The Metastructure Model

Zachary King   Ali Ebrahim

August 19, 2015

A Metastructure model is:

- A genomic knowledge base
- A collection of genome-oriented annotations that include:

    - Genes
    - Transcription units (TU)
    - Transcription start sites (TSS)
    - Ribosome binding sites (RBS)
    - (Translation pause sites)
    - Sigma factor binding sites
    - Transcription factor binding sites

- A self-consistent representation of the genome with rules for consistency

The Metastructure model includes the content of a traditional genome annotation, but it has a different purpose. Eventually, the Metastructure model should include all the knowledge about a genome that has positional information.

Experimental error is considered in the model by assigning a confidence interval for each locus in the model.

A Metastructure model is constructed using a reference genome annotation and omics datasets. A workflow can be described for this process. It includes a number of manual and automated steps.

Rules for consistency are defined. These rules identify groups of features whose locations are not consistent with our general understanding of the genome. The inconsistencies identified by these rules require manual curation. They lead to improvements in the model and improvements in our understanding of genome structure and organization.