README

The file “GBS\_6K\_CUR\_MER\_32\_566\_dbSNPs\_MSUv7\_20160212.txt” contains genotypic data from the Introgression Line (IL) population derived from the cross between Curinga (*O. sativa tropical japonica*) and OR44 (or W2112, *O. meridionalis*). This population consists of 32 ILs, MER1 to MER32, the recurrent parent CURINGA and the donor wild parent OR44. The data set contains 566 informative SNPs. Each SNP ID is listed in column 1 (***snp\_id***). Each SNP allele is listed in column 2 (***alleles***), the chromosome (***chr***) and MSU.v7 physical position (***pos***) are listed in columns 3 and 4. The SNP genotypes for the Curinga, OR44, and each IL (MER1-MER32) are listed in the columns named after each genotype.

The file “GBS\_6K\_CUR\_RUF\_48\_2272\_dbSNPs\_MSUv7\_20160212.txt” contains genotypic data from the Introgression Line (IL) population from the cross between Curinga (*O. sativa tropical japonica*) and IRGC105491 (*O. rufipogon*). This population consists of 48 ILs, RUF1 to RUF48, the recurrent parent CURINGA and the donor wild parent IRGC105491. The data set contains 2,272 informative SNPs. Each SNP ID is listed in column 1 (***snp\_id***). Each SNP allele is listed in column 2 (***alleles***), the chromosome (***chr***) and MSU.v7 physical position (***pos***) are listed in columns 3 and 4. The SNP genotypes for the Curinga, IRGC105491, and each IL (RUF1-RUF48) are listed in the columns named after each genotype.

**6K\_GBS final combined SNP data sets**

The final GBS.6K dataset consists of 566 SNPs defining introgressions in the CUR/MER population and 2,272 SNPs defining introgressions in the CUR/RUF population. These SNPs were selected from the original 6K SNP and GBS datasets as follows:

* Polymorphic SNPs were identified independently in the CUR/MER and CUR/RUF populations using the 6K chip and GBS. There were 1092 SNPs from the 6K chip and 107,445 GBS-SNPs polymorphic between Curinga and *O. meridionalis (*OR44), and 1,769 from the 6K chip and 89,577 GBS-SNPs polymorphic between Curinga and *O. rufipogon* (IRGC 105491).
* SNPs detected using both platforms were combined into a single dataset for each population and sorted by chromosome and physical position in the genome (MSUv7).
* An R scrip was used to identify informative SNPs for each population; this set was thinned to identify a minimum number of informative SNPs that could detect and define the regions of introgression in each population.
* Because there were many more, smaller introgressions in the CURUF population, the minimum number of informative SNPs was 2,272 for CUR/RUF population and 566 for the CUR/MER population.