10:50 *Janaka Edirisinghe*, *José Faria*, *Filipe Liu*, *Joana Xavier*, *Samuel Seaver*, *Pamela Weisenhorn*, *James Jeffryes*, *Tian Gu*, *Qizh Zhang*, *Isabel Rocha* and *Christopher Henry*

Automated pathway curation and improving metabolic model reconstruction based on phylogenetic analysis of pathway conservation

SPEAKER: *Janaka Edirisinghe*

ABSTRACT. Metabolic models generated by automated reconstruction pipelines are widely used for high-throughput prediction of microbial phenotypes. However, the generation of accurate in-silico phenotype predictions based solely on genomic data continues to be a challenge as metabolic models often require extensive gapfilling in order to produce biomass. As a result, the true physiological profile of an organism can be altered by the addition of non-native biochemical pathways or reactions during the gapfilling process. In this study, we constructed draft genome-scale metabolic models for ~1000 diverse set of reference microbial genomes currently available in GenBank, and we decomposed these models into a set of classical biochemical pathways. We then determine the extent to which each pathway is either consistently present or absent in each region of the phylogenetic tree, and we study the degree of conservation in the specific steps where gaps exist in each pathway across a phylogenetic neighborhood. Based on this analysis, we improved the reliability of our gapfilling algorithms, which in turn, improved the reliability of our models in predicting auxotrophy. This also resulted in improvements to the genome annotations underlying our models. We validated our improved auxotrophy predictions using growth condition data collected for a diverse set of organisms. Our improved gapfilling algorithm will be available for use within the DOE Knowledgebase (KBase) platform (https://kbase.us).