國立臺灣大學生物資源暨農學院生物產業機電工程學系

碩士論文

Department of Bio-Industrial Mechatronics Engineering
College of Bioresources and Agriculture
National Taiwan University
Master Thesis

# 辨識不同種類之稻米以及研究其基因型與表現型之關聯

# Identifying Rice Grains of Various Varieties and Studying the Genotype-Phenotype Association of Rice Grains

郭子毅

Tzu-Yi Kuo

指導教授：郭彥甫 博士

Advisor: Yan-Fu Kuo, Ph.D.

中華民國 105 年 1 月

January, 2016

# ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor Yan-Fu Kuo, my research advisor, for his patient guidance and encouragement during my master program. I would also like to thank Szu-Yu Chen, Heng-An Lin and Dr. Chia-Lin Chung for their support in providing rice samples and the expertise in biological field. I would like to acknowledge the member in Lab 304, Cheng-Liang, Tzu-Kuei, Walter, Han, Cheng-Chun, Jerry and David with their friendship, who had accompanied me during my master program. Last but not least, I want to express my deeply appreciation to my family and Pennie Chen, especially my parents, for their support and encouragement throughout my study.

# 摘要

水稻是全世界許多人的主食，每年在國際市場上交易的數量十分龐大。不同的品種的水稻在外觀上存在著差異，這些外觀差異可以藉由分析其與水稻基因型的關聯來了解造成外觀差異的原因。本研究利用影像處理以及稀疏表達分類器等非破壞性檢測分辨 30 種不同的水稻，同時也對於 255 種水稻的外觀以及基因型做出其關聯性的分析。稀疏表達分類器則可以利用過度充分基底來捕捉具有代表性的外觀特徵。在實驗中，種子取自於 Genetic Stocks Oryza，基因資訊取自於公開資料庫，利用顯微鏡與高解析度數位相機提高影像畫質。量化的外觀特徵大致上被分為水稻種子以及護穎的形態、顏色以及紋理等特徵。接下來利用線性模型對上述特徵分析其與基因型的關係。稀疏表達分類器藉著輸入量化的外觀特徵來辨識其中 30 種水稻品種，稀疏表達分類器對於 30 種品種的辨識準確率可達到 89.1%。

**關鍵詞**：水稻辨識、稀疏編碼、影像處理、機器視覺、數量性狀基因座

# ABSTRACT

Rice (*Oryza sativa L.*) is a major staple food and is traded globally in considerable amount. Rice shows remarkable variation in grains. The phenotypic information of the rice grains need to be quantified as the first step to investigate the association between the phenotypes and genotypes. This study proposed to distinguish the rice grains of 30 varieties nondestructively using image processing, sparse representation based classification (SRC) and a procedure to phenotype rice grains of 255 varieties in high precision. SRC is a method that uses over-complete bases to capture the representative traits of rice grains. In the experiments, rice seeds were acquired from Genetic Stocks Oryza germplasm collection. The genotypic information (i.e., SNPs) of these seeds are publicly available. The images of the grains were acquired in high resolution using microscopy (approximately 2413 dots per inch). Morphological, color, and textural traits of the grain body, sterile lemmas, and brush were quantified. The traits were subsequently fit into a unified mixed linear model for investigating the association between the phenotypic and genotypic variations of the varieties. An SRC classifier was developed to identify the varieties of the grains using the traits as the inputs. The proposed approach could discriminate the varieties of the rice grains with an accuracy of 89.1%.

**Keywords.** Variety identification, sparse coding, locality constraint, machine vision, machine learning, image processing, phenotyping

# TABLE OF CONTENTS

# LIST OF FIGURES

# CHAPTER 1. INTRODUCTION

## 1.1 Rice Phylogeny

*Oryza sativa* has hundreds of varieties with different grain color, size, and shape. The Food and Agriculture Organization of the United Nations estimates that the global rice production in 2015 was 491.4 million tons. China and India were the leading producers, followed by Indonesia, Bangladesh, and Vietnam. In 1992, Zhang et al. applied restriction fragment length polymorphism to analysis separate *Oryza sativa* into two subspecies, *indica* and *japonica*. In 2007, the improved differentiate methods for the two groups of *Oryza sativa* were proposed by Kovach et al.. *Oryza sativa* was further sorted into five groups, *indica, aus, temperate japonica*, *tropical japonica* and *aromatic*, by using simple sequence (SSR) repeats markers (Garris et al., 2005; Huang et al., 2010), while *indica* and *aus* were categorized from the *indica* varieties and *temperate japonica*, *tropical japonica* and *aromatic* classified from the *japonica* varieties.

