**Harnessing natural variation through pan genome and pan proteome unravel the mysteries of seed quality traits in rice**

Kanika Narula, Pragya Nalwa, Atreyee Sengupta, Nisha and Subhra Chakraborty

National Institute of Plant Genome Research, New Delhi, India

Accessing the natural genetic diversity of crop plants unravels hidden genetic traits, elucidates gene functions and allows the expansion of bioeconomy impacting food security. Determining the drivers of natural variations in germplasm is critical for improving quality traits in crop plants. Protein deficiency affects physical growth and development, and increases morbidity and mortality in human. Rice (*Oryza sativa* L.), the most important food source that provides more than 50% of total calories and 20% dietary proteins consumed by half the world's population. However, rice grain quality is compromised with low grain protein content (GPC), essential amino acids (EAAs), high glycemic index, presence of some metabolites and inhibitory compounds that adversely affect protein digestibility and bioavailability. Rice has wide genetic base due to evolutionary history, domestication and modern breeding techniques. Availability of 3K panel has facilitated discovery of allelic variants, but it captures only fraction of genetic diversity in rice. Here, to examine scale of yet untapped functional biomolecule diversity linked to proteostasis, we generated proteo-metabolomic resource and merged it with genomic and transcriptomic data in rice. Next, we determined protein content of 1500 rice germplasm. Using graph-based clustering, we grouped these cultivars into geographical niche, gene neighbourhood distributions, sequence clusters and gene-protein-metabolite network associated with proteostasis. Furthermore, we illustrate storage protein mediated molecular mechanism of GPC variation by developing high protein rice with increased EAAs. Overall, our results uncover diverse functional space and characterize positive regulation of GPC and EAAs by exploring the dark matter for complex seed quality traits.

Keywords; Genetic diversity, Protein quality and quantity, gene-protein-metabolite network