



SOFTWARE TOOL ARTICLE

SnakeMAGs: a simple, efficient, flexible and scalable workflow to reconstruct prokaryotic genomes from metagenomes

[version 1; peer review: awaiting peer review]

Nachida Tadrent¹, Franck Dedeine ¹, Vincent Hervé ^{1,2}

¹Institut de Recherche sur la Biologie de l'Insecte, UMR 7261, CNRS-Université de Tours, Tours, 37200, France

²Université Paris-Saclay, INRAE, AgroParisTech, UMR SayFood, Palaiseau, 91120, France

V1 First published: 15 Dec 2022, 11:1522
<https://doi.org/10.12688/f1000research.128091.1>

Latest published: 15 Dec 2022, 11:1522
<https://doi.org/10.12688/f1000research.128091.1>

Abstract

Background: Over the last decade, we have observed in microbial ecology a transition from gene-centric to genome-centric analyses. Indeed, the advent of metagenomics combined with binning methods, single-cell genome sequencing as well as high-throughput cultivation methods have contributed to the continuing and exponential increase of available prokaryotic genomes, which in turn has favored the exploration of microbial metabolisms. In the case of metagenomics, data processing, from raw reads to genome reconstruction, involves various steps and software which can represent a major technical obstacle.

Methods: To overcome this challenge, we developed *SnakeMAGs*, a simple workflow that can process Illumina data, from raw reads to metagenome-assembled genomes (MAGs) classification and relative abundance estimate. It integrates state-of-the-art bioinformatic tools to sequentially perform: quality control of the reads (illumina-utils, Trimmomatic), host sequence removal (optional step, using Bowtie2), assembly (MEGAHIT), binning (MetaBAT2), quality filtering of the bins (CheckM), classification of the MAGs (GTDB-Tk) and estimate of their relative abundance (CoverM). Developed with the popular Snakemake workflow management system, it can be deployed on various architectures, from single to multicore and from workstation to computer clusters and grids. It is also flexible since users can easily change parameters and/or add new rules.

Results: Using termite gut metagenomic datasets, we showed that *SnakeMAGs* is slower but allowed the recovery of more MAGs encompassing more diverse phyla compared to another similar workflow named ATLAS.

Conclusions: Overall, it should make the reconstruction of MAGs more accessible to microbiologists. *SnakeMAGs* as well as test files and an extended tutorial are available at

<https://github.com/Nachida08/SnakeMAGs>.

Open Peer Review

Approval Status AWAITING PEER REVIEW

Any reports and responses or comments on the article can be found at the end of the article.

Keywords

Snakemake, metagenomics, microbiology, genomics, bioinformatics, microbial ecology



This article is included in the **Bioinformatics** gateway.

Corresponding author: Vincent Hervé (vincent.herve@inrae.fr)

Author roles: **Tadrent N:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Review & Editing; **Dedeine F:** Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing; **Hervé V:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the FEDER grant InFoBioS n°EX011185 (région Val de Loire, France), the Centre National de la Recherche Scientifique (CNRS) and the Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE). Ph.D. work of Nachida Tadrent (Ph.D. scholarship) was funded by the University of Tours.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Tadrent N *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Tadrent N, Dedeine F and Hervé V. ***SnakeMAGs: a simple, efficient, flexible and scalable workflow to reconstruct prokaryotic genomes from metagenomes*** [version 1; peer review: awaiting peer review] F1000Research 2022, 11:1522 <https://doi.org/10.12688/f1000research.128091.1>

First published: 15 Dec 2022, 11:1522 <https://doi.org/10.12688/f1000research.128091.1>

Introduction

Over the last years, microbial ecology has progressively made the transition from gene-centric to genome-centric analyses,¹ allowing the clear assignment of (sometimes novel) microbial taxa to specific functions and metabolisms.^{2–5} Indeed, technical and technological progresses such as binning methods applied to metagenomics,⁶ single-cell genome sequencing⁷ as well as high-throughput cultivation methods⁸ have contributed to the continuing and exponential increase of available prokaryotic genomes.⁹ This is particularly true for metagenomics that offers the possibility to reconstruct metagenome-assembled genomes (MAGs) on a large scale and from various environments, and thus has generated a huge amount of new prokaryotic genomes.^{10,11}

