

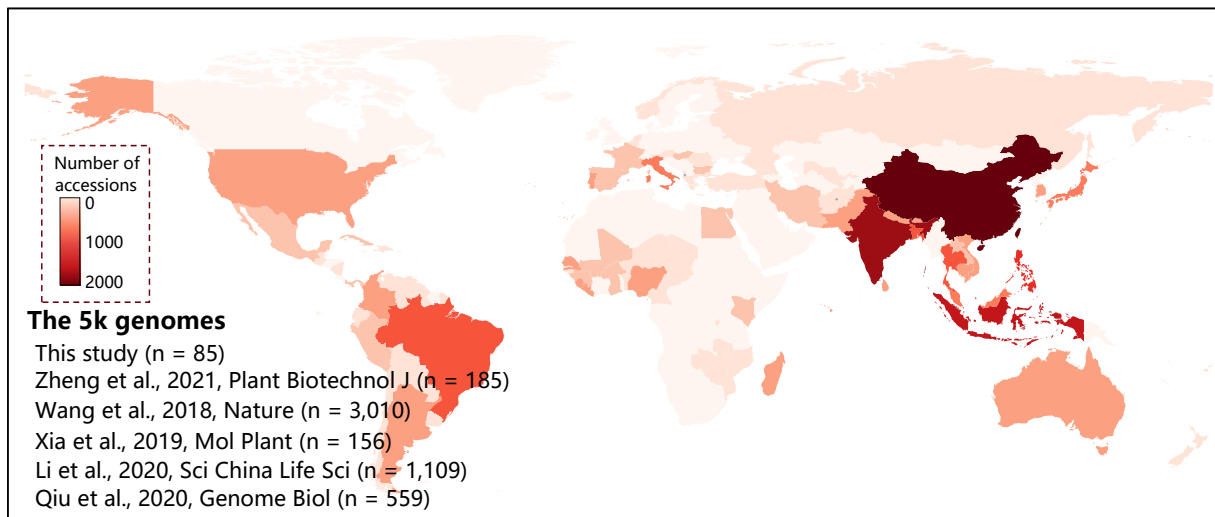
Plant Communications, Volume 5

Supplemental information

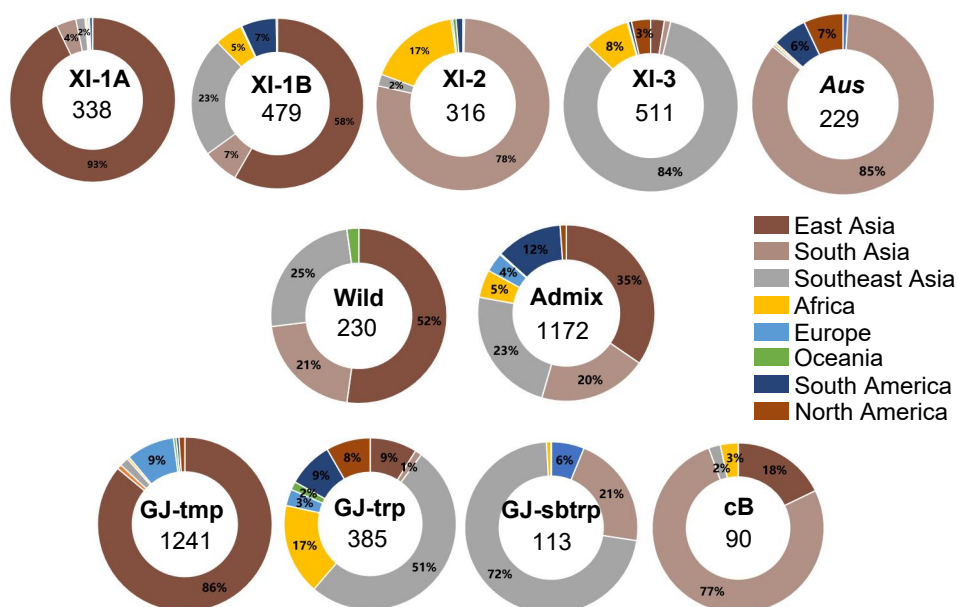
Population genomic analysis unravels the evolutionary roadmap of pericarp color in rice

