LoRDEC: a tool for correcting errors in long sequencing reads

Eric Rivals

LoRDEC: a tool for correcting errors in long sequencing reads

Eric Rivals

21st August 2015

Abstract for keynote presentation at "German Conference in Bioinformatics", Dortmund, Germany

Address:
Laboratoire d'Informatique de Microélectronique et de Robotique de Montpellier (LIRMM)
and Institute of Computational Biology (IBC)
CNRS and Université de Montpellier, France – http://www.lirmm.fr/řivals/

High-throughput DNA/RNA sequencing is a routine experiment in molecular biology and life sciences in general. For instance, it is increasingly used in the hospital as a key procedure of personalized medicine. Compared to the second generation, third generation sequencing technologies produce longer reads with comparatively lower throughput and higher error rate. Those errors include substitutions, indels, and they hinder or at least complicate downstream analysis like mapping or de novo assembly. However, these long read data are often used in conjunction with short reads of the 2nd generation.

I will present a hybrid strategy for correcting the long reads using the short reads that we introduced last year. Unlike existing error correction tools, ours, called LoRDEC, avoids aligning short reads on long reads, which is computationally intensive. Instead, it takes advantage of a succinct graph to represent the short reads, and compares long reads to paths in the graph. Experiments show that LoRDEC outperforms existing methods in running time and memory while achieving a comparable correction performance. It can correct both Pacific Biosciences and MinION reads from Oxford Nanopore.

LoRDEC is available at http://atgc.lirmm.fr/lordec; joint work with L. Salmela and A. Makrini.