Genome-wide association studies of 14 agronomic traits in rice landraces

Xuehui Huang1,2,10, Xinghua Wei3,10, Tao Sang4,10, Qiang Zhao1,2,10, Qi Feng1,10, Yan Zhao1, Canyang Li1, Chuanrang Zhu1, Tingting Lu1, Zhiwu Zhang5, Meng Li3,6, Danlin Fan1, Yunli Guo1, Ahong Wang1, Lu Wang1, Liuwei Deng1, Wenjun Li1, Yiqi Lu1, Qijun Weng1, Kunyan Liu1, Tao Huang1, Taoying Zhou1, Yufeng Jing1, Wei Li1, Zhang Lin1, Edward S Buckler5,7, Qian Qian3, Qi-Fa Zhang8, Jiayang Li9 & Bin Han1,2

Uncovering the genetic basis of agronomic traits in crop landraces that have adapted to various agro-climatic conditions is important to world food security. Here we have identified ~3.6 million SNPs by sequencing 517 rice landraces and constructed a high-density haplotype map of the rice genome using a novel data-imputation method. We performed genome-wide association studies (GWAS) for 14 agronomic traits in the population of Oryza sativa indica subspecies. The loci identified through GWAS explained ~36% of the phenotypic variance, on average. The peak signals at six loci were tied closely to previously identified genes. This study provides a fundamental resource for rice genetics research and breeding, and demonstrates that an approach integrating second-generation genome sequencing and GWAS can be used as a powerful complementary strategy to classical biparental cross-mapping for dissecting complex traits in rice.

Here we have genotyped rice landraces through direct resequencing of their genomes by adopting sequencing-by-synthesis technology, which represents a step forward from the oligonucleotide array technology widely used for GWAS11-12. More than 500 diverse rice landraces, representing a large collection of rice accessions, were sequenced at approximately onefold genome coverage. The resulting data set captures more of the common sequence variation in cultivated rice than any other data set to date. Using a highly accurate imputation method, we constructed a high-density rice haplotype map and performed GWAS for 14 agronomic traits to identify a substantial number of loci potentially important for rice production and improvement. Some loci were mapped at close to gene resolution, indicating that GWAS of rice landraces could provide an effective approach for gene identification.

**RESULTS**

**Genome sequencing and SNP identification**

From a collection of ~50,000 rice accessions originating in China, we have undertaken an effort to build a large sample of morphologically, genetically and geographically diverse landraces for genetic studies. In this study, a total of 517 landraces were selected and comprehensively phenotyped (see Online Methods). We genotyped...