Lingjuan Xie, Dongya Wu, Yu Fang, Chuyu Ye, Qian-Hao Zhu, Xinghua Wei, and Longjiang Fan

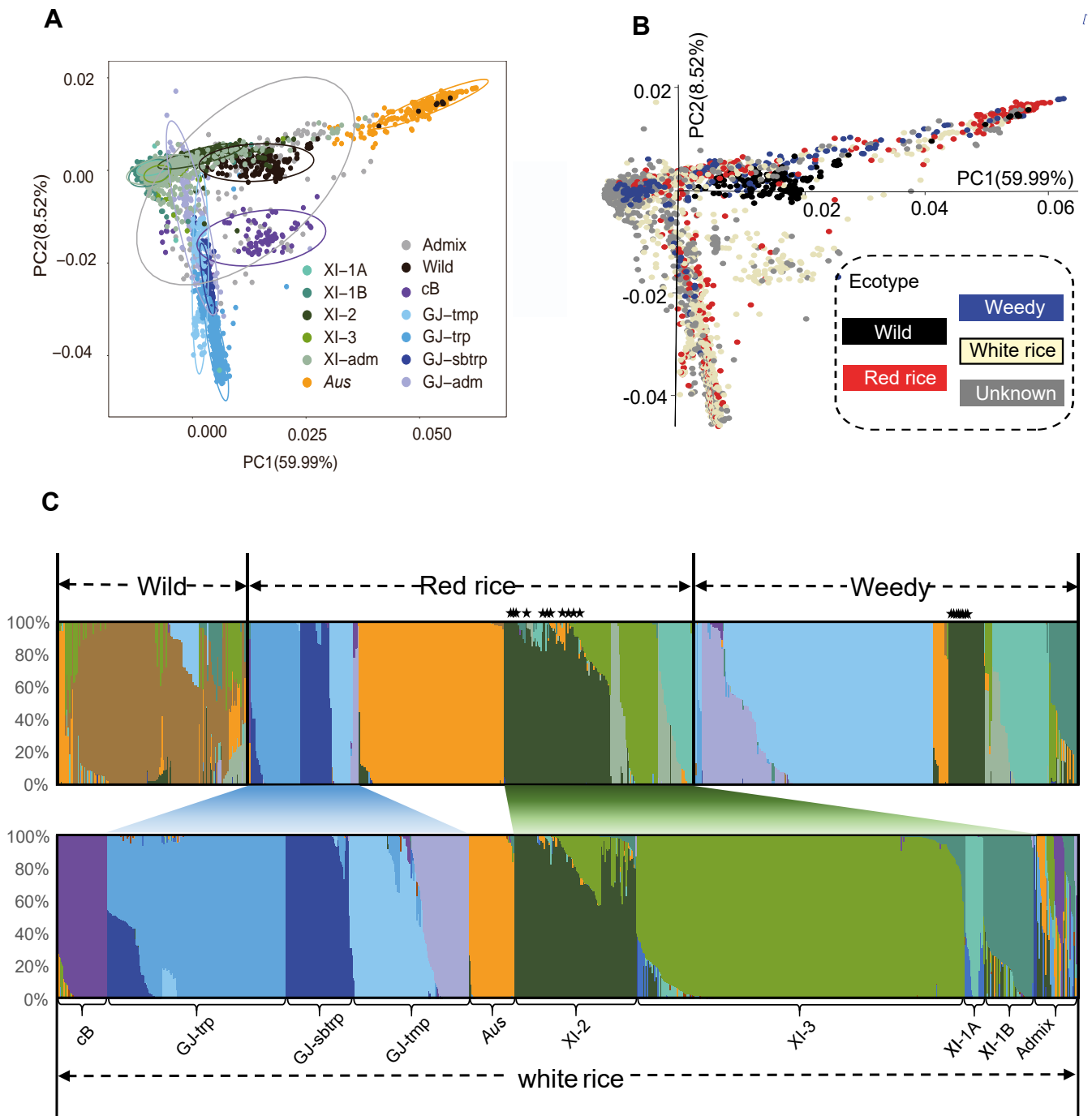
A



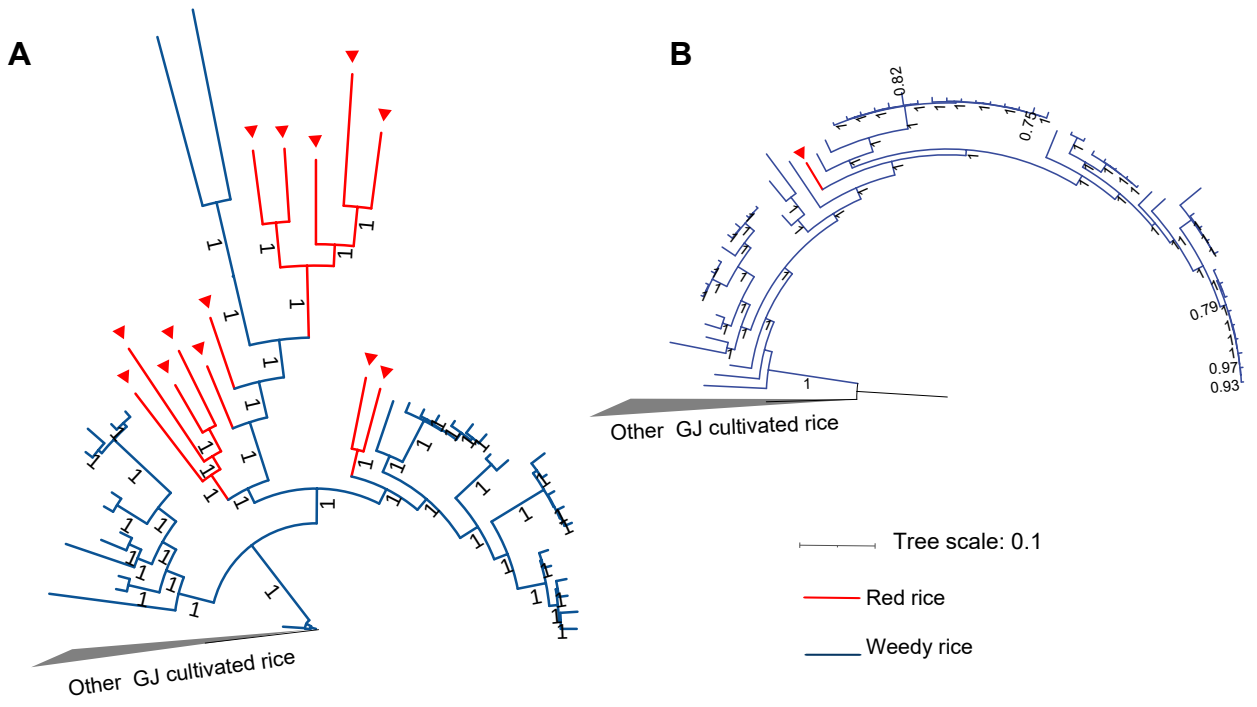
B



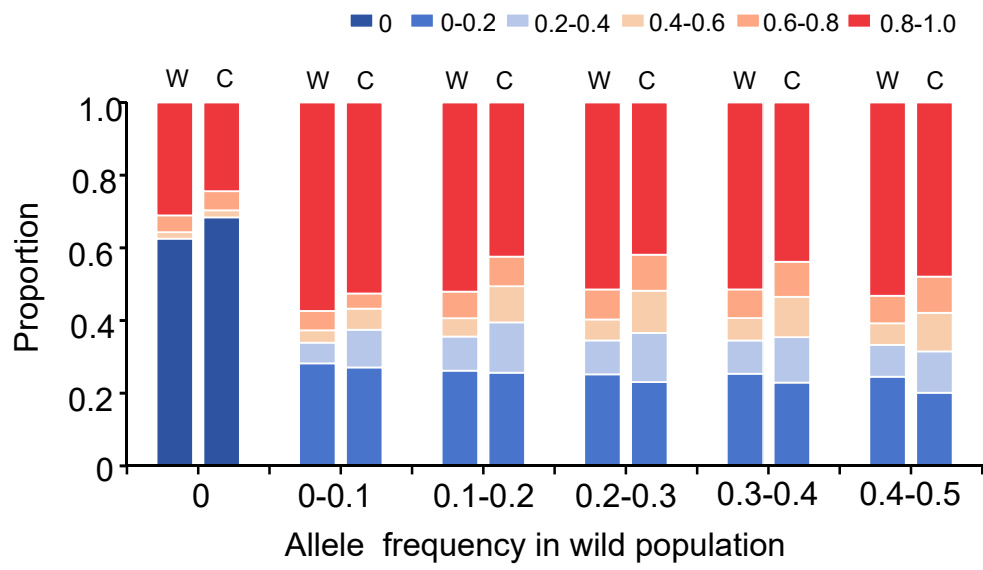
Supplementary Fig.1 Overview of sampling. (A) Geographical distribution and origin of 