DNA Gynotyping for Assessing Variety Area Estimates based on Farmer Identification: Case of Rice in Bangladesh

T. Yamano, M.L. Malabayabas, M.A. Habib, S.K. Das, Z. Huelgas, G. Carino, T. Kretzschmar, and others

Introduction

Based on farmers' identification of variety names, areas under different rice varieties have been estimated by previous studies. Farmers' knowledge on rice variety names, however, is considered unreliable in developing countries where farmers obtain seeds from various sources. Only a small portion of farmers buy seeds in certified packages. Even certified seeds could be incorrectly labeled or adulterated. In Bangladesh, rice variety names are more confused than in other countries, because many Indian rice varieties have been brought to the country unofficially. This situation makes it difficult to track diffusion of new rice varieties, such as submergence-tolerant rice varieties. To mitigate flood damages on rice production, two submergence-tolerant rice varieties, called BR11-Sub1 and Swarna-Sub1, have been developed and distributed in Bangladesh since 2010.[1] The submergence-tolerant rice varieties have a single major quantitative trait locus (QTL) responsible for submergence tolerance, named Sub1 QTL, allowing rice varieties to withstand up to 14 days of complete submergence. To assess the area estimates based on farmers' identification of variety names, we collected 1,289 rice seed samples from 554 farmers in 2014 and 2015. The seed samples were collected from a pooled samples of 3,000 farmers who were interviewed either in 2014 or 2015. The gynotyping of the farmer and breeder seed samples was conducted by using Illumina Infinium 6K SNP chips (Illumina Infinium 6K SNP - http://gsl.irri.org/services/infinium-6k).

Data

For the surveys, we randomly selected 16 districts out of 57 districts in Bangladesh, after excluding remote districts where rice is not grown. In the 16 districts, we randomly selected 75 thanas out of 117 in the districts. Two villages were randomly selected from each sample thana, and ten farmers were randomly selected in each village (see Appendix Figure 1). One of the two villages in each thana was selected for seed selection, and randomly selected four out of 10 sample households were selected for seed collection. Finally, from each of the selected households, we listed all rice varieties that they produced in the last aman season, and collected aman rice seeds/grains up to four varieties.

DNA Fingerprinting - Method

The collected seeds were kept at the IRRI office in Dhaka and germinated at once in early 2016. When rice plants were grown at knee high, leaf samples were taken and small

[1] BR11-Sub1 is called BRRI dhan 52, and Swarna-Sub1 is called BRRI dhan51 in Bangladesh

pieces were packed in plastic tubes with Silica gel to keep moisture low. Boxes of plastic tubes with leaf samples were sent to the Genotyping Services Laboratory (GSL) located at the International Rice Research Institute (IRRI) Head Quarters in the Philippines. Gynotyping was conducted by using Illumina Infinium 6K SNP chips. From 6K data pointes, about 4K data points were selected for identifications. This suggests that 100% match indicates that only less than 20 SNP points are difference between two samples. It is rare but still possible for two different samples to share more than 3980 data points, especially for closely related varieties, such as Swarna and Swarna-Sub1. Therefore, we also check for availability of SUB1 QTL markers to identify submergence-tolerant rice varieties with the Sub1 QTL.

Results

Area estimated based on a pooled farmer surveys

Based on the pooled farmer surveys of 3,000 farmers, we estimated areas under different rice varieties using the farmer variety identification. The results indicate that the most popular variety in Bangladesh is an Indian variety called Swarna (1 million ha – 21%), followed by BR11 (276'000ha, 5.9%), Binadhan 7 (240,400ha, 5.1%), and Sadamota 206,450ha, 4.4%). The estimated area under BR11-Sub1 (BRRI dhan52) was 51,750ha (1.1%), and that of Swarna-Sub1 (BRRI dhan51) was 37,750ha (0.8%). Thus, the submergence-tolerant rice varieties combined covered 89,500 ha (1.9%). Note that, through the DNA gynotyping, we can only identify modern varieties that we have breeder seeds (we do not have reference data for traditional and hybrid rice varieties). According to the farmer identification, less than 50% of the total areas is under modern varieties.

DNA Fingerprinting Results

Out of the 1,289 seed samples, we identified 186 samples (14.4%) with breeder seeds. We used 100% match as a cut-off point for all varieties, except for Swarna. For Swarna, we used 96% as a cut-off point because we used 3K SNP data as reference for Swarna, and none of the seeds samples is expected to have 100% match with the 3K SNP data of Swarna. On other varieties, we found: BR11 (38 samples), BR11-Sub1 (23), BR23 (17), BR10 (13), and Swarna-Sub1 (5). On the submergence-tolerant rice varieties, we found that the share of the submergence-tolerant rice varieties is 2.2% of the total sample size, and the share is close to the area share estimate of 1.9%. Regarding the Sub1-QTL, it was found on all of the Swarna-Sub1 samples and 78.3% of the BR11-Sub1 samples. It is also found a few samples of BR11, BR10, and BR23. Notably, it is found on about 8% of un-identified samples. Farmers' variety identification is found poor. Only 31% of seed samples named Swarna (with some variations in their names) were named correctly. The rest, 69%, were wrongly identified as Swarna (False-positive, Type I error). At the same time, 30 seed samples were named otherwise but identified as Swarna (False negative – Type II error). Regarding other varieties, 30% of BR11 seed samples were correctly identified, and so were 13.8% of BR10. None of the new varieties, such

as BR11-Sub1, Swarna-Sub1, and Binadhan 7 were correctly identified. Further analysis is required, however, to confirm the preliminary findings provide in this study.

Caveats and Plan for Additional Analyses

Breeder seeds used for reference need to be examined. Which breeder seed should be treated as “true”?

Rice seeds used by farmers might have gone through mutations or cross pollinations. How should we consider this?

We merge the seed data with household data and find determinants of correct identification against the main seed source, household characteristics, locations, etc.