
GraffiTE: a Unified Framework to Analyze Transposable Element Insertion Polymorphisms using Genome-graphs

Cristian Groza¹, Xun Chen², Travis Wheeler³, Guillaume Bourque⁴, and Clément Goubert*³

¹Quantitative Life Sciences, McGill University – Montréal, Canada

²Kyoto University – Kyoto, Japan

³University of Arizona – Tucson, Arizona, United States

⁴McGill University – Montréal QC, Canada

Abstract

Transposable Elements (TEs) are abundant and mobile repetitive DNA sequences evolving within and across their hosts' genomes. Active TEs cause insertion polymorphism and contribute to genomic diversity. Here, we present GraffiTE, a flexible and comprehensive pipeline for detecting and genotyping polymorphic mobile elements (pMEs). By integrating state-of-the-art SV detection algorithms and graph-genome frameworks, GraffiTE enables the accurate identification of pMEs from genomic assemblies and long-read as well as the precise genotyping of these variants using short-or long-read data. Performance evaluations using simulated and benchmark datasets demonstrate high precision and recall rates. Notably, we demonstrate the versatility of GraffiTE by analyzing the human reference pangenome, 30 *Drosophila melanogaster* genomes, and multiple cultivars of the emerging crop model *Cannabis sativa*, where pMEs are undocumented. These analyses reveal the landscapes of pMEs and their frequency variations across individuals, strains, and cultivars. GraffiTE provides a user-friendly interface, allowing non-expert users to perform comprehensive pME analyses, including in models with limited TE prior knowledge. The pipeline's extensible design and compatibility with various sequencing technologies make it a valuable integrative framework for studying TE dynamics and their impact on genome evolution. GraffiTE is freely available at <https://github.com/cgroza/GraffiTE>.

Keywords: Transposable elements, insertion polymorphism, graph, genomes, structural variation, genotyping, genome evolution

*Speaker