5,104 rice accessions (termed as 5k genomes) used for this study. (B) Proportion of geographic composition in each subspecies.



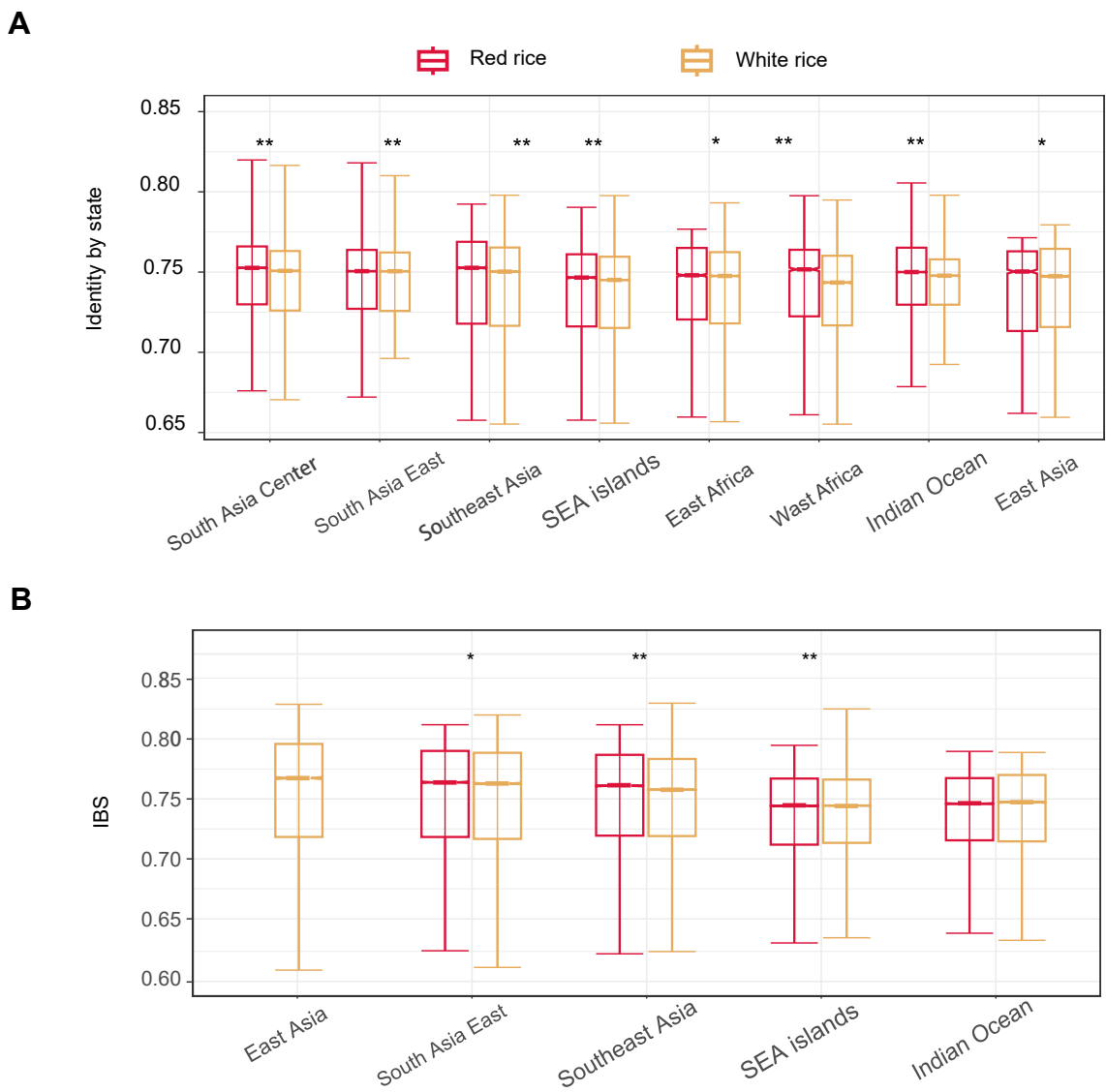
Supplementary Fig.2 Principal-component analyses and population structure of 5,104 accessions. (A) Rice accessions distinguished by subpopulations, including XI rice, GJ rice, aus rice, aromatic rice and wild rice. (B) Rice accessions distinguished by ecotypes, including red rice, white rice, weedy rice, wild rice and unknown type. (C) Individual ancestry coefficients of $K = 12$ of 154 wild rice, 294 weedy rice, 361 red rice, and 1,238 white rice.



