



**28**  
**MAR.**  
2019

🕒 de 13h à 14h

## SÉMINAIRE

# Some contributions to the estimation of genetic distances between populations

**Tristan Mary Huard**

INRA-Moulon

We consider the problem where one wants to evaluate the level of divergence between  $K$  populations. Each population is characterized by its allelic frequency profile, where allelic frequencies are assumed to be estimated from a sample at several (typically thousands/millions of) markers. In this context the  $F_{ST}$  is a widely used criterion for the quantification of the divergence between two populations, that can also be adapted to the question of detecting genomic regions that exhibit a divergence level substantially higher than the rest of the genome. Still, the concept of  $F_{ST}$  remains ambiguous - with different available definitions assumed to be "connected" in some sense - and the strategy to estimate the  $F_{ST}$  when there are more than 2 populations is still an open question, the most popular strategy being to consider all possible pairs of population successively. In this presentation we will first propose a hierarchical model for the history of population divergence and show that the two classical definitions of the  $F_{ST}$  (as provided by Hudson and Weir & Cockerham) actually measure independent quantities. We will then provide an estimation procedure based on the moment estimators suggested by Bhatia (in the case of 2 populations) and show how both the  $F_{ST}$  components and the history of population divergence may be jointly estimated. Lastly, we will consider the problem of detecting genomic regions under selection and provide a segmentation procedure for the identification of such regions. Both the estimation and the segmentation procedures will be illustrated on the 1KG human genome dataset that gathers several human populations sampled over the world.