

A new tool for non-hybrid correction of long noisy reads

LIRMM

Eric Rivals, LIRMM & Institut de Biologie Computationnelle (IBC)

CNRS et Université de Montpellier, France

Nowadays, several long read sequencing technologies are available and long reads are theoretically advantageous not only to assemble a genome, but also to investigate the linkage of genetic variations or the diversity of transcriptomes. However, current levels of sequencing errors hamper the use of long reads. For instance, even the simple task of read alignment on a reference sequence becomes less reliable, more complex and time consuming than with short reads. Several hybrid error correction methods, such as LoRDEC were recently proposed: they require and take advantage of short reads to correct long reads. Here, we present a non hybrid error correction method, which only uses long reads. This method, embodied in a software called LoRMA, relies on LoRDEC to iteratively correct long reads using several De Bruijn graphs of increasing order. Then it performs multiple alignment to compute a long read consensus. LoRMA was tested on bacterial and yeast datasets and provides reliable correction in reasonable computing times.

LoRMA is available at: <https://www.cs.helsinki.fi/u/lmsalmel/LoRMA/>

LoRDEC is available at: <http://www.lirmm.fr/~rivals/lordec/>

Work in collaboration with L. Salmela, R. Wake, E. Ukkonen de l'Université d'Helsinki, Finlande.

Publication in Bioinformatics: [doi: 10.1093/bioinformatics/btw321](https://doi.org/10.1093/bioinformatics/btw321)