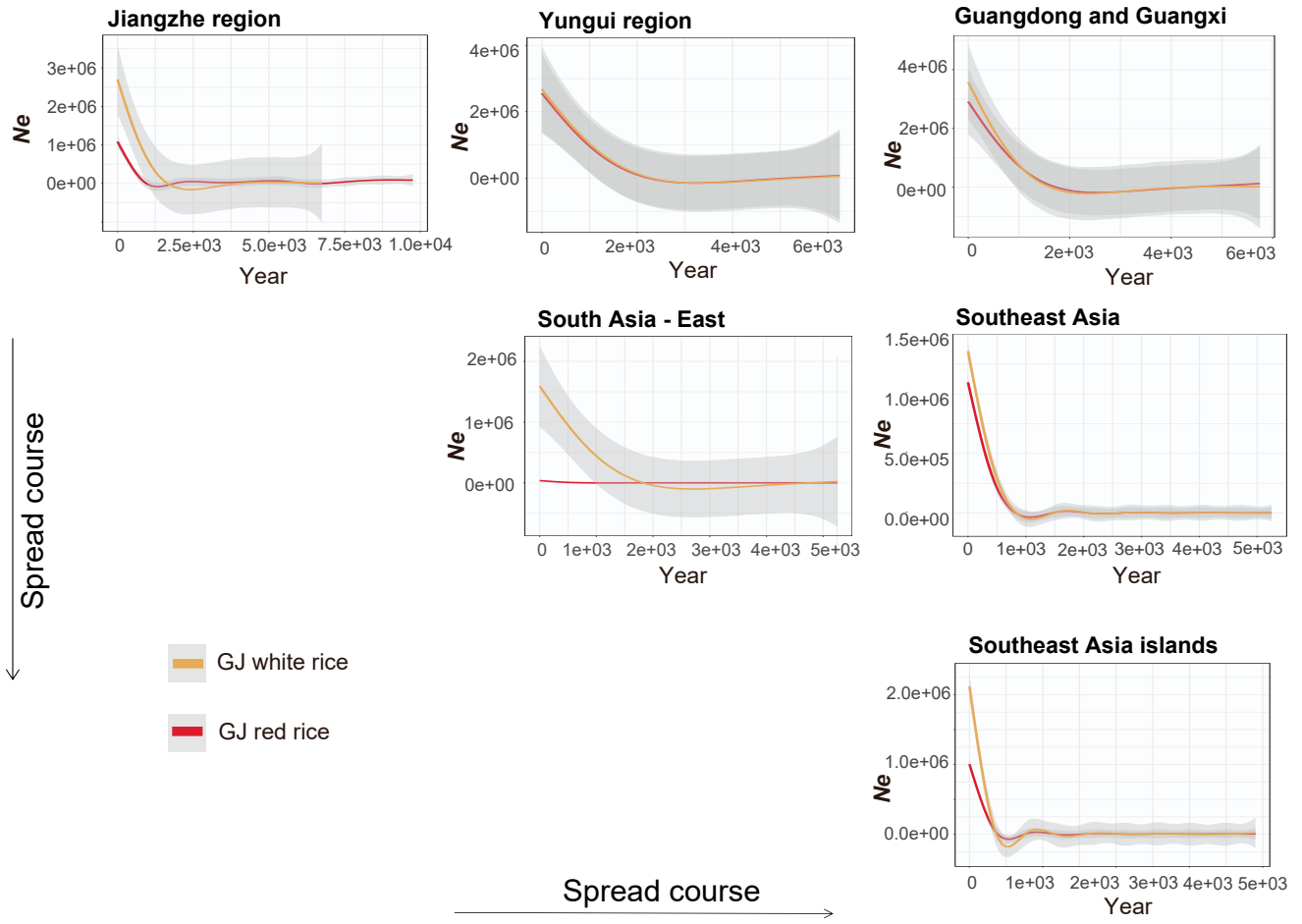
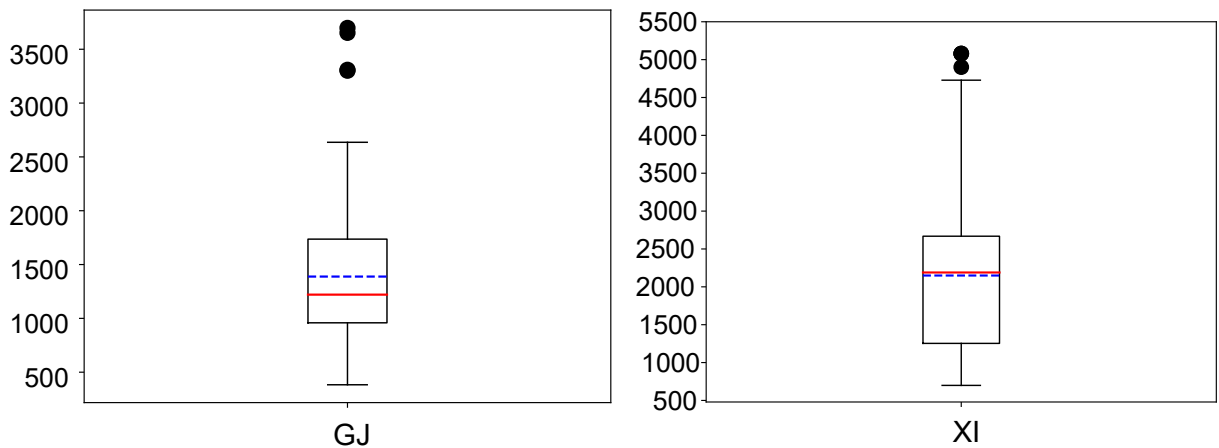
Supplementary Fig.3 Detailed phylogenetic tree of the rice accessions which were marked with red triangle in Figure 1B to show the relationship between a cultivated red rice (represented by the red line) with weedy red rice. The cultivated red rice is surrounded by weedy red rice (blue line). Other cultivated rice was represented by a gray cluster.



