# **HiC-TE: a pipeline for HiC data analysis in the context of repeats and genome organization**

Matej Lexa1,2, Monika Cechova1, Son Hoang Nguyen1, Pavel Jedlicka2, Zdenek Kubat2, Roman Hobza2, Eduard Kejnovsky2

1 *Faculty of Informatics, Masaryk University, Brno, Czech Rep*

2 Institute of Biophysics, Czech Academy of Sciences, Brno, Czech Rep

High-throughput chromosome conformation capture (Hi-C) has become a well established sequencing-based method to detect physical proximity of DNA segments in nuclei, with thousands of Hi-C experiments now available in public repositories (e.g. NCBI Sequence Read Archive (SRA)). We combined this data with increasingly precise public plant reference sequences and tools that characterize the repetitive fraction of genomes, such as Tandem Repeat Finder, PlantSat database, TE-greedy-nester or Repeat Explorer 2. We built a Nextflow pipeline with two main inputs, a SRA sequencing run ID and a corresponding reference genome. The pipeline processes the Hi-C reads, maps and clusters them to identify Hi-C pairs that can be attributed to specific repeat classes, quantifying them within and between repeat families. Results of the analysis are conveniently visualized as heatmaps, circular chromosome plots and exported as data tables. The pipeline is available for use via a gitlab repository (https://gitlab.fi.muni.cz/lexa/hic-te). First experiments show biologically important interactions of ribosomal DNA clusters or centromeric repeats that are clearly visible in most plant species. We discovered that LTR retrotransposon families with high interaction rates are often species-specific. This pipeline represents a novel and reproducible way to analyze the role of repetitive elements in the 3D organization of genomes.