Although the use of MAGs in microbial ecology is becoming a common practice nowadays, processing raw metagenomic reads up to genome reconstruction involves various steps and software which can represent a major technical obstacle, especially for non-specialists. To face this problem, several workflows such as MetaWRAP,¹² its Snakemake version called SnakeWRAP,¹³ ATLAS¹⁴ and more recently MAGNETO,¹⁵ have been developed to automatically reconstruct genomes from metagenomes. However, these workflows contain various modules and perform more tasks than only generating MAGs. For instance, they will taxonomically assign the metagenomic reads, create gene catalog or perform functional annotations. They rely on numerous dependencies, require significant computational resources and regenerate a lot of outputs which are not essential to most research projects. To simplify this procedure and make it more accessible while remaining efficient, reproducible and biologically relevant, we developed with the popular Snakemake workflow management system,¹⁶ a configurable and easy-to-use workflow called *SnakeMAGs* to reconstruct MAGs in just a few steps. It integrates state-of-the-art bioinformatic tools to sequentially perform from Illumina raw reads: quality filtering of the reads, adapter trimming, an optional step of host sequence removal, assembly of the reads, binning of the contigs, quality assessment of the bins, taxonomic classification of the MAGs and estimation of the relative abundance of these MAGs.

Methods

Creation

Our tool was built by integrating a set of software needed to process metagenomic datasets, utilizing Snakemake. There are no additional equations/math needed to recreate this tool.

Implementation

The workflow has been developed with the workflow management system Snakemake v7.0.0¹⁶ based on the Python language. Snakemake enables reproducible and scalable data analyses as well as an independent management of the required software within a workflow. *SnakeMAGs* is composed of two main files:

The Snakefile, named “SnakeMAGs.smk”, contains the workflow script. It is divided into successive rules which correspond to individual steps. Our workflow includes a total of 15 distinct rules. Each rule requires input files and relies on a single software installed independently when starting the workflow in a dedicated conda v4.12.0 environment. At the end of each rule, output files will be generated in a dedicated folder, as well as a log file (stored in the logs folder) summarizing the events of the software run and a benchmark file (stored in the benchmarks folder) containing the central processing unit (CPU) run time, the wall clock time and the maximum memory usage required to complete the rule. Thanks to Snakemake wildcards, our rules are generalized, so one can process multiple datasets in parallel without having to adjust the source code manually.

The configuration file,⁴⁰ named “config.yaml”, is used to define some variable names (*e.g.* names of the input files), paths (*e.g.* working directory, location of the reference databases), software parameters and computational resource allocations (threads, memory) for each of the main steps.

To run the workflow, the user only requires Snakemake. It can be easily installed, for instance *via* Conda, as explained in the GitHub repository:

```
conda create -n snakemake_7.0.0 snakemake=7.0.0
```

After that, the user will only have to edit the config file (an example is provided on the GitHub repository) and then run *SnakeMAGs*:

```
#Example of command on a Slurm cluster
snakemake --snakefile SnakeMAGs.smk --cluster \
'sbatch -p <cluster_partition> --mem -c \
-o "cluster_logs/{wildcards}.{rule}.{jobid}.out" \
```

```
-e "cluster_logs/{wildcards}.{rule}.{jobid}.err" ' \
--jobs --use-conda --conda-frontend \
conda --conda-prefix/path/to/SnakeMAGs_conda_env/ \
--jobname "{rule}.{wildcards}.{jobid}" --configfile/path/to/config.yaml
```

During the first use of the workflow, a dedicated Conda environment will be installed for each of the bioinformatic tool to avoid conflict. Then the input files will be processed sequentially. Output files will be stored in eight dedicated folder: logs, benchmarks, QC_fq (containing FASTQ files), Assembly, Binning, Bins_quality (all three containing FASTA files), Classification (containing FASTA files and text files with the taxonomic information), and MAGs_abundances (text files).