Supplementary Fig.4 Allele frequency in cultivated red rice and weedy rice, with the frequency range in wild rice shown on the x axis. "C" and "W" refer to cultivated red rice and weedy rice, respectively.



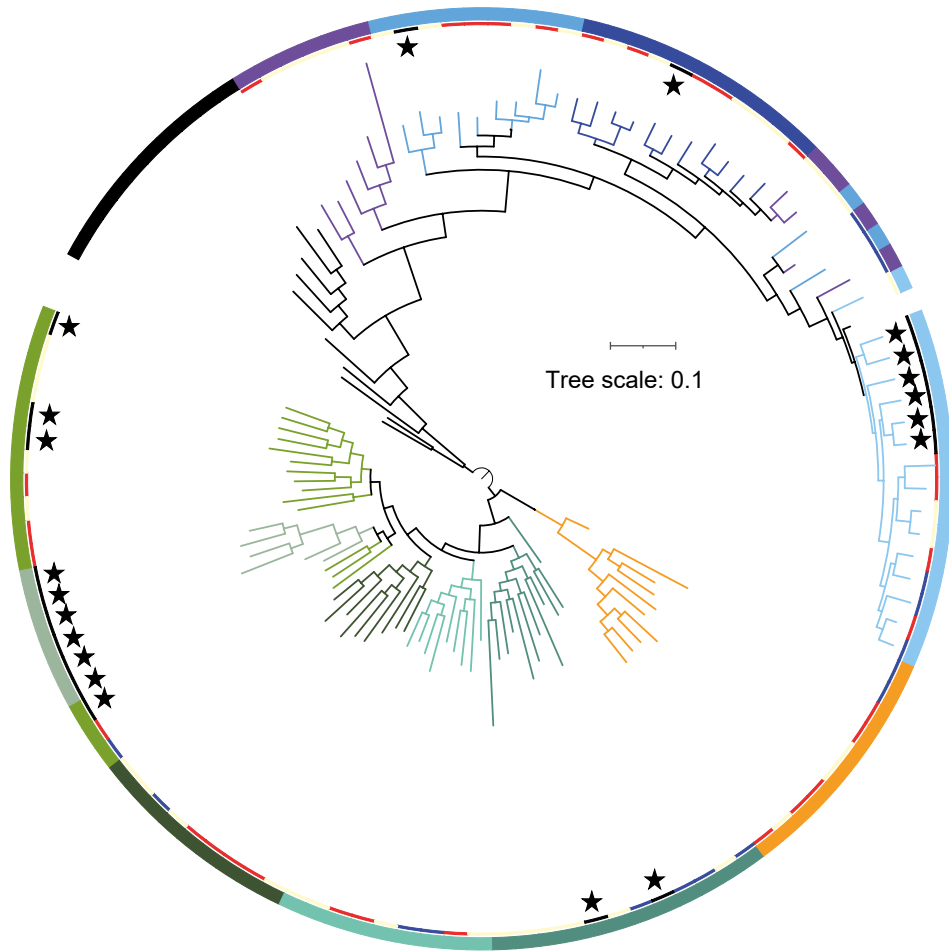
Supplementary Fig.5 (A) Identity-by-state (IBS) distance between wild rice and XI red rice or XI white rice in different regions. (B) Identity-by-state (IBS) distance between wild rice and GJ red rice or GJ white rice in different regions. * and ** indicate significant difference at $P < 0.05$ and $P < 0.01$, respectively, determined by the Wilcoxon test.

A**B**

Supplementary Fig.6 (A) Effective population size history of red rice and white rice subgroups assessed by the coalescent approach SMC++. The thick line represents the mean and the grey-shaded area represents the quartiles of random individuals ($n = 15$ random pairs). (B) Boxplots of estimated split times of red rice and white rice in GJ and XI population. Center black line: median; Center blue line: mean; box limits: upper and lower quartiles; dots: outliers.

 M16_R G G A A G G C C G G - - G G G G T T C C C C T T T T G G G G T T A G T T
 Q100_R G G A A G G C C G G T T G G G G T T C C C C T T T T G G G G T T A G T C
 Q76_W -
 297_W - - - - - C C - - - - G G G G T T C C - - - - - - - - - - - - - - - A A T T

Supplementary Fig.7 The transcripts of *Os01g0152951* in the two red rice (M16_R, Q100_R) and two white rice (Q76_W, 297_W).



Supplementary Fig.8 Phylogenetic tree of the 120 genomes inferred from whole-genome SNPs, including 10 wild rice, 32 red rice, 20 black rice, 16 weedy rice, and 42 white rice. Different subspecies are indicated by color lines and the innermost circle represents ecotypes of accessions. Red, blue, black, and light yellow represent red rice, weedy rice, black rice, and white rice, respectively.