# dnaPipeTE, a new bioinformatic pipeline to assemble, annotate estimate abundance and dynamics of repetitive DNA in low coverage sequencing: application to Aedes mosquitoes

Clément Goubert[*1], Laurent Modolo , Cristina Vieira , Claire Valiente Moro , Patrick Mavingui , and Matthieu Boulesteix

[1]UMR CNRS 5558 Laboratoire de Biométrie et de Biologie Evolutive (LBBE) – CNRS : UMR5558, Université Claude Bernard - Lyon I, VetAgro Sup, L'Institut National de Recherche en Informatique et e n Automatique (INRIA) – France

**Résumé**

Repetitive DNA, including transposable elements (TEs), is found throughout eukaryotic genomes. Annotating and assembling this "repeatome" during genome-wide analysis often poses a challenge. To address this problem, we present dnaPipeTE - a new bioinformatics pipeline that uses a small amount of raw genomic reads. It produces precise estimates of repeated DNA content and TE consensus sequences, as well as an overview of the relative ages of TE families. dnaPipeTE can annotate and quantify repeats in any genome, using very low coverage sequencing as input ($<$ 0.5X) and works without reference genome assembly. We applied this pipeline to the genome of the Asian tiger mosquito, Aedes albopictus, an invasive species of human health interest, for which the genome size is estimated to be over 1 Gbp but whose sequence has not been released yet in spite of multiple ongoing projects. Using dnaPipeTE, we showed that this species harbours a large (50% of the genome) and potentially active repeatome with an overall TE class composition similar to that of Aedes aegypti, the yellow fever mosquito. However, intra-class dynamics shows clear distinctions between the two species, with differences at the TE family level that are compatible with the theory of genome ecology. Our pipeline's ability to manage the repeatome annotation problem will make it helpful for new or ongoing assembly projects, and our results will benefit future genomic studies of Ae. albopictus.

---

[*]Intervenant