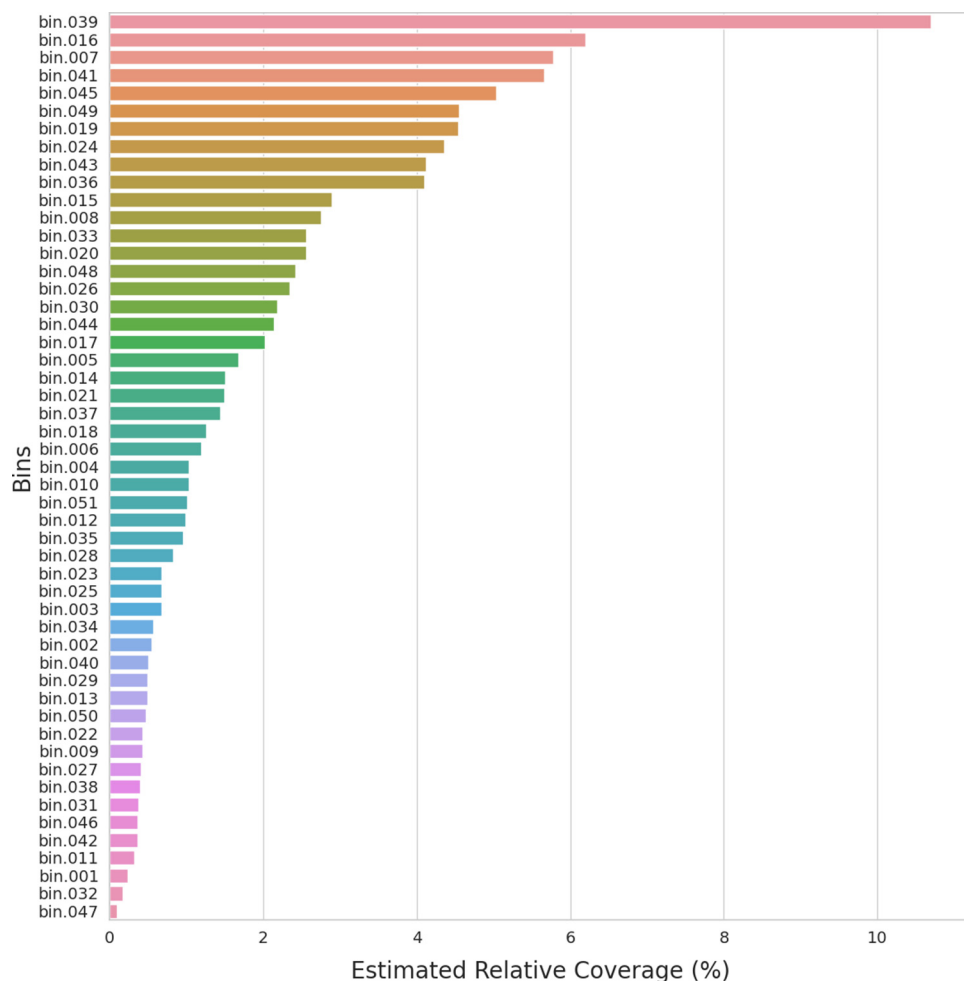
<sup>d</sup> Genome completeness calculated by CheckM.

<sup>e</sup> Contamination determined by CheckM.

<sup>f</sup> Total number of contigs in the MAG.

<sup>g</sup> Length of the shortest contig in the set of largest contigs making up at least 50% of the total assembly.

<sup>h</sup> NCBI-assigned taxonomy and original bin numbers from the KBase analysis.



**FIG 1** Relative abundance in the most abundant bins. The relative abundance is the number of reads per bin divided by the total number of reads in all of the KBase bins from both metagenome samples. The bin numbers can be found next to the NCBI identifications in Table 1. The most abundant bin is bin 039, a *Muribaculum* sp. strain.

to the assembled contigs and determining how many times the contigs aligned to the reads. The two read libraries were combined before the Salmon mapping process. All tools used in the analysis were run with default parameters unless otherwise specified.

