

NextDenovo & NextPolish Best Assembly Tool for ONT Platform

The Road to One-Contig Chromosome Assembly

INTRODUCTION

With the continual improvement of third generation sequencing technology (TGS), the throughput and read length increase dramatically, especially on Nanopore sequencing platforms.

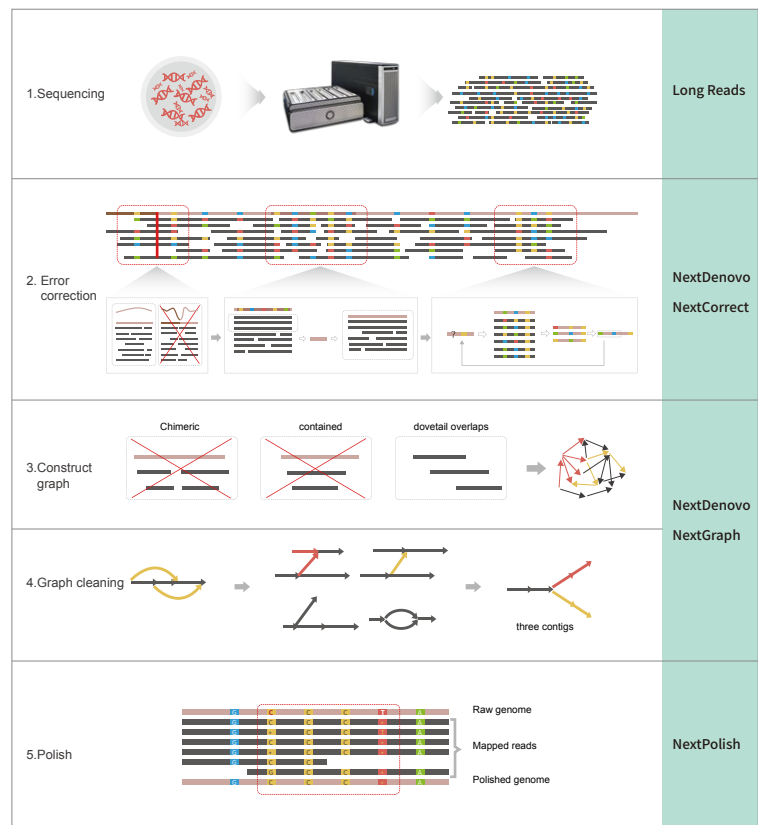
The major challenge of TGS nowadays is how to fully leverage ultra long reads with novel or optimized algorithm to achieve the high-quality and one-contig assembly of a chromosome, which is the milestone of genome assembly.

NextDenovo and NextPolish, developed by GrandOmics, are novel tools of correction, assembly and polish specifically for TGS long reads.

NextDenovo dramatically improves the assembly quality and reduce the computing resources usage by jointing correction and assembly. Additionally, NextPolish can improve the assembly quality further by fast polishing of draft assembly. Here we present the performance of NextDenovo and NextPolish by assembly from 273X long reads of Rice 9311. The final assembly has 18 Contigs (Oryza sativa L. 2n=24) and Contig N50 is 29.43M. It indicates half of 12 chromosomes are one-contig chromosomes.

In order to assess accuracy of genome, Illumina whole-genome sequence data was aligned to the assembly and sequence differences were estimated to yield a genomic accuracy of 99.99%^[1]. Further, BUSCO evaluation is applied on the assembly and about 98.1% full gene elements are found, which indicates the assembly is reliable and better than other assemblies of Rice 9311.

METHODS



<https://github.com/Nextomics>

RESULTS

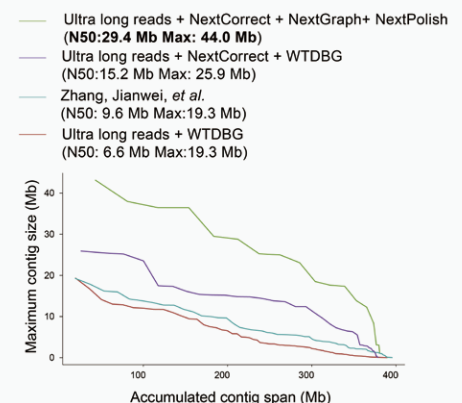
Rice 9311 Sequencing Metrics

Sample	Raw reads (bp)	Raw reads_ number	Clean reads (bp)	Clean reads_ number	Clean reads_ Depth (X)
9311	109,358,561,918	4,873,030	109,329,718,937	4,830,019	273.32

Rice 9311 Assembly Metrics

Contig N50 (bp)	Contig Number	Longest Contig (bp)	Total Length (bp)	Genomic accuracy (%)	BUSCO (%)
29,433,936	18	44,026,398	389,046,306	99.99	98.1

Comparison of various assemblies of Rice 9311



Reference

[1]. Gordon D, et al. Long-read sequence assembly of the gorilla genome. Science. 2016;352 6281:aae0344.

[2]. Zhang, J. et al. Building two indica rice reference genomes with PacBio long-read and Illumina paired-end sequencing data. Sci. Data 3:160076 doi: 10.1038/sdata.2016.76 (2016).