



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

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

P.C. Veettil, I. Gupta, T. Yamano, Z. Huelgas, G. Carino, T. Kretzschmar, and others

Introduction

To assess the area estimates based on farmers' identification of variety names in eastern India, we have collected 2,797 rice seed samples from 1,380 farmers in 2015. To verify the identity of the seed samples, we have also collected breeder seeds from seed companies and public institutes. 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>). By using this approach, we effectively compare more than 4,000 SNP points across seed samples.

Data

The sampling of the 2015 survey used a simple self-weighting design across states. The total number of villages in each state was determined based on the total rice area in each state. A simple random sampling was used to select villages within each state by using the 2001 Census. In each of the selected villages, 10 households were randomly selected after listing households in the village. The total number of households interviewed was 6,740. From the sample households, 20 percent of them were selected for rice seed collection. In this paper, we use data from 2,797 seed samples collected from 1,380 households. The average number of samples collected per household was 1.8. More samples were collected from rice farmers in Odisha because they produce more varieties in kharif 2015.

DNA Fingerprinting - Method

The collected seeds were sent to a private company called SciGenom (<http://www.scigenom.com/>) located in Chennai, India, and DNA gynotyping was conducted by using Illumina Infinium 6K SNP chips (Illumina Infinium 6K SNP - <http://gsl.irri.org/services/infinium-6k>). From 6K data points, 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 is still possible for two different samples to share more than 3,980 data points. It is possible for two closely related varieties, such as Swarna and Swarna-Sub1. Therefore, we also check for availability of SUB1 QTL markers to identify Sub1 varieties.

Results

Area Estimated Based on a Pooled Farmer Surveys

Based on the farmer survey, we estimated areas under different rice varieties based on the farmer variety identification. The results indicate that the most popular variety in eastern India was Swarna (4.2 million ha – 29%), followed by Mahsuri (1.3 million ha, 8.9%), Pooja (1.0 million ha, 6.6%), and Lalat (0.9 million ha, 6.0%). The estimated area under Swarna-Sub1 was 0.4 million ha (2.7%). Note that, through the DNA gnotyping, 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 73% of the total areas is under modern varieties.

DNA Fingerprinting Results

Out of the 2,797 seed samples, we identified 650 samples (23.2%) with breeder seeds. We used 96% match as a cut-off point for all varieties. Because the results are still preliminary and can have commercial implications, we use pseudo names for variety names of breeder seeds, except for a few varieties such as Swarna and Sabhagidhan. In Table 1, we identified 241 samples (8.6%) as Swarna, 92 samples (3.3%) as Variety 1, and 74 samples (2.6%) as Variety 2. These varieties are called by different names. Swarna, for example, called as Swarna (122 samples), Mahsuri (13), and others. Out of 92 samples of Variety 1, 41 samples were mistaken as MTU1010, but 10 samples were called Lalat, although Lalat is a common name used for Variety 2: 39 sample out of 74 sampled identified as Variety 2 was called Lalat. Well known variety names, such as Swarna, Lalat, and Puja, are used for different rice varieties.

Table 1 Major Varieties Identified by the 6K SNP analysis

Variety	Number of seed samples identified ^A	Variety names commonly used by farmers
	(A)	(C)
	Number (%)	
Swarna	241 (8.6)	Swarna (122), Mahsuri (13), Niranjan (4), Moti Gold (2), others
Variety 1	92 (3.3)	MTU1010 (41), Lalat (10), Swarna (4), Moti (1), others
Variety 2	74 (2.6)	Lalat (39), MTU1001 (4), Swarna (2), Puja (2), others
Variety 3	74 (2.6)	Lalat (48), Puja (2), MTU1001 (2), others
Variety 4	38 (1.4)	Moti (8), others
Variety 5	28 (1.0)	Swarna (16), Ranjit (2), Puja (1), others
Variety 6	22 (0.8)	MTU1010 (16), Lalat (2), others
Variety 7	16 (0.6)	Sarju 52 (14), others
Sabhagi Dhan	15 (0.5)	Sahbhagi Dhan (4), Swarna (2), others
...		
Not identified	2,147 (76.8)	
Total	2,797 (100)	

Note: ^A A cut-off point of 96% is used (accepted if 96% or higher).

By farmers, 382 seed samples were called Swarna. Among them, only 32% were identified correctly named as Swarna, and the rest were identified wrongly as Swarna (False-positive, Type I error). In addition, 119 samples were called otherwise and were identified as Swarna (False negative – Type II error). Other samples have been matched with breeder seeds, but we need to conduct additional analysis to cross-check if the names given by farmers were correct.

Caveats and Plan for Additional Analyses

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

We are in the process of identifying factors associated with the correct identification of varieties. The factors examined include main seed source, household characteristics, and locations.