Sabreur: Fast, Reliable and Handy Barcode Demultiplexing of Fasta and Fastq Files

Anicet Ebou Dominique Koua Adolphe Zézé

Background

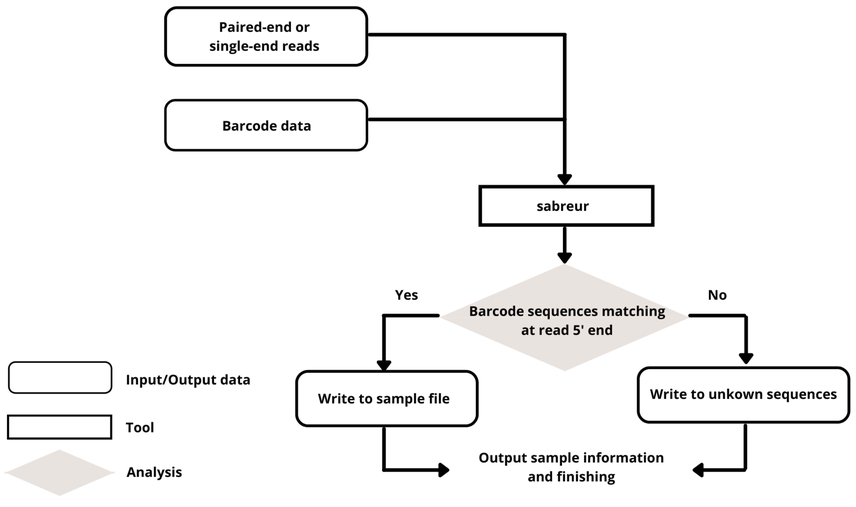
Next-generation sequencing tools are able to generate millions to billions of reads in a single run. To reach such a high rate in a cost-efficient manner, next-generation sequencers often take advantage of the barcoding of multiple samples or species.

To demultiplex raw data, only a few tools are available from which Idemp and sabre are the most used. Most of the current tools are:

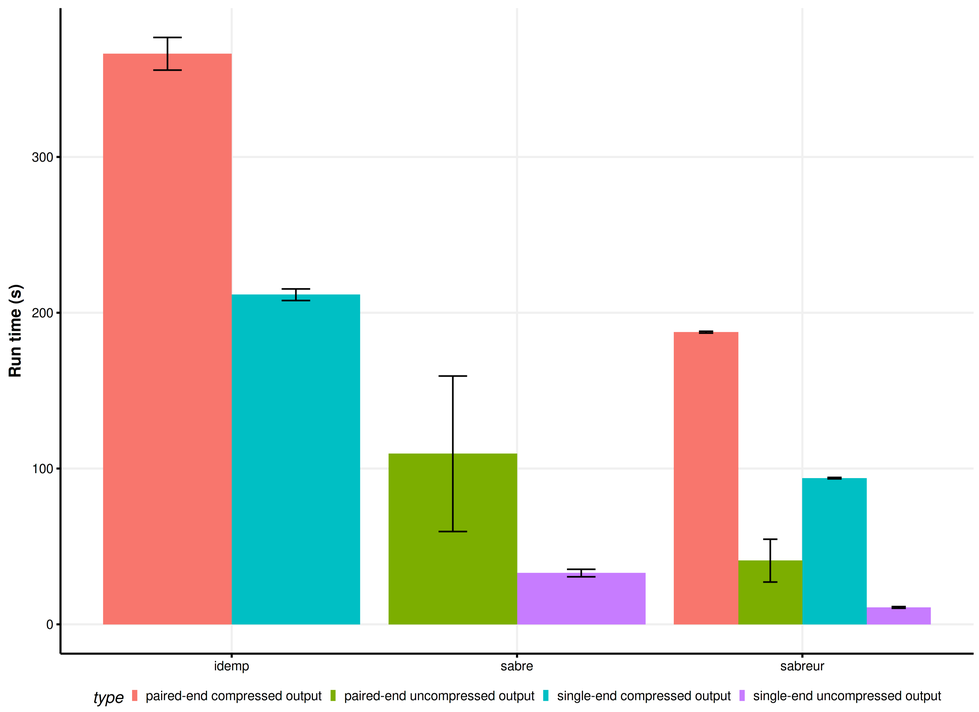
- Not actively maintained
- Support only gzip'd files for or output
- Prone to memory violations like use-after-frees, buffer overflows, and out-of-bounds reads/write [MITRE, 2020].

Algorithm

Bio-rust [Köster, 2020]



Benchmark



References

[MITRE, 2020]: MITRE 2020 CWE Top 25 Most Dangerous Software Weaknesses; 2020; https://cwe.mitre.org/top25/archive/2020/2020_cwe_top25.html

[Köster, 2016]: Köster, J. Rust-Bio: A Fast and Safe Bioinformatics Library. Bioinformatics 2016, 32, 444–446, doi:10.1093/bioinformatics/btv573.

Poster Code Contact