Table 1 details information on the bins, including the GenBank accession numbers, the closest genome in the Genome Taxonomy Database (GTDB), the completeness of the genomes, the number of contigs, and the NCBI-assigned taxonomy. The KBase workflow also produced a plot of MAG genome quality, a phylogenetic analysis of the MAGs, and a visualization of their functional pathways (16, 17). The Salmon analysis determined the relative abundances of the MAGs (Fig. 1). The most abundant MAG according to the Salmon analysis was a novel *Muribaculum* sp. strain (formerly S24-7) (bin 039) (Fig. 1 and Table 1).

**Data availability.** Raw metagenomic sequences and the MAGs can be found at NCBI under BioProject accession number [PRJNA876044](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA876044). Details on the analysis of the assemblies can be found as a KBase narrative (<https://doi.org/10.25982/116829.47/1887425>).

#### ACKNOWLEDGMENTS

We thank Annie Chen for collecting the fecal samples and performing the DNA extractions. Gut microbial metagenomes were sequenced by the University of California, San Diego Institute for Genomic Medicine Genomics Center (grant P30 DK063491) utilizing an Illumina NovaSeq 6000 system that was purchased with funding from a NIH Shared Instrument Grant (SIG) award (grant S10 OD026929). This work was funded by NIH grant R01 HD095412 to V.G.T. and S.T.K.

## REFERENCES

- Brandscheid C, Schuck F, Reinhardt S, Schäfer K-H, Pietrzik CU, Grimm M, Hartmann T, Schwiertz A, Endres K. 2017. Altered gut microbiome composition and tryptic activity of the 5×FAD Alzheimer's mouse model. *J Alzheimers Dis* 56:775–788. <https://doi.org/10.3233/JAD-160926>.
- Candon S, Perez-Arroyo A, Marquet C, Valette F, Foray A-P, Pelletier B, Milani C, Ventura M, Bach J-F, Chatenoud L. 2015. Antibiotics in early life alter the gut microbiome and increase disease incidence in a spontaneous mouse model of autoimmune insulin-dependent diabetes. *PLoS One* 10:e0125448. <https://doi.org/10.1371/journal.pone.0125448>.
- Fujimura KE, Demoor T, Rauch M, Faruqi AA, Jang S, Johnson CC, Boushey HA, Zoratti E, Ownby D, Lukacs NW, Lynch SV. 2014. House dust exposure mediates gut microbiome *Lactobacillus* enrichment and airway immune defense against allergens and virus infection. *Proc Natl Acad Sci U S A* 111:805–810. <https://doi.org/10.1073/pnas.1310750111>.
- Kelley ST, Skarra DV, Rivera AJ, Thackray VG. 2016. The gut microbiome is altered in a letrozole-induced mouse model of polycystic ovary syndrome. *PLoS One* 11:e0146509. <https://doi.org/10.1371/journal.pone.0146509>.
- Torres PJ, Ho BS, Arroyo P, Sau L, Chen A, Kelley ST, Thackray VG. 2019. Exposure to a healthy gut microbiome protects against reproductive and metabolic dysregulation in a PCOS mouse model. *Endocrinology* 160:1193–1204. <https://doi.org/10.1210/en.2019-00050>.
- Ho B, Ryback D, Benson B, Mason CN, Torres PJ, Quinn RA, Thackray VG, Kelley ST. 2021. Gut metabolites are more predictive of disease and cohoused states than gut bacterial features in a polycystic ovary syndrome-like mouse model. *mSystems* 6:e01149-20. <https://doi.org/10.1128/mSystems.01149-20>.
- Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ pre-processor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- 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>.
- Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, et al. 2018. KBase: the United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>.
- Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32:605–607. <https://doi.org/10.1093/bioinformatics/btv638>.
- Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 3:836–843. <https://doi.org/10.1038/s41564-018-0171-1>.
- Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365. <https://doi.org/10.1038/srep08365>.
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
- 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>.
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14:417–419. <https://doi.org/10.1038/nmeth.4197>.
- Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodríguez-Ramos J, Bolduc B, Gazitúa MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan MB, Roux S, Wrighton KC. 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* 48:8883–8900. <https://doi.org/10.1093/nar/gkaa621>.
- Stothard P, Wishart DS. 2005. Circular genome visualization and exploration using CGView. *Bioinformatics* 21:537–539. <https://doi.org/10.1093/bioinformatics/bti054>.