



LoRDEC: a tool for correcting errors in long sequencing reads

Eric Rivals

► To cite this version:

Eric Rivals. LoRDEC: a tool for correcting errors in long sequencing reads. Montpellier Omics Days, Université de Montpellier, Master BCD, Feb 2016, Montpellier, France. lirmm-01446431

HAL Id: lirmm-01446431

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-01446431>

Submitted on 25 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LoRDEC: a tool for correcting errors in long sequencing reads

Montpellier Omics Days 2016

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

CNRS et Université de Montpellier, France

High-throughput DNA and RNA sequencing has become a routine experiment in molecular biology and life sciences in general. 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 tool, ours, called LoRDEC, avoids the alignment of short reads on long reads, which is computationally very intensive. Instead, it takes advantage of a succinct graph to represent the short reads, and compares the 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. In conclusion, i will comment on the impact of read error correction.

Work in collaboration with L. Salmela from University of Helsinki. LoRDEC is available at: <http://www.lirmm.fr/~rivals/lordec/>

