Strategies and difficulties in assembling highly recombinogenic plant organelle genomes: a case study

Concita Cantarella[1], Rachele Tamburino[2], Nunzia Scotti[2], Teodoro Cardi[1], Nunzio D'Agostino[1].
[1] *Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Centro di ricerca per l'orticoltura (CREA-ORT)*
[2] *Consiglio Nazionale delle Ricerche, Istituto di Bioscienze e BioRisorse (CNR-IBBR)*
nunzio.dagostino@entecra.it

Abstract

Mitochondrial genomes in plants are larger and more complex than in other eukaryotes due to their recombinogenic nature as widely demonstrated. The mitochondrial DNA (mtDNA) is usually represented as a single circular map, the so-called master molecule. This molecule includes repeated sequences, some of which are able to recombine, generating sub-genomic molecules in various amounts, depending on the balance between their recombination and replication rates. Recent advances in DNA sequencing technology gave a huge boost to plant mitochondrial genome projects. Conventional approaches to mitochondrial genome sequencing involve extraction and enrichment of mitochondrial DNA, cloning, and sequencing. Large repeats and the dynamic mitochondrial genome organization complicate *de novo* sequence assembly from short reads. The PacBio RS long-read sequencing platform offers the promise of increased read length and unbiased genome coverage and thus the potential to produce genome sequence data of a finished quality (fewer gaps and longer contigs). However, recently published articles revealed that PacBio sequencing is still not sufficient to address mtDNA assembly-related issues.

Here we present a preliminary hybrid assembly of a potato mtDNA based on both PacBio and Illumina reads and debate the strategies and obstacles in assembling genomes containing repeated sequences that are recombinationally active and serve as a constant source of rearrangements.

Acknowledgements

This work was funded by Ministry of Education, University and Research, PON Project GenHORT (PON02_00395_3215002) and was supported by the PON R&C 2007-2013 grant funded by the Italian Ministry of Education, University and Research in cooperation with the European Funds for the Regional Development (FESR).