LTC software for physical mapping: contig assembly, MTP selection and verification of clone overlaps at sequence level

1,2Zeev Frenkel, 2Vladimir Glikson, 1Dina Raats, 3Bujie Zhan, 4Etienne Paux, 1Tzion Fahima, 4Catherine Feuillet, 3Simen Rød Sandve, and 1Abraham Korol

1University of Haifa, Institute of Evolution, Haifa, Israel
2MultiQTL Ltd, Haifa, Israel
3Norwegian University of LifeSciences, Department of Plant and Environmental Sciences
4INRA, Genetics, Diversity and Ecophysiology of Cereals, Clermont-Ferrand, France

With the advent of NGS technologies, genome sequencing through whole genome shotgun approaches has become more and more affordable. However, for large and complex genomes, such as the one of wheat, this approach is still hampered by polyploidy and the presence of repeats that cause problems during sequence assembly. For such genomes, the BAC-by-BAC approach remains the best option to achieve a high quality reference sequence that can be used for positional cloning, structural and functional analyses. This approach relies on the construction of physical maps that are anchored to genetic maps. To date, FPC has been the most commonly used software to construct physical maps. Recently, we developed an analytical system for contig assembly called Linear Topology Contig (LTC) as an alternative or complementary tool to FPC.

Algorithms implemented in LTC provide more effective detection of problematic clones and clone overlaps and better ordering of clones within contigs. These features reduce the risks associated with the problems of chimerical contigs classically encountered with FPC. The effectiveness of LTC was checked on real data of different species including wheat and on simulated data based on sequenced genomes of maize and rice. Comparing LTC and FPC physical maps demonstrates that LTC can produce longer and more reliable contigs than FPC. For example, for wheat 1BS chromosome arm, average contig length (kb) and N50 (kb) were 1076 and 2430 for LTC contigs, 4564 and 8515 for LTC scaffolds, and 531 and 1033 for FPC contigs. For simulated data, the proportion of clones found in chimerical LTC contigs was much lower than in chimerical FPC contigs (e.g., 3% and 62% for LTC and FPC, respectively, in the case of simulated coverage ~10x and simulated high proportion=11% of chimerical clones). LTC is also able to overcome problems caused by presence of large sets of highly overlapped clones and enables working with large BAC libraries (e.g., ~600,000 BACs).

In addition to de novo contig assembly and minimal tilling path (MTP) selection, LTC enables curing gaps, reviewing, reordering, and elongating of contigs and MTPs obtained by standard FPC package. LTC can utilize available marker and sequence information in manual editing, verifying, and merging of contigs and MTPs. LTC includes tools for simplification of positive clone identification (during the anchoring procedure) based on the list of positive pools. LTC also can be used for testing the quality of clone sequencing, checking overlaps of MTP clones at sequence level and suggesting solutions to repair the observed gaps.