**Fig. 1.** Rice grains of various varieties.

## 1.2 Genome Architecture

The genome of *Oryza sativa* consists of 430 Mb across 12 chromosomes. Variations on DNA sequences, such as insertion, deletion, short tandem repeats, and single nucleotide (SNPs), can be found between populations, subpopulations, varieties, and individuals. The most common type of variation is the SNP. SNPs have been widely utilized to

characterize genetic diversity at the sequence level and look for genetic variation associated with phenotypic traits.

## 1.3 Objectives

This study aimed to differentiate rice grains of 30 varieties using locality constrained SRC and study the genotype-phenotype association of rice grains. The specific objectives of the research were to (1) establish a microscopic imaging system for acquiring images of the grains, (2) quantify the morphological, color, and textural traits of the grain body and parts, (3) develop a locality constrained SRC classifier for identifying varieties of the rice grains, and (4) fit the traits into a unified mixed linear model for investigating the association between the phenotypic and genotypic variations of the varieties.

## 1.4 Organization

The remaining of this document is organized as follows. In Chapter 2, some methods for identifying rice grain and some studies which analyzed genome-wide association were reviewed. In Chapter 3, the process and results of distinguishing the rice grains of 30 varieties nondestructively using image processing and SRC were presented. In Chapter 4, the association between the phenotypic and genotypic variations of the varieties was revealed. Conclusions of this research are given in Chapter 5.

# CHAPTER 2. LITERATURE REVIEW

## 2.1 Genetic Marker-based Methods to Identify Varieties of Rice Grains

Genetic marker-based methods have been applied for identifying varieties of rice grains. Steele et al. (2008) selected insertion and deletion markers to distinguish Basmatis rice grains from some other fragrant rice varieties. Cirillo et al. (2009) applied random amplified polymorphic DNA approach to fingerprint rice grains of 13 Italian varieties. Becerra et al. (2015) determined the genetic variability of some Chilean and foreign commercial rice varieties using SSR markers. SSR markers were also used in another work to distinguish rice grains of 36 varieties from different countries (Chuang et al., 2011). Although these chemical methods are accurate, they can be time-consuming, destructive, labor-intensive, and costly for field application.

## 2.2 Nondestructive Method to Identify Varieties of Rice Grains

Image-based approaches, in contrast, are nondestructive and rapid. The approaches combine image analysis and machine learning techniques to achieve automatic inspection and evaluation. Image-based approaches have been applied for discriminating varieties of cereal grains using one or a combination of morphological, color, textural traits. Camelo-Méndez et al. (2012) characterized rice grains of 9 Mexican varieties using principle component analysis (PCA) and hierarchical analysis. Kong et al. (2013) classified rice seeds of 4 varieties using a near-infrared hyperspectral imaging system and various machine learning algorithms. Mebatsion et al. (2013) distinguished barley, oat, and, rye using a least squares classification approach. Another work applied multi-layer perceptron and neuro-fuzzy classification networks for identifying 5 Iranian rice varieties (Pazoki et al., 2014). Although the results of the studies were promising, they only included a relatively limited numbers of varieties for discrimination.

## 2.3 Sparse Representation Based Classifier to Identify Divergent Objects

SRC is a machine learning algorithm suitable for solving high-dimensional problems. SRC encodes representative characteristics (i.e., atoms) of training samples into a dictionary. The dimension of the traits space increases during the encoding process. When a query sample is provided, the sample is coded as a sparse combination of the atoms. The query sample is then assigned to the class that yields the least coding error. SRC is robust to noises and is computationally inexpensive. The method has been used for detecting flowers of various species (Yuan et al., 2012), recognizing field crop insects (Xie et al., 2015), and assisting the diagnosis of Alzheimer's disease (Liu et al., 2012).

## 2.4 Genome-wide Association Studies on Rice Grains

Genome-wide association study is an analysis of SNPs in different individuals to see the relationship between SNPs and phenotypic traits. Rice SNP arrays were successfully used for variety verification and trait introgression. The accurate high-throughput genotyping tool to enhance density and quality of rice SNP arrays were proposed by Chen et al. (2014). In GWAS, a large number of accessions for genetic variation can be screened underlying diverse complex traits. In the population of *indica*, 14 agronomic traits including heading data, grain size and starch content have been identified by sequencing 517 rice landraces (Huang et al., 2010). Huang, Zhao, et al. (2012) identified 32 new loci associated with flowering time and with ten grain-related traits by using GWAS. The domestication-associated traits in cultivated rice were analysed through high-resolution genetic mapping (Huang et al., 2012).