The workflow has been successfully used on a workstation with Ubuntu 22.04 as well as on high-performance computer clusters with Slurm v18.08.7 and SGE v8.1.9.

Operation

The minimal system requirements to run the workflow will depend on the size of the metagenomic dataset. Small datasets (e.g. the test files provided on the GitHub repository) have been successfully analyzed on a workstation with an Intel Xeon Silver 4210, 2.20GHz (10 cores/20 threads) processor and 96GB of RAM. Larger datasets should be processed on cluster computing or within a high-performance infrastructure. For instance, performance evaluation of publicly available

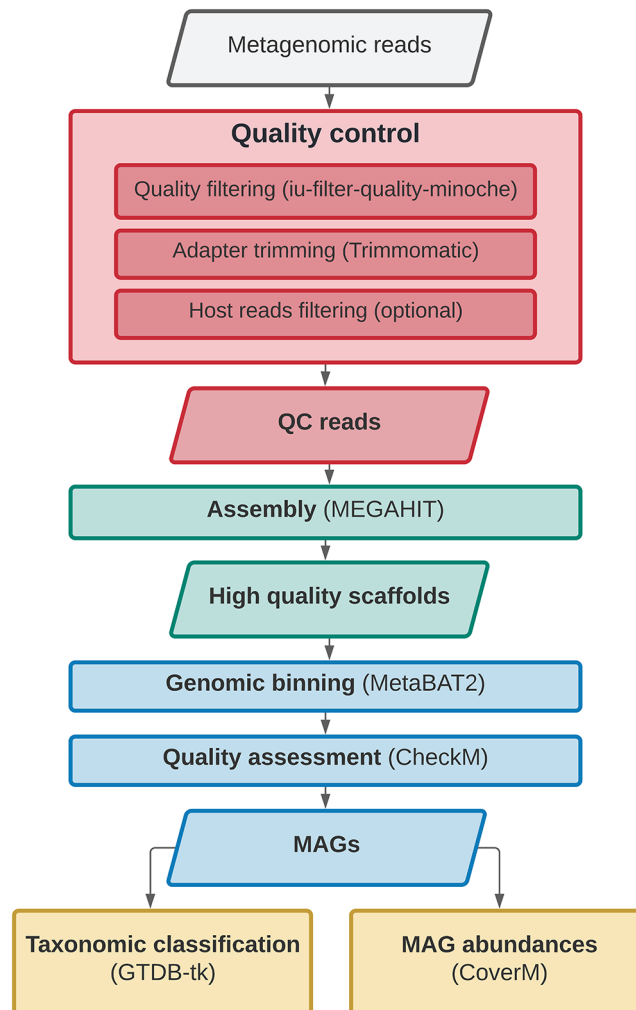


Figure 1. Directed acyclic graph describing the main steps performed by *SnakeMAGs*. The names of the software used for each step are showed in parentheses.

metagenomes (see below) was performed on a computer cluster under CentOS Linux release 7.4.1708 distribution with Slurm 18.08.7, on a node possessing an Intel Xeon CPU E7-8890 v4, 2.20GHz (96 cores/192 threads) and 512 GB RAM.

SnakeMAGs integrates a series of bioinformatic tools to sequentially perform from Illumina raw reads: quality filtering of the reads with *illumina-utils* v2.12,¹⁷ adapter trimming with *Trimmomatic* v0.39¹⁸ (RRID:SCR_011848), an optional step of host sequence removal (*e.g.* animal or plant sequences) with *Bowtie2* v2.4.5¹⁹ (RRID:SCR_016368), assembly of the reads with *MEGAHIT* v1.2.9²⁰ (RRID:SCR_018551), binning of the contigs with *MetaBAT2* v2.15²¹ (RRID:SCR_019134), quality assessment of the bins with *CheckM* v1.1.3²² (RRID:SCR_016646), classification of the MAGs with *GTDB-Tk* v2.1.0²³ (RRID:SCR_019136) and estimation of the relative abundance of these MAGs with *CoverM* v0.6.1. An overview of the workflow is presented in [Figure 1](#).

