De Novo Assembly of Mitochondrial Genomes from Low Coverage Whole-Genome Sequencing Reads

Fahad Alqahtani, Ion Mandoiu
Computer Science and Engineering Department - University of Connecticut
Storrs, CT 06269, USA
{fahad.alqahtani,ion}@engr.uconn.edu

Advances in sequencing technology makes it feasible to generate large amounts of DNA sequencing data quickly and inexpensively. However, accurately assembling genomes de novo remains challenging. In this work, we present a scalable bioinformatics pipeline for de novo assembly of mitochondrial genomes sequences from low coverage whole-genome sequencing (WGS) data. By taking advantage of the fact that cells contain many more copies of mitochondrial genomes compared to the nuclear genome, we develop a k-mer coverage based read classifier that selects a subset of reads highly enriched in mitochondrial sequences. Experiments on low-coverage WGS reads from multiple species show that de novo assembly of reads using Geneious routinely results in nearly-perfect circular mitochondrial contigs.