Pan-genome analysis identifies complex genetic variation in cultivated and wild rice

With the support by the National Natural Science Foundation of China, a collaborative study by the research groups led by Prof. Huang XueHui (黄学辉) from Shanghai Normal University and Prof. Han Bin (韩斌) from the Institute of Plant Physiology and Ecology, Chinese Academy of Sciences constructed a pan-genome dataset of *O. sativa*-*O. rufipogon* species complex, which was published in *Nature Genetics* (2018, 50(2): 278—284).

Rice is one of the major crops in China. The rich genetic diversity in cultivated and wild rice serves as the main sources in rice breeding through continuously uncovering and utilizing new alleles. With the application of high-throughput sequencing technologies, many diverse rice accessions have been re-sequenced and phenotyped during recent few years. However, in these re-sequencing studies, the identifications of sequence variants all rely on high levels of sequence similarity to map the short reads onto the rice reference Nipponbare genome, in which the information from highly polymorphic regions would be often lost. Moreover, previous studies have found that there are functionally important genes absent in the reference Nipponbare genome but present in other rice varieties. Hence, one or a few rice genomes cannot include all important genomic contents.

Their group selected a total of 66 rice accessions for deep sequencing and whole-genome *de novo* assembling. Comparative analyses and genome annotations of the genome assemblies enabled the identification of diverse alleles and the functional consequences of various polymorphisms at the fine-scale level. Particularly, 10,872 genes absent in Nipponbare reference genome were identified, many of which have transcript evidence and well-characterized protein domains. They systemically investigated the whole set of coding genes using this pan-genome data, which reveals extensive presence and absence variation among rice accessions.

These findings provide new insights into the genetic diversity in cultivated and wild rice and will further promote evolutionary and functional studies in rice.

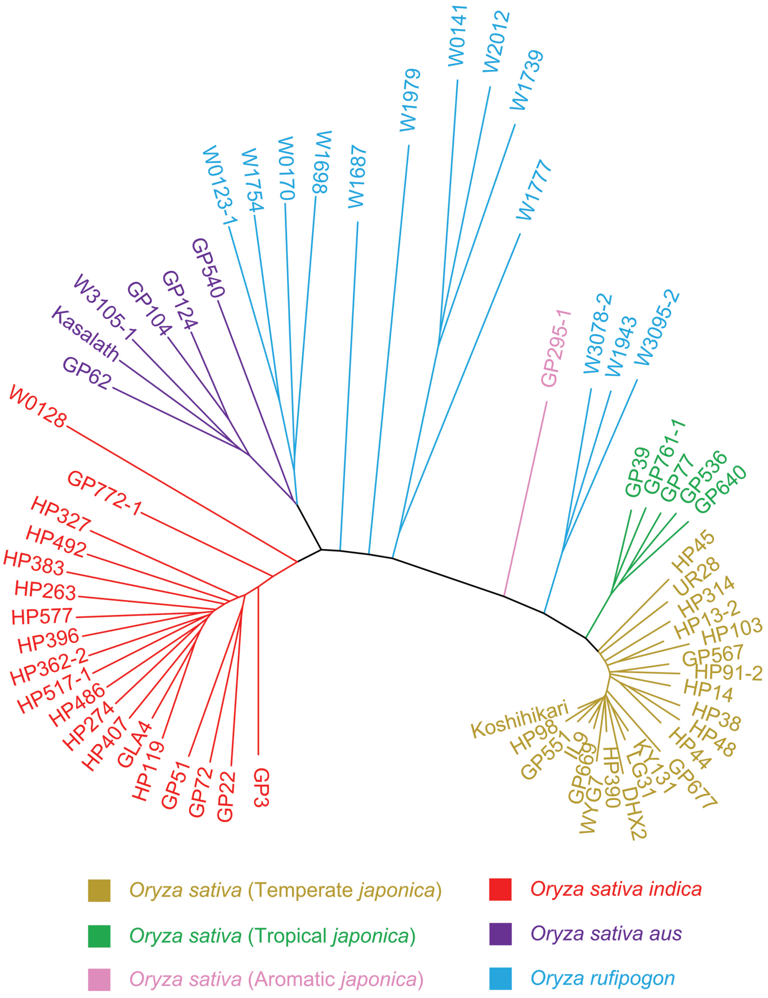

Figure Phylogenetic tree of the 66 rice accessions in the pan-genome data. The accessions within different groups are indicated by different colors.