# CHAPTER 3. IDENTIFYING RICE GRAINS OF VARIOUS VARIETIES

## 3.1 Material and Methods

### 3.1.1 Grain sample preparation

Rice grains of 30 varieties were used in this study. The varieties used in this study were selected from these subpopulations, including 7 from *indica*, 8 from *aus*, 3 from *aromatic*, 7 from *template japonica*, and 5 from *tropical japonica*. The grain samples were acquired from Genetic Stocks Oryza germplasm collection (Agricultural Research Service, United States Department of Agriculture) and reproduced in a local greenhouse (Kaohsiung District Agricultural Research and Extension Station, Taiwan) in 2013. After harvest, the grains were dried to the moisture content of approximately 13% and were stored in refrigerators at 4°C. Fifty grains of each variety were prepared.

### 3.1.2 Rice grain exterior

A grain body is typically composed of husks and sterile lemmas (Fig. 2). In some varieties, brush covers the surface of the husks. These organs (e.g., husk, sterile lemma, and brush) possess morphological, textural, and color traits that can be used for identifying the varieties of the grains. The awns of the grains typically fall off during the drying. Hence, the traits of awn were not considered in this study.

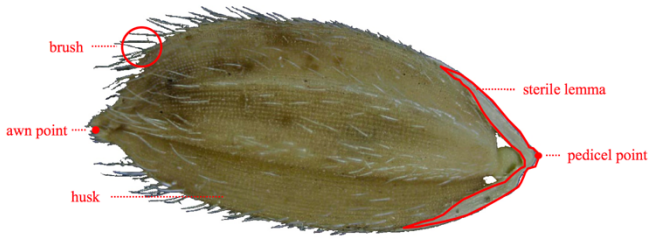



**Fig. 2.** Husks, sterile lemmas, and brush of a rice grain. Husks are the outermost layer. Sterile lemmas are the 2 flowerless bracts connected to the pedicel. Brush is the hair on husk and is readily observed in some varieties.

3.1.3 Imaging system and image acquisition

A system was developed to acquire images of the grains (Fig. 3). The system comprised a digital camera (EOS 450D, Canon; Tokyo, Japan), a microscope (BXFM, Olympus; Tokyo, Japan), a 2X objective lens (PLN UIS2, Olympus; Tokyo, Japan), and a ring-shaped LED illuminator. The LED illuminator was placed 15 mm above the surface of the sample placement platform of the microscope. The system was enclosed in a dark chamber to prevent stray light. Before the image acquisition, the system was calibrated using a standard color reference board (Color Checker Passport, X-rite; Grand Rapids, USA) to estimate device-independent color parameters of the grains. The camera was set in manual mode for image acquisition, using an ISO of 400 and a shutter of 1/30 s.

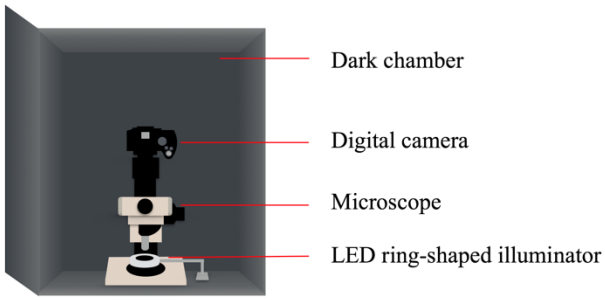



**Fig. 3.** Image acquisition system.

3.1.4 Multi-focus image fusion and background removal

Multi-focus image fusion (Wang and Chang, 2011) was applied to improve the quality of the grain images. Lenses of optical microscope typically have limited depth of field. Two micrographic images of the same rice grain, one focusing at the center of the grain and the other focusing at the edge of the grain, were taken. The two photos were merged to obtain an image with all pixels in focus. The fusion involved matching, registration, and consolidation. In the matching process, the same characteristic points of the rice grain in the two photos were identified by using speeded up robust features (Bay et al., 2006).

Next, the two images were registered using the identified characteristic points and fast approximate nearest-neighbor search algorithm (Muja and Lowe, 2009). Subsequently, the two images were consolidated into one using Laplacian pyramid transformation (Jin et al., 2005). The fused images were 1068×712 pixels and were at a resolution of approximately 95 dots per millimeter (2413 dots per inch). The image fusion algorithm was implemented using programs written in C language and was developed in Xcode (Apple Inc.; Cupertino, CA, USA).

The rice grain in each image was segmented from the background. First, k-means operation (Hartigan and Wong, 1979) was applied to the hue channel in the hue-saturation-value color space. The number of clusters in the k-means operation was set to two. The algorithm labeled each pixel in the image as foreground or background. Next, connected-component labeling (Dillencourt et al., 1992) was applied to identify the largest foreground object as the grain body. For some varieties, the brush could be recognized as a part of the grain body. This false recognition reduced the accuracy in estimating the grain perimeter. Therefore, morphological closing (Gonzalez and Woods, 2007) was performed to remove the brush outside the grain contour from the grain body. After the grain body segmentation was completed, the awn and pedicel points were determined as the two points on the grain contour with the largest distance apart. The area of the sterile lemmas of the grains were also identified using color thresholding.

