Title:

Rice Pan-genome studies: How can bioinformatics findings be applied in the field?

Author's name:

Yong Zhou[1,2*], Keerthana Manickam[1], Manjula P Thimma[1], Zhichao Yu[3], Dmytro Chebotarov[4], Kapeel Chougule[5], Luis F. Rivera[1], Nagarajan Kathiresan[1], Ramil Mauleon[4], Waseem Hussain[4], Doreen Ware[5], Kenneth McNally[4*], Jianwei Zhang[3,*], Rod A Wing[1,2*]

Affiliation:

[1]Biological, Environmental Sciences and Engineering (BESE), KAUST, 23955, Saudi Arabia
[2]Arizona Genomics Institute (AGI), University of Arizona, Tucson, Arizona 85721, USA
[3]National Key Laboratory of Crop Genetic Improvement, HZAU, Wuhan 430070, China
[4]International Rice Research Institute (IRRI), Los Baños, 4031 Laguna, Philippines
[5]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
Email*: yong.zhou@kaust.edu.sa, K.McNally@irri.org, jzhang@mail.hzau.edu.cn, rod.wing@kaust.edu.sa

Keywords:

pan-genome, inversions, hidden variants, GWAS, field

Abstract:

Bioinformatics analysis of pan-genomes has significantly advanced our understanding of genetic diversity, large structural variation, and hidden variants. However, the application or transition of bioinformatics to the field remains to be fully explored. To facilitate this challenge, we built a pan-genomic resource from 16 platinum standard *O. sativa* reference sequences (a.k.a. PSRefSeqs) that represent the population genetic diversity of the 3K-RGP, and systematically explored variants that could be applied in the field. First, we built a pan-genomic rice inversion index to study large inversions (>100 bp) and identified 32 haplotypes of inversions that were significantly ($p < 10^{-4}$) associated with 11 traits. Second, we built a full SNP dataset of the 3K-RGP samples against the 16 PSRefSeqs and identified 27 QTLs for grain length, including four novel peaks. In addition, we discovered 2 million hidden SNPs that were not possible to study using a single reference genome alone (*i.e.* IRGSP-RefSeq), 11% of which were associated with gene function or expression. Hidden variants in the *Sub1A* gene were refined, and a total of 180 accessions were predicted to have potential submergence tolerance based on the presence of a favorable allele. These accessions will be tested this season in the field (IRRI) for submergence tolerance response, as a pilot study test our ability to transfer pangenome variant identification to the field. Our bioinformatics analysis demonstrates significant potential for transferring pan-genomic findings to the field, and we look forward to collaborative efforts to design optimal rice varieties for future climate and nutritional scenarios.