Use cases

To demonstrate the benefits and potential of our workflow, we compared it to another Snakemake workflow named *ATLAS* v2.9.1.¹⁴ To produce a fair comparison, *ATLAS* was run with the *MEGAHIT* assembler, without co-binning and dereplicating only 100% similar MAGs. To test these two workflows, we downloaded and analyzed ten publicly available termite gut metagenomes (accession numbers: SRR10402454; SRR14739927; SRR8296321; SRR8296327; SRR8296329; SRR8296337; SRR8296343; DRR097505; SRR7466794; SRR7466795) from five studies^{24–28} and belonging to ten different termite species.

SnakeMAGs requires only a limited number of inputs files: the raw metagenomic reads in FASTQ format from the 10 above-mentioned metagenomes, a FASTA file containing the adapter sequences,⁴⁰ a YAML configuration file specifying the variable names, paths and computational resource allocations (available on the GitHub repository and on Zenodo), and here since we worked with host-associated metagenomes a FASTA file containing the termite genome sequences.³⁹ Regarding the outputs, *SnakeMAGs* produced quality-controlled FASTQ files without adapters nor termite sequences, in the QC_fq folder. Then the reads assembled into contigs and scaffolds (FASTA files) were saved in the

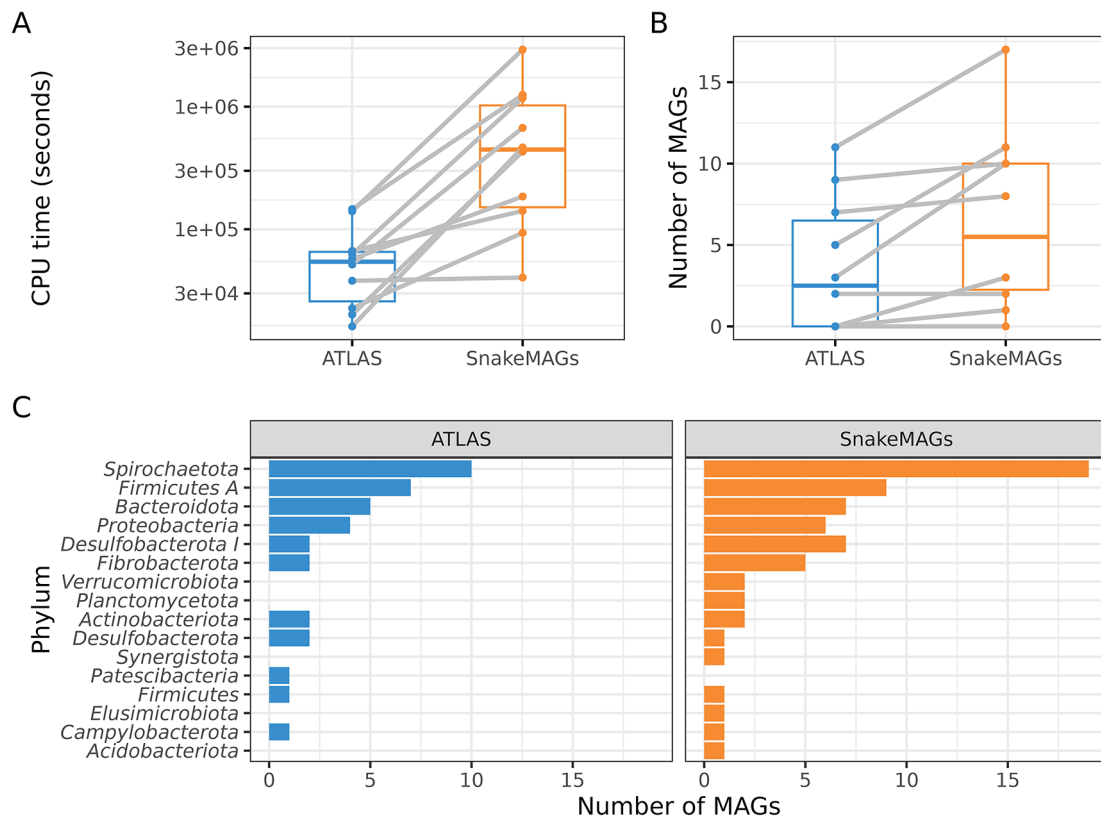


Figure 2. Comparison of the performance of *SnakeMAGs* v1.0.0 with another workflow, namely *ATLAS* v2.9.1¹⁴ using 10 termite gut metagenomes. A. CPU time (in seconds) required to process each metagenome. B. Number of MAGs reconstructed from each metagenome. On both boxplots, gray lines link the result obtained with *ATLAS* and the one obtained with *SnakeMAGs* for each of the 10 analyzed termite metagenomes. C. Number of bacterial MAGs at the phylum level recovered from each workflow.

Assembly folder. Products of the binning procedure were stored in the Binning folder. Bins with >50% completeness and <10% contamination (according to CheckM) were considered as MAGs and stored in the Bins_quality folder. Subsequently, the results of the MAGs classification and relative abundance estimation were sent to the Classification and MAGs_abundances folders, respectively. ATLAS requires similar input files and produces, among others, similar outputs files.

ATLAS appeared to be faster than *SnakeMAGs* to reconstruct MAGs from metagenomes (Figure 2A). However, *SnakeMAGs* always recovered more MAGs (>50% completeness and <10% contamination according to CheckM) per metagenome or at least as much as ATLAS (Figure 2B). From the ten metagenomes, *SnakeMAGs* produced a total of 65 MAGs while ATLAS generated only 37 MAGs. Additionally, *SnakeMAGs* was able to recover MAGs encompassing a higher diversity of bacterial phyla ($n = 15$ phyla) compared to ATLAS ($n = 11$ phyla). Only one phylum, namely *Patescibacteria*, represented by a single MAG was recovered by ATLAS and not by *SnakeMAGs*. On the contrary, ATLAS failed to reconstruct MAGs belonging to *Verrucomicrobiota*, *Planctomycetota*, *Synergistota*, *Elusimicrobiota* and *Acidobacteriota* when *SnakeMAGs* succeeded (Figure 2C).

Discussion

Using metagenomic datasets from the gut of various termite species, our analyses revealed that while being slower, *SnakeMAGs* allowed the recovery of more MAGs encompassing more diverse phyla compared to ATLAS, another similar Snakemake workflow. More importantly our results showed that *SnakeMAGs* was able to recover MAGs encompassing the major bacterial phyla found in termite guts,^{29,30} and that some of these phyla were not recovered by ATLAS. Indeed, taxa belonging to *Verrucomicrobiota*,³¹ *Planctomycetota*,^{30,32} *Synergistota*,³³ *Elusimicrobiota*³⁴ and *Acidobacteriota*^{35,36} have been repeatedly found in the gut of various termite species. As such, they would represent relevant targets for genome-centric analyses of the termite gut microbiota. Therefore, we showed that *SnakeMAGs* has the potential to retrieve quantitatively more genomic information from metagenomes but also to extract genomic features of biological interest.

Thanks to the inherent flexibility of Snakemake, *SnakeMAGs* offers the possibility to the users to easily tune the parameters of the workflow (*e.g.* resource allocations for each rule, options of a specific tools) to adapt their analysis to the datasets and to the computational infrastructure. Additionally, advanced users will have the opportunity to edit or add new rules to the workflow. Regarding the future of *SnakeMAGs*, several avenues will be considered for the next versions of the workflow. Firstly, the workflow could give more freedom to the users by offering the choice of different tools to perform the same task (*e.g.* different trimming, assembly or binning software). Secondly, with the current emergence of metagenomic datasets generated with long-read DNA sequencing,³⁷ it might be relevant to adjust our workflow for long-read sequencing technology by including specific bioinformatic tools for this technology.³⁸ Meanwhile, since the majority of the metagenomic datasets have been and are still currently generated with Illumina short-read technology, *SnakeMAGs* can be widely used to explore the genomic content of various ecosystems *via* metagenomics.

Software availability

Source code available from: <https://github.com/Nachida08/SnakeMAGs>

Archived source code at time of publication: <https://doi.org/10.5281/zenodo.7334838>.³⁹

License: CeCILL v2.1

Data availability

Source data

Termite genome references used for removing host sequences and their Bowtie2 index are available at: <https://zenodo.org/record/6908287#.Y1JLANJBzUR>

The termite gut metagenomes analyzed in the present study are available on NCBI with the following accession numbers: [SRR10402454](https://ncbi.nlm.nih.gov/sra/SRR10402454); [SRR14739927](https://ncbi.nlm.nih.gov/sra/SRR14739927); [SRR8296321](https://ncbi.nlm.nih.gov/sra/SRR8296321); [SRR8296327](https://ncbi.nlm.nih.gov/sra/SRR8296327); [SRR8296329](https://ncbi.nlm.nih.gov/sra/SRR8296329); [SRR8296337](https://ncbi.nlm.nih.gov/sra/SRR8296337); [SRR8296343](https://ncbi.nlm.nih.gov/sra/SRR8296343); [DRR097505](https://ncbi.nlm.nih.gov/sra/DRR097505); [SRR7466794](https://ncbi.nlm.nih.gov/sra/SRR7466794); [SRR7466795](https://ncbi.nlm.nih.gov/sra/SRR7466795).

Underlying data

Zenodo. Reconstruction of prokaryotic genomes from ten termite gut metagenomes using two distinct workflows: *SnakeMAGs* and ATLAS: <https://doi.org/10.5281/zenodo.7334397>.⁴⁰

- *SnakeMAGs_config.yaml* (The configuration file used to analyze the 10 termite gut metagenomes with *SnakeMAGs*)

- ATLAS_config.yaml (The configuration file used to analyze the 10 termite gut metagenomes with ATLAS)
- MAGs_SnakeMAGs.zip (A zipped folder containing the genomes of the 65 MAGs reconstructed with *SnakeMAGs*)
- MAGs_ATLAS.zip (A zipped folder containing the genomes of the 37 MAGs reconstructed with ATLAS)
- taxonomic_assignment_MAGs.csv (A text file containing the taxonomic assignment of all the MAGs reconstructed by both workflows)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0).

Acknowledgments

The authors thank Emmanuelle Morin and H el ene Gardon for their valuable advice and feedback during the workflow development.

References

1. Prosser JI: **Dispersing misconceptions and identifying opportunities for the use of "omics" in soil microbial ecology.** *Nat. Rev. Microbiol.* 2015 Jun 8; **13**(7): 439–446. [PubMed Abstract](#) | [Publisher Full Text](#)
2. Evans PN, Parks DH, Chadwick GL, et al.: **Methane metabolism in the archaeal phylum *Bathyarchaeota* revealed by genome-centric metagenomics.** *Science.* 2015 Oct 23; **350**(6259): 434–438. [PubMed Abstract](#) | [Publisher Full Text](#)
3. Engelberts JP, Robbins SJ, de Goeij JM, et al.: **Characterization of a sponge microbiome using an integrative genome-centric approach.** *ISME J.* 2020 Jan 28; 1–11. [Publisher Full Text](#)
4. Loh HQ, Herv e V, Brune A: **Metabolic potential for reductive acetogenesis and a novel energy-converting [NiFe] hydrogenase in *Bathyarchaeia* from termite guts – A genome-centric analysis.** *Front. Microbiol.* 2021 Feb 3; **11**: 3644. [Publisher Full Text](#)
5. Bay SK, Dong X, Bradley JA, et al.: **Trace gas oxidizers are widespread and active members of soil microbial communities.** *Nat. Microbiol.* 2021 Feb 4; **6**(2): 246–256. [PubMed Abstract](#) | [Publisher Full Text](#)
6. Sedlar K, Kupkova K, Provaznik I: **Bioinformatics strategies for taxonomy independent binning and visualization of sequences in shotgun metagenomics.** *Comput. Struct. Biotechnol. J.* 2017 Jan 1; **15**: 48–55. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Woyke T, Doud DFR, Schulz F: **The trajectory of microbial single-cell sequencing.** *Nat. Methods.* 2017 Oct 31; **14**(11): 1045–1054. [PubMed Abstract](#) | [Publisher Full Text](#)
8. Overmann J, Abt B, Sikorski J: **Present and future of culturing bacteria.** *Annu. Rev. Microbiol.* 2017 Sep 8; **71**(1): 711–730. [Publisher Full Text](#)
9. Almeida A, Nayfach S, Boland M, et al.: **A unified catalog of 204,938 reference genomes from the human gut microbiome.** *Nat. Biotechnol.* 2021 Jan 20; **39**(1): 105–114. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Nayfach S, Roux S, Seshadri R, et al.: **A genomic catalog of Earth's microbiomes.** *Nat. Biotechnol.* 2021 Apr 9; **39**(4): 499–509. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Pasolli E, Asnicar F, Manara S, et al.: **Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle.** *Cell.* 2019 Jan; **176**(3): 649–662.e20. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Uritskiy GV, DiRuggiero J, Taylor J: **MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis.** *Microbiome.* 2018 Dec 15; **6**(1): 158. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Krapohl J, Pickett B: **SnakeWRAP: a Snakemake workflow to facilitate automated processing of metagenomic data through the metaWRAP pipeline [version 2; peer review: 1 approved].** *F1000Res.* 2022; **11**(265). [Publisher Full Text](#)
14. Kieser S, Brown J, Zdobnov EM, et al.: **ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data.** *BMC Bioinformatics.* 2020 Dec 22; **21**(1): 257. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Churchward B, Millet M, Bihou e A, et al.: **MAGNETO: An automated workflow for genome-resolved metagenomics.** *mSystems.* 2022; **7**(4): e00432–e00422. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
16. M older F, Jablonski KP, Letcher B, et al.: **Sustainable data analysis with Snakemake.** *F1000Res.* 2021; Vol. **10**: [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [Reference Source](#)
17. Eren AM, Vineis JH, Morrison HG, et al.: **A filtering method to generate high quality short reads using Illumina paired-end technology.** *PLoS One.* 2013; **8**(6). [Publisher Full Text](#)
18. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics.* 2014 Aug 1; **30**(15): 2114–2120. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat. Methods.* 2012 Apr 4; **9**(4): 357–359. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Li D, Liu CM, Luo R, et al.: **MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph.** *Bioinformatics.* 2015 May 15; **31**(10): 1674–1676. [PubMed Abstract](#) | [Publisher Full Text](#)
21. Kang DD, Li F, Kirton E, et al.: **MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies.** *PeerJ.* 2019 Jul 26; **7**: e7359. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Parks DH, Imelfort M, Skennerton CT, et al.: **CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.** *Genome Res.* 2015 Jul; **25**(7): 1043–1055. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Chaumeil PA, Mussig AJ, Hugenholtz P, et al.: **GTDB-TK v2: memory friendly classification with the Genome Taxonomy Database.** *Bioinformatics.* 2022 Oct 11; btac672. [Publisher Full Text](#)
24. Calusinska M, Marynowska M, Bertucci M, et al.: **Integrative omics analysis of the termite gut system adaptation to *Miscanthus* diet identifies lignocellulose degradation enzymes.** *Communications Biology.* 2020 Dec 1; **3**(1): 275. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Moreira EA, Persinoti GF, Menezes LR, et al.: **Complementary contribution of fungi and bacteria to lignocellulose digestion in the food stored by a neotropical higher termite.** *Front. Ecol. Evol.* 2021 Apr 26; **9**: 248. [Publisher Full Text](#)

26. Romero Victorica M, Soria MA, Batista-García RA, *et al.*: **Neotropical termite microbiomes as sources of novel plant cell wall degrading enzymes.** *Sci. Rep.* 2020 Dec 2; **10**(1): 3864.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Tokuda G, Mikaelyan A, Fukui C, *et al.*: **Fiber-associated spirochetes are major agents of hemicellulose degradation in the hindgut of wood-feeding higher termites.** *Proc. Natl. Acad. Sci.* 2018 Dec 18; **115**(51): E11996–E12004.
[PubMed Abstract](#) | [Publisher Full Text](#)
28. Waidele L, Korb J, Voolstra CR, *et al.*: **Ecological specificity of the metagenome in a set of lower termite species supports contribution of the microbiome to adaptation of the host.** *Animal Microbiome.* 2019 Dec 24; **1**(1): 13.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Arora J, Kinjo Y, Šobotník J, *et al.*: **The functional evolution of termite gut microbiota.** *Microbiome.* 2022 Dec; **10**(1): 78.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Hervé V, Liu P, Dietrich C, *et al.*: **Phylogenomic analysis of 589 metagenome-assembled genomes encompassing all major prokaryotic lineages from the gut of higher termites.** *PeerJ.* 2020 Feb; **8**: e8614.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Wertz JT, Kim E, Breznak JA, *et al.*: **Genomic and physiological characterization of the *Verrucomicrobia* isolate *Diplosphaera colitermitum* gen. nov., sp. nov., reveals microaerophily and nitrogen fixation genes.** *Appl. Environ. Microbiol.* 2012 Mar 1; **78**(5): 1544–1555.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Köhler T, Stingl U, Meuser K, *et al.*: **Novel lineages of *Planctomycetes* densely colonize the alkaline gut of soil-feeding termites (*Cubitermes* spp.).** *Environ. Microbiol.* 2008 May; **10**(5): 1260–1270.
[PubMed Abstract](#) | [Publisher Full Text](#)
33. Ahmad F, Yang G, Zhu Y, *et al.*: **Tripartite symbiotic digestion of lignocellulose in the digestive system of a fungus-growing termite.** *Microbiology Spectrum.* 2022 Oct 17; e01234–e01222.
[PubMed Abstract](#) | [Publisher Full Text](#)
34. Herlemann DPR, Geissinger O, Ikeda-Ohtsubo W, *et al.*: **Genomic analysis of "*Elusimicrobium minutum*," the first cultivated representative of the phylum "*Elusimicrobia*" (formerly termite group 1).** *Appl. Environ. Microbiol.* 2009 May 1; **75**(9): 2841–2849.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Hongoh Y, Deevong P, Inoue T, *et al.*: **Intra- and interspecific comparisons of bacterial diversity and community structure support coevolution of gut microbiota and termite host.** *Appl. Environ. Microbiol.* 2005 Nov 1; **71**(11): 6590–6599.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Bourguignon T, Lo N, Dietrich C, *et al.*: **Rampant host switching shaped the termite gut microbiome.** *Curr. Biol.* 2018 Feb; **28**(4): 649–654.e2.
[PubMed Abstract](#) | [Publisher Full Text](#)
37. Bickhart DM, Kolmogorov M, Tseng E, *et al.*: **Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities.** *Nat. Biotechnol.* 2022 Jan 3; **40**: 711–719.
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Feng X, Cheng H, Portik D, *et al.*: **Metagenome assembly of high-fidelity long reads with hifiasm-meta.** *Nat. Methods.* 2022 Jun; **19**(6): 671–674.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Tadrent N, Dedeine F, Hervé V: **SnakeMAGs (v1.0.2).** [Code] *Zenodo.* 2022.
[Publisher Full Text](#)
40. Tadrent N, Dedeine F, Hervé V: **Reconstruction of prokaryotic genomes from ten termite gut metagenomes using two distinct workflows: SnakeMAGs and ATLAS.** [Data]. *Zenodo.* 2022.
[Publisher Full Text](#)

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research