MITGARD: an automated pipeline for mitochondrial genome assembly based on RNA-seq data

Pedro Gabriel Nachtigall, Felipe Gobbi Grazziotin, Inácio L.M. Junqueira-de-Azevedo

Laboratório Especial de Toxinologia Aplicada (LETA), Instituto Butantan, São Paulo, Brazil

Abstract

In Metazoa, mitochondrial genes are the most commonly used markers for molecular species determination and phylogenetic studies due to their extremely low rate of recombination, maternal inheritance, ease of use and fast substitution rate in comparison to nuclear DNA. In this sense, the assembly of the mitochondrial genome is an important step to proceed with rapid species identification in biodiversity surveys as also a key tool to identify hidden lineages or cryptic species. Over the past decade, the emerging field of next-generation sequencing (NGS) has seen dramatic advances in methods and a decrease in costs. Consequently, we noticed a big expansion on data being generated by NGS, most of them from RNA-seq experiments aiming at different objectives. Since mitochondrial genes are expressed at different levels in the majority of animal tissues, mRNA sequences are usually co-sequenced within the target transcriptome, generating a sequence data that is commonly underused or discarded. Then, the design of a computational pipeline that can be easily and automated applicable to assembly mitochondrial genomes from RNA-seq data is a valuable tool in the constant expansion of high-throughput data generation. Here, we present MITGARD, an automated pipeline that reliably recovers and assembles the mitochondrial genome from RNA-seq data from various sources. MITGARD was developed using Python and third-party tools, by taking the RNA-seq data as input and confident mitochondrial genome assembly as output. The preliminary results, using RNA-seq data from venom glands of several species from the snake genus Bothrops, showed that MITGARD reliably assembled the mitochondrial genomes, which could be used in alignments and construction of phylogenetic trees. Our assemblies together with available mitochondrial genomes resulted in confident phylogenetic inferences in snake species. In this sense, MITGARD is a helpful approach to studies focusing to assemble and annotate the mitochondrial genome and/or sequences of specific mitochondrial genes that can be used for species identification and evolutionary studies.