

## Enhancing Gibbs sampling method for motif finding in DNA with initial graph representation of sequences

---

ŽIVA STEPANČIČ

*Arctur d.o.o.*

*Nova Gorica (Slovenia)*

`ziva.stepancic@gmail.com`

Finding short patterns with residue variation in a set of sequences is still an open problem in genetics, since motif finding techniques on DNA and protein sequences are inconclusive on real data sets and their performance varies on different species. Hence finding new algorithms and evolving established methods are vital to further understand genome properties and the mechanisms of protein development. In this talk we present an approach to search for possible motifs in connection to Gibbs sampling method. Starting points in the search space are partly determined via graphical representation of input sequences opposed to completely random initial points with the standard Gibbs sampling. Our algorithm is evaluated on synthetic as well as on real data sets by using several statistics, such as sensitivity, positive predictive value, specificity, performance and correlation coefficient. Additionally, a comparison between our algorithm and basic standard Gibbs sampling algorithm is made to show improvement in accuracy, repeatability and performance.