**PanOryza – defining high-quality gene models in the rice pan-genome, using new data types**

*Bruno Contreras Moreira, Eshan Sharma, Shradha Saraf, Guy Naamati, Parul Gupta, Justin Elser, Dmytro Chebotarov, Kapeel Chougule, Zhenyuan Lu, Sharon Wei, Yong Zhou, Jianwei Zhang, Zhichao Yu, Sandeep Amberkar, Kawinnat Sue-Ob, Kenneth L. McNally, Doreen Ware, Rod Wing, Dario Copetti, Eric W Deutsch, Maria Martin, Pankaj Jaiswal, Sarah Dyer and* ***Andrew R Jones***

**Presenter contact details:**

**andrew.jones@liverpool.ac.uk**

University of Liverpool, Institute of Systems, Molecular and Integrative Biology, Biosciences Building, Crown Street, Liverpool. L69 7ZB UK

**Keywords:** Bioinformatics, databases, gene models, pan-gene set, proteomics, protein structures.

**Abstract**

There are multiple sets of plausible gene models for the Nipponbare reference rice genome, as well models for high-quality assemblies capturing the variation across domesticated Asian rice (*Oryza sativa*), such as the MAGIC-16 set of genomes. Our teams (University of Liverpool, EMBL-EBI, Oregon State, Arizona Genomics Institute, IRRI, KAUST, Cold Spring Harbour) have built a consistent and aligned pan-gene set, and assigned persistent identifiers to pan-genes. In the presentation, I will describe the profiles of pan-genes, depending on the genomes in which they are detected, using multiple types of experimental and computationally-derived data that have been aligned.

We are now also working towards generating a set of metrics for individual gene models to assess their quality, using experimental evidence, orthology and, via new methods, support from protein 3D structures to determine the best set of rice gene models. We will describe evidence supporting cases where gene models can be ranked and improved in the reference genome, and then propagated across the pan-gene set, to deliver data-driven curation of rice gene models.