



**17**  
**JAN.**  
2013

🕒 de 11h à 12h

## SÉMINAIRE

# Reconstruction of ancestral genome organization in bacteria

**Murray Patterson**

LBBE

**Background.** Recovering the structure of ancestral genomes can be formalized in terms of properties of binary matrices such as the Consecutive-Ones Property (C1P). The Linearization Problem asks to extract, from a given binary matrix, a maximum weight subset of rows that satisfies such a property. This problem is in general intractable, and in particular if the ancestral genome is expected to contain only linear chromosomes or a unique circular chromosome. In the present work, we consider a relaxation of this problem, which allows ancestral genomes that can contain several chromosomes, each either linear or circular. **Result.** We show that, when restricted to binary matrices of degree two, which correspond to adjacencies, the genomic characters used in most ancestral genome reconstruction methods, this relaxed version of the Linearization Problem is polynomially solvable using a reduction to a matching problem. This result holds in the more general case where columns have bounded multiplicity, which models possibly duplicated ancestral genes. We also prove that for matrices with rows of degrees 2 and 3, without multiplicity and without weights on the rows, the problem is NP-complete, thus tracing sharp tractability boundaries. I also give a preliminary result on a method for generating these binary matrices of degree 2, i.e., a method for inferring ancestral adjacencies, and how the Linearization Problem fits into this larger context. **Conclusion.** As it happened for the breakpoint median problem, also used in ancestral genome reconstruction, relaxing the definition of a genome turns an intractable problem into a tractable one. The relaxation is adapted to some biological contexts, such as bacterial genomes with several replicons, possibly partially assembled. Algorithms can also be used as heuristics for hard variants. More generally, this work opens a way to better understand linearization results for ancestral genome structure inference.