### 3.1.5 Trait quantification

Traits were quantified to describe the characteristics of the rice grains. The traits were categorized into 4 groups: morphological traits, color traits, textural traits, and Fourier descriptors. Twelve morphological traits were calculated for the grain body. The traits were perimeter, surface area, length of major axis, length of minor axis, axis aspect ratio,

arc ratio, standard deviation (SD) of radii, maximum radius, minimum radius, radius ratio, Haralick ratio, and thinness ratio (Fig. 4) (Camelo-Méndez et al., 2012; Majumdar and Jayas, 2000). The major axis was defined as the line connecting the awn and pedicel points. The minor axis was defined as the line perpendicular to the major axis with the longest segment intersecting the grain. The axis aspect ratio was the ratio of the major axis length to the minor axis length. The arc ratio was the ratio of major arc length to minor arc length. The major and minor arcs were the long and short contour segments, respectively, along the grain contour between the awn and pedicel points. The radii were the distances between contour points and geometric centroid of the grain. The radius ratio was the ratio of the maximum radius to the minimum radius. The Haralick ratio was the ratio of the mean radius to the SD of the radii. The thinness ratio was defined as the ratio of the area to the perimeter. A grain of a larger thinness ratio is more circular, whereas a grain of a smaller thinness ratio is more pointy. Morphological traits were also quantified for the sterile lemmas. The traits were the lengths along the major arc (L1 in Fig. 4) and the minor arc (L2 in Fig. 4), ratio of L1 to the major arc, ratio of L2 to the minor arc, area of the sterile lemmas, and ratio of the sterile lemmas area to the grain body area.

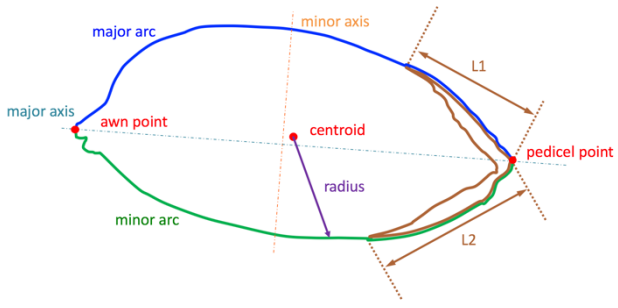



**Fig. 4.** Morphological traits of grain body. L1 and L2 were lengths of the sterile lemmas along the major and minor arcs, respectively.

Nine color traits were acquired for the husk: red, green, and blue (RGB) parameters; hue, saturation, and value parameters; L*, a*, and b* in the CIE color space (Hunter, 1975). The color traits were mean color parameters of the pixels in a region of interest (ROI). The ROI was a rectangle centered at the grain centroid. The edges of the rectangle was parallel to the major and minor axes of the grain. The length and width of the ROI were 50% of the lengths of the major and minor axes, respectively. The L*, a*, and b* parameters were converted from the ROI RGB values using the transform functions obtained during the calibration. These 9 color parameters were also quantified for the sterile lemmas using the entire area of the sterile lemmas as the ROI.

Seven textural traits were assessed for the husk: brush ratio and 6 gray level co-occurrence matrix (GLCM; (Haralick et al., 1973) traits. The brush ratio was defined as the percentage of the ROI covered with brush. The 6 GLCM traits were: mean, variance, uniformity, entropy, contrast, and correlation (Galloway, 1975). To quantify the GLCM traits, the ROI was first converted to a grayscale image of 3-bit word length. The GLCM of this image was then calculated using a displacement vector with a direction parallel to the major axis. The 6 traits were subsequently calculated from the GLCM following their definitions (Haralick et al., 1973).

Fourier descriptors were quantified for the grain body. Fourier descriptors are a set of sine and cosine harmonics at various frequencies for encoding the outline of an object (Rohlf and Archie, 1984). The low-frequency harmonics characterize coarse outlines of the grains, whereas the high-frequency harmonics depict fine details of the grain contours. To calculate the descriptors, grain contours were first represented as a sequential connected points in a Cartesian coordinate system. The coordinates of the connected points in each dimension were then converted into Fourier descriptors by using discrete

Fourier transform (Harris, 1978). The first 20 low-frequency descriptors were retained to characterize the grain shapes. The harmonics at high frequencies were disregarded because that the grain images were of limited resolution. The high-frequency harmonics may depict noise rather than the native structures of the grain contours. The Fourier power of the first 20 harmonics was 99.9% (Costa et al., 2009). The grain images were two-dimensional, and there was a sine and a cosine harmonic in each dimension. A total of 80 coefficients of the descriptors were collected as the traits.

### 3.1.6 Variations in grain shape

The shape variations between the grains of the 30 varieties were examined. PCA was first applied to the Fourier descriptors for obtaining principal components (PCs). Each PC was associated with a particular grain shape variation. The PCs were arranged in descending order by percentage variance. The first few PCs accounted for a large proportion of the variance and could represent the major shape variations. Next, the shape variation associated with a PC was visualized by reconstructing grain contours. In the process, the mean and SD of the PCs were calculated. Fourier descriptors were next derived using inverse PCA with a specific PC value being manipulated, whereas other PC values were maintained at the mean values. The manipulated PC values were set at the mean or mean ± 2 SD. Grain contours were then reconstructed using the resulting Fourier descriptors and inverse Fourier transform.

### 3.1.7 Variety identification

Locality constrained SRC was applied for identifying the varieties of the rice grains. SRC aimed to encode the grain traits as a sparse linear combination of a dictionary. The dictionary represented the essence of the grain traits and was developed using the collected traits. The dictionary learning process is explained as the follows. Considered

$N \in$ training samples (i.e., number of grains) of the $j^{th}$ cultivar, where $j = 1 \quad J \in$ , were gathered. Each training sample was associated with $M \in$ traits. Assumed that the dictionary has $K \in$ atoms. Let $\boldsymbol{X}_j \in {}^{M \times N}$, $\boldsymbol{D}_j \in {}^{M \times K}$, and $\boldsymbol{A}_j \in {}^{K \times N}$ denote the matrix of the training samples, dictionary to be trained, and sparse coefficient matrix, respectively, of the $j^{th}$ cultivar. The dictionary was obtained by solving the equation:

$$\boldsymbol{D}_j = \underset{\boldsymbol{D}_j, \boldsymbol{A}_j}{\arg\min} \left\| \boldsymbol{X}_j - \boldsymbol{D}_j \boldsymbol{A}_j \right\|_2^2 + \lambda \sum_{i=1}^{N} \left\| \boldsymbol{p}_{ji} \quad \boldsymbol{\alpha}_{ji} \right\|_2^2 \quad , \qquad (1)$$
$$s.\mathrm{t.}\ \ \mathbf{1}^{\mathrm{T}} \boldsymbol{\alpha}_{ji} = 1,\ i = 1,...,N$$

where $\boldsymbol{\alpha}_{ji} \in {}^{K \times 1}$ is the $i^{th}$ column of $\boldsymbol{A}_j$, $\lambda \in$ is regularization parameter that controls the sparsity of $\boldsymbol{\alpha}_{ji}$, $\boldsymbol{p}_{ji} \in {}^{K \times 1}$ is locality adaptor, and the symbol denotes element-wise multiplication. The locality adaptor $\boldsymbol{p}_{ji}$ was defined as a vector consisting of the Euclidean distances between a training sample (i.e., a column of $\boldsymbol{X}_j$) and the columns of $\boldsymbol{D}_j$. This regularization formulation using the locality adaptor considered the underlying manifold structures of the grain traits (Wang et al., 2010). Hence, SRC could produce dictionary that encodes the essential patterns of the grains. The dictionary development was performed using the locality-sensitive dictionary learning algorithm proposed by Wei et al. (2013). The optimal regularization parameter $\lambda$ and dictionary size $d$ were determined using grid search and ten-fold cross-validation (Arlot and Celisse, 2010).

Once developed, the dictionary $\boldsymbol{D}_j$ was employed to classify the grains using the traits as the inputs. Let $\boldsymbol{y} \in {}^{r \times 1}$ be a query sample. The sparse coefficients $\boldsymbol{\alpha}^j \in {}^{d \times 1}$ were calculated for all the varieties ( $j = 1 \quad J$ ) to assemble the sparse reconstructions

$\hat{\boldsymbol{y}} = \boldsymbol{D}_j \boldsymbol{\alpha}^j$ of the query sample. The variety of the query sample was then determined as the variety associated with the dictionary that gives the minimum reconstruction error:

$$\begin{aligned} \text{cultivar}(\boldsymbol{y}) &= \arg\min_j \left\| \boldsymbol{y} - \boldsymbol{D}_j \boldsymbol{\alpha}^j \right\|_2^2 + \lambda \left\| \boldsymbol{p}_j \quad \boldsymbol{\alpha}^j \right\|_2^2, \\ & s.t. \ \ \mathbf{1}^{\mathrm{T}} \boldsymbol{\alpha}^j = 1, \ j = 1, ..., J \end{aligned} \tag{2}$$

where $\boldsymbol{p}_j$ is the Euclidean distances between the sample $\boldsymbol{y}$ and the columns of $\boldsymbol{D}_j$. The dictionary training and variety identification were performed using MATLAB (The MathWorks; Natick, MA, USA).

3.2 Results

3.2.1 Image pre-processing

Figures 5(a) and 5(b) show the images with focuses at the center and edge, respectively, of a rice grain. Figure 5(c) displays the image fused from Figs. 5(a) and 5(b). All the pixels of the fused image were in focus. Figures 5(d) and 5(e) demonstrate the foreground rice image and brush eliminated image. The contour of the grain can be readily detected in Fig. 5(e).

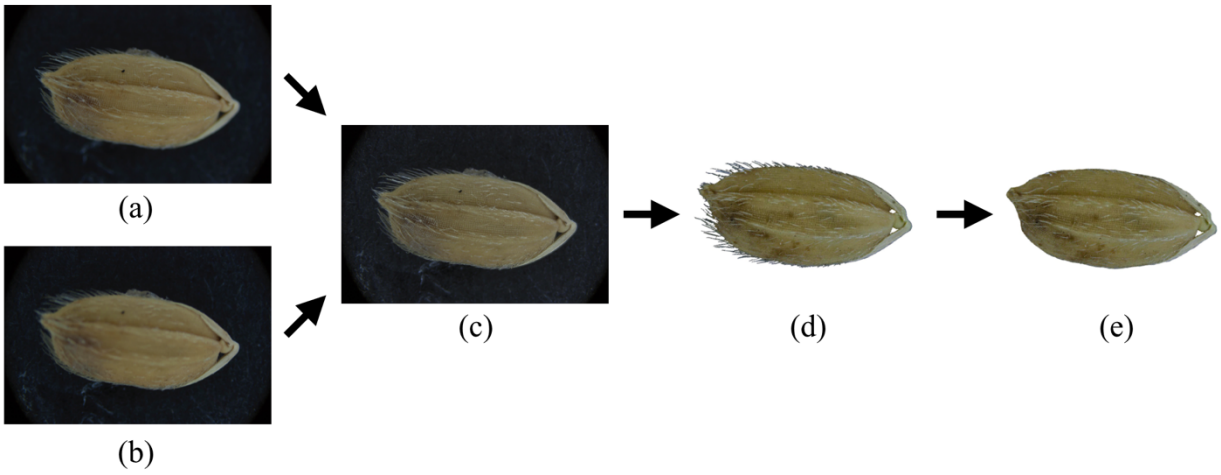



**Fig. 5.** Grain images with the (a) center and (b) edge in focus; (c) Image fused from (a) and (b); (d) Foreground image; (e) Brush-eliminated image.

3.2.2 Illustration of grain shape variation

The grain shape variations between the 30 varieties were illustrated. PCA indicated that the first two PCs accounted for 98.6% and 0.5% of the total variance, respectively. Each of the rest PCs accounted less than 0.3%. Hence, only the shape variations associated with the first two PCs were illustrated. Figure 6 shows the grain contour variations. Major variations were observed in grain length, width, and roundness. Figure 6 indicates that PC1 primarily corresponded to the grain size. The grains of large PC1 values were greater in volume compared with the grains of small PC1 values. PC2 principally corresponded to the roundness. The grains of large PC2 values were relatively slender, whereas the grains of small PC2 values were more spherical. Figure 7 shows the sample grain images of varieties Guan-Yin-Tsan, NSF-TV 107, Dourado Agulha, T1, and Tainung 67 (Appendix 1). The grains of Guan-Yin-Tsan were associated with mean contour of the 30 varieties. Moreover, the grains of Dourado Agulha and NSF-TV 107 were related to extreme PC1 values and the grains of Tainung 67 and T1 were linked to extreme PC2 values (-0.87, 1.88, -1.61 and 2.31, respectively).

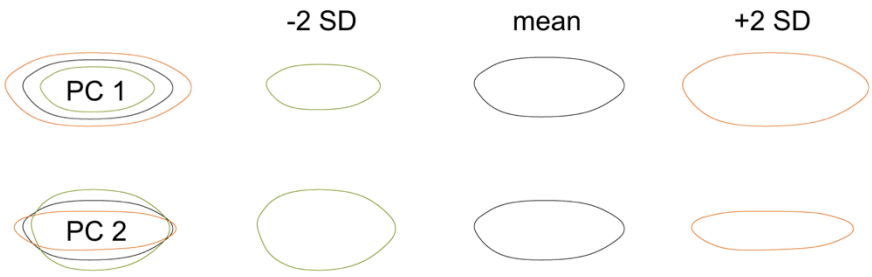



**Fig. 6.** Major grain shape variations. Each column shows the grain contours with altered PC values (mean−2SD, mean, and mean+2SD) as labelled. The left-hand column shows the three contours stacking together. It can be observed that PC2 primarily corresponds to the roundness of the grain.

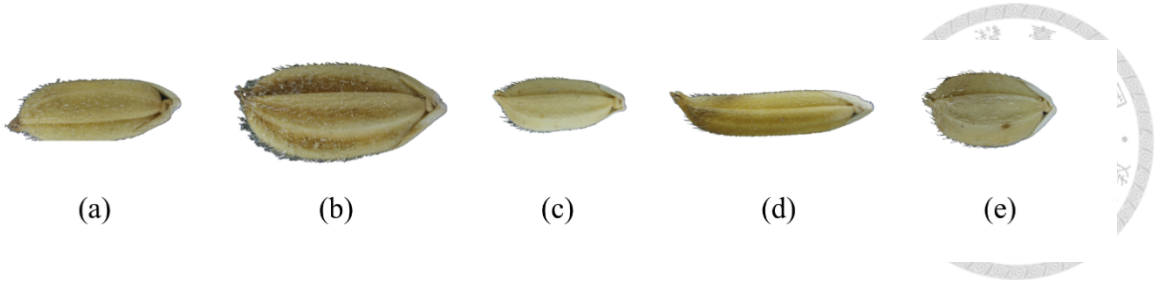



**Fig. 7.** Grains of varieties (a) Guan-Yin-Tsan, (b) NSF-TV 107, (c) Dourado Agulha, (d) T1, and (e) Tainung 67. The grains are dissimilar in size and roundness.

3.2.3 Grain color discrepancies

The color discrepancies between the grains of the 30 varieties were examined. In the analysis, PCA was applied to summarize the color distribution of each variety in the CIE L*a*b* color space (Fig. 8). The color distribution was illustrated using an ellipsoid. The ellipsoid centered at the mean L*, a*, and b* color parameters of a cultivar. The principal axes of the ellipsoid were set parallel to the first three PCs and were two SD of the PC scores in length. Figures 8(a) and 8(b) show the ellipsoids of the 30 varieties projected onto the chromaticity (a*-b*) and chromaticity-lightness (b*-L*) planes, respectively. The figures indicate that the grains were associated with a considerable variation in lightness (L*) and yellow-blue (b*). The colors of some varieties (e.g., Dosel, NSF-TV 107, Dular, and NSF-TV 160) were apparently distinct from the colors of some other varieties.

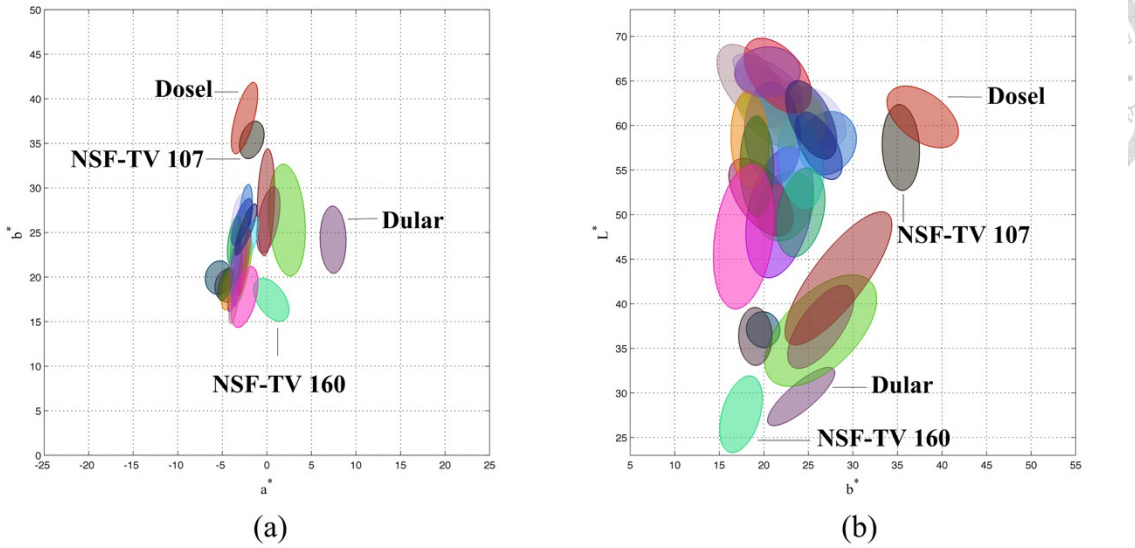



**Fig. 8.** Color distribution of the grains in the (a) chromaticity and (b) chromaticity-lightness planes. The ellipsoids represent the color ranges of the varieties. The ellipsoids are pseudo colored.

Figure 9 shows sample grain images of varieties (a) Dosel, (b) NSF-TV 107, (c) Dular, and (d) NSF-TV 160. The colors of Dosel and NSF-TV 107 were evidently brighter than the colors of Dular and NSF-TV 160 (Fig. 8b). The grain of Dosel was associated with stronger yellow compared with the grain of NSF-TV 160 (Fig. 8a). These observations indicated that the color traits can be used to effectively differentiate grains of certain varieties.


**Fig. 9.** The grains of varieties (a) Dosel, (b) NSF-TV 107, (c) Dular, and (d) NSF-TV 160. The grains are dissimilar in color. It can also be observed that traits such as the brush coverage and sterile lemma color are considerably different between varieties.

3.2.4 Classification performance

The dictionary of the SRC classifier was developed using the collected traits. The SRC classifier reached an overall accuracy of 89.1% in identifying the 30 varieties. The identification accuracies for the 30 varieties were showed in Fig. 10. The mean accuracy of the 30 varieties was 89.0%, and the standard deviation was 7.0%. Most accuracies were over 80.0%. Only the accuracy of Aswina 330 was relatively low, 74.0%. Few grains of Aswina 330 were misclassified as R101 owing to the similar appearance (Fig. 11). One limitation of using image-based approach for identifying the variety of rice grains is that the grains require to have distinct morphological and color traits. The SRC classifier achieved a relatively mediocre accuracy in classifying some varieties because the grain appearances of these varieties were similar to the appearances of some other varieties.

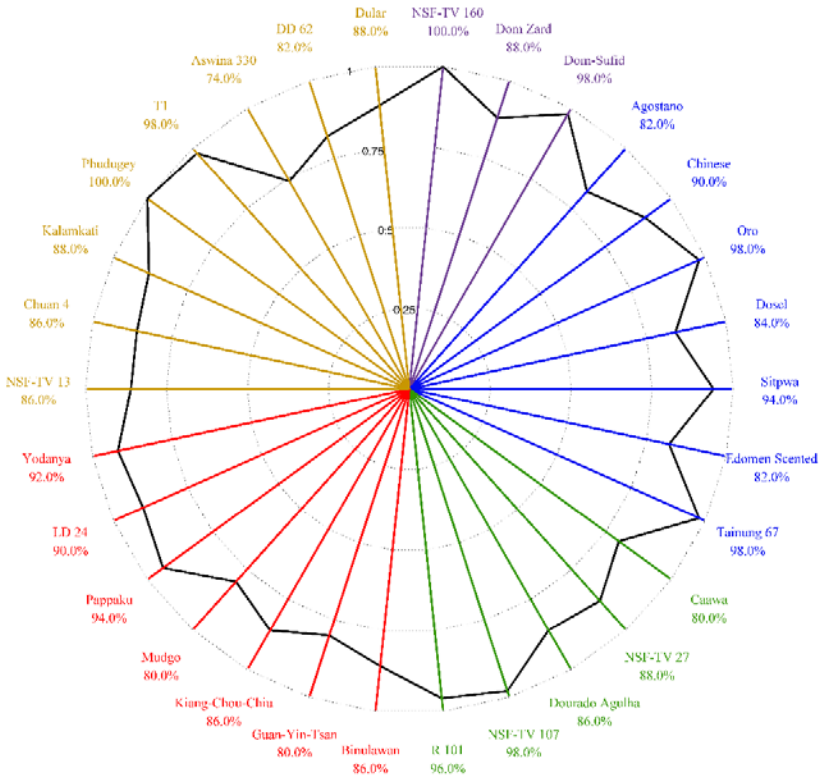



**Fig. 10.** The accuracies for the 30 varieties. Five different color, purple, blue, green, red, and yellow, stands for aromatic, temperate japonica, tropical japonica, indica, and aus, respectively. The black polygon shows the variation of accuracies.

(a) (b)

**Fig. 11.** The grains of varieties (a) R101, and (b) Aswina 330. The grains are similar in appearance.

3.3 Concluding Remarks

This study nondestructively distinguished the rice grains of 30 varieties using image analysis and SRC techniques. Morphological and color variations between rice grains of different varieties were observed. This prompted using image-based approaches for differentiating the rice grains. In the proposed approach, images of the rice grains were acquired using microscopy at a resolution of approximately 95 dots per millimeter. The high resolution made it possible to observe the fine details of the rice grains. Morphological, textural, and color traits of the grains were quantified. An SRC classifier was then developed to predict the varieties of the grains using the traits as the inputs. The classifier achieved an overall accuracