

ALGORITHMS AND TOOLS FOR METAGENOMIC DATA ANALYSES

LIQING ZHANG

1. INSTRUCTIONS

Studying the microbiome (the collection of microbes) in a community through metagenomic sequencing data has become a standard practice in recent years. Identifying the organisms from the sequences fast and accurately remains a major computational challenge. Here several reviews and latest papers are included as a starting point for you to become familiar with this field and learn different algorithms/methods used to process and annotate metagenomic data computationally. As expected, there are pros and cons with different methods. You are supposed to start from the references below and identify all other relevant papers. Write a review paper on the papers you read, summarize and classify the algorithms used in processing metagenomic data, enumerate all the issues that need to be considered by different types of methods (e.g., binning vs. assembly), discuss limitations of the current algorithms and programs, and suggest future improvement in terms of how you may address the limitations. Your paper should be like a standard review paper, with at least these sections:

- Abstracts
- Introduction or Background
- Problem Statement and Challenges
- Previous Work
- Strengths and Weaknesses of the Current State of Research ? Proposal for Future Research
- References

2. PAPERS

Here are a few review papers to get started. You are supposed to expand from the papers listed here and read other relevant papers on the topic.

- Comparison of microbiome samples: methods and computational challenges.
<https://academic-oup-com.ezproxy.lib.vt.edu/bib/advance-article/doi/10.1093/bib/bbaa121/5861761>
- Metagenomics: Tools and Insights for Analyzing Next-Generation Sequencing Data Derived from Biodiversity Studies.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4426941/>
- Alignment-free sequence comparison: benefits, applications, and tools.
<https://www.ncbi.nlm.nih.gov/pubmed/28974235>

- Bioinformatics strategies for taxonomy independent binning and visualization of sequences in shotgun metagenomics.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5148923/>
- MetaFast: fast reference-free graph-based comparison of shotgun metagenomic data.
<https://www.ncbi.nlm.nih.gov/pubmed/27259541>
- Entropy-Scaling Search of Massive Biological Data.
[http://www.cell.com/cell-systems/pdf/S2405-4712\(15\)00058-7.pdf](http://www.cell.com/cell-systems/pdf/S2405-4712(15)00058-7.pdf)