M-100

Comparative genome analysis of soybean germplasms by optical mapping

*Ting-Fung Chan**, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China
Alden King-Yung Leung, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China
Claire Yik-Lok Chung, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China
Eugene Yui-Ching Chow, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China
Nana Jin, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China
Hon-Ming Lam, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China

Optical mapping is an imaging technique for capturing enzymatic patterns along DNA molecules of hundred-kilobase scale. It has been applied to different genomic applications including genome assembly, microbial strain typing, and structural variation detection. Recently, high-throughput technology for optical mapping data generation raises interest on its applications in comparative genomics. Previous studies using optical mapping in comparative genomics mainly focuses on detecting structural variations between two genome assemblies, or between multiple genome assemblies together with a reference. However, result interpretation in the former becomes difficult when three or more genomes are analyzed. While the latter is able to analyze more genomes, it is limited by the quality and completeness of the reference. In addition, both approaches cannot handle well multiple types of variations within the same region, or regions frequently confounded by repetitive elements, inversions or other large-scale rearrangements. Such regions are usually variation hotspots that carry important biological meanings.

We introduce an analysis framework using a reference-free approach based on multiple alignment of optical map contigs. Segments of DNA molecules with signal patterns conserved among samples are grouped as collinear blocks. Samples with higher similarity as sharing more conserved regions are clustered together, providing hints for phylogeny. Blocks contributing to most variations between sample clusters are found by statistical analysis. Further annotation reveals the identities of these blocks as the most representative features. We anticipate this method can facilitate comparative genome analysis on soybean, which has a high degree of diversity between different germplasms.