

Algorithmic challenges in emerging fields of “complete” genomics and metagenomics

Abstract. Even though the first draft human genome was generated two decades ago, many important regions (such as centromeres) represented the *terra incognita* of the genome sequencing until recently. The long-read DNA sequencing technologies promise to transform genome sequencing toward generating complete genomes and lead to a critical demand for fast and accurate genome assembly algorithms. They also raised novel algorithmic problems that represent the key bottlenecks for the ongoing Big Data sequencing projects, such as the Vertebrate Genomes Project (aimed at sequencing genomes of all vertebrates) and the Human Pangenome Project (aimed at population-scale sequencing of complete human genomes). I will describe the new algorithmic challenges in genome assembly, discuss the recently developed algorithms for their solution, and show how these algorithms bring us closer to the goal of complete genome and metagenome sequencing.