Metagenomic proxy assemblies of single cell genomes

Andreas Bremges[1,2,*], Jessica Jarett[2], Tanja Woyke[2], Alexander Sczyrba[1,2]

[1] Center for Biotechnology and Faculty of Technology, Bielefeld University, 33615 Bielefeld, Germany
[2] U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA

Background

Over 99% of the microbial species observed in nature cannot be grown in pure culture, making it impossible to study them using classical genomic methods. Metagenomics and single cell genomics are two complimentary approaches to study the microbial dark matter.

Metagenomics can obtain genome sequences from uncultivated microbes through direct sequencing of environmental DNA. Each genome's metagenomic coverage is constant and depends only on its abundance. A complementary approach to sequencing DNA of a whole microbial community is single cell genomics. Prior to sequencing of a single cell, its DNA needs to be amplified. This usually is done by multiple displacement amplification (MDA), introducing a tremendous coverage bias. Poorly amplified regions result in extremely low sequencing coverage or physical sequencing gaps. These parts of the genome cannot be reconstructed in the subsequent assembly step, and therefore genomic information is lost.

Results

Frequently, single amplified genomes (SAGs) and shotgun metagenomes are generated from the same environmental sample. We developed a fast, *k*-mer based recruitment method to sensitively identify metagenomic "proxy" reads representing the single cell of interest, using the raw single cell sequencing reads as recruitment seeds. By assembling metagenomic proxy reads instead of the single cell reads, we circumvent most challenges of single cell assembly, such as the aforementioned coverage bias and chimeric MDA products. In a final step, the original single cell reads are used for quality assessment of the proxy assembly.

On real and simulated data we show that, with sufficient metagenomic coverage, assembling metagenomic proxy reads instead of single cell reads significantly improves assembly contiguity while maintaining the original accuracy. By applying our method iteratively, we span physical sequencing gaps and are able to recover genomic regions that otherwise would have been lost. However, careful contamination screening is needed.

Conclusions

We developed kgrep, a new tool that naturally exploits the complementary nature of single cells and metagenomes to improve *de novo* assembly of single cell genomes.

* abremges@cebitec.uni-bielefeld.de