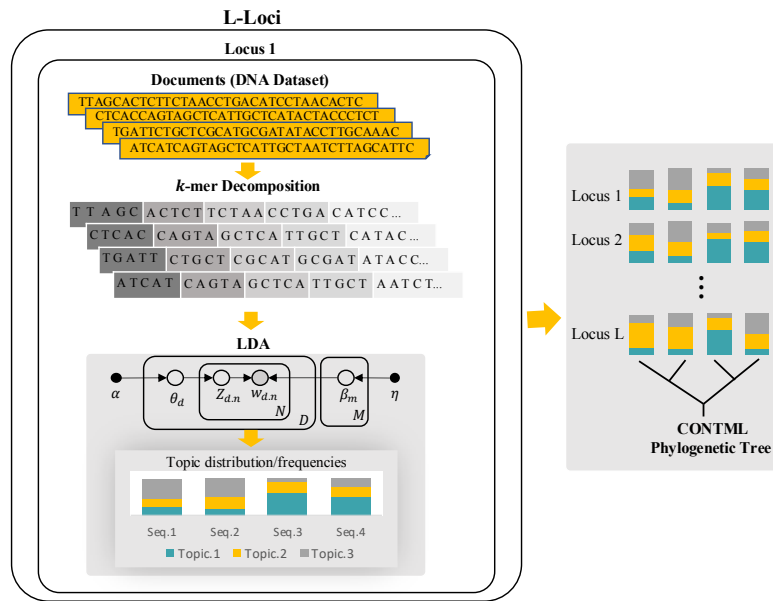


# Oct 30, 2024 Colloquium with Tara Khodaei

- **Date:** Oct 30, 2024
- **Name:** Tara Khodaei
- **University/Affiliation:** Postdoctoral Research Associate, Scientific Computing, FSU
- **Talk Title:** TOPICCONTML: Genome-wide Phylogeny Reconstruction Using Topic Modeling
- **Abstract:** Inferring the evolutionary history of species or populations with genome-wide data is gaining ground, but computational constraints still limit our abilities in this area. We developed an alignment-free method, TOPICCONTML, to infer the genome-wide species tree. The method uses probabilistic topic modeling, an unsupervised machine learning method inspired by natural language processing, to extract topic frequencies from k-mers derived from multilocus DNA sequences, which are then used by Contml to generate a species tree. The approach operates in two phases: (1) The multi-locus or genome-wide data is broken down into k-mers; these k-mers are then used to learn a probabilistic topic model and extracts the topic frequencies of sequences using Latent Dirichlet Allocation (LDA) model for each locus. (2) These topic frequencies from multiple loci are then used to estimate a phylogeny with in the PHYLIP package. We evaluate our method with different biological datasets.
- **Picture/Graphic:**



TOPICCONTML workflow