

**Lukasz Olczak**

INSTITUTE OF INFORMATICS, SILESIA UNIVERSITY OF TECHNOLOGY

e-mail: Lukasz.Olczak@polsl.pl

**Rafał Pokrzywa**

INSTITUTE OF INFORMATICS, SILESIA UNIVERSITY OF TECHNOLOGY

**Andrzej Polaski**

INSTITUTE OF INFORMATICS, SILESIA UNIVERSITY OF TECHNOLOGY

## **Mathematical model of tandem repeat evolution based on comparisons of Homo sapiens and Homo neanderthalensis genomes**

Tandem repeats are genomic markers well suited for studying evolutionary scenarios for closely related species, due to their high mutation rates. There are many studies concerned with fitting evolutionary models to data on short tandem repeats with conclusions leading to estimates of parameters of tandem repeats mutation process, evolutionary and demographic scenarios of different species and populations etc.

In this talk we present coalescence based mathematical model of evolution of tandem repeats based on comparison of genomes of homo sapiens and Homo neanderthalensis. In the coalescence model we assume the deterministic moment of speciation event leading to Homo sapiens and Homo neanderthalensis species. The results of the coalescence model of evolution are probability distributions of differences between numbers of repeats in two species. These probability distributions depend on parameters, mutation intensities, different for models for evolution of loci with different motif length.

The obtained models are then fitted to data on locations and structures of tandem repeat loci of homo sapiens and Homo neanderthalensis genomes obtained by using the recently developed genome browsing tool BWtr and the appropriately designed alignment algorithm. Due to imperfections of the assembly process for Homo neanderthalensis genome the model with censored observations is applied and the appropriate EM procedure is designed.

Estimates of mutations rates for different sizes of repeat motifs are compared to results of other population dynamics studies. Possible sources of biases in different approaches are highlighted and possible future improvements of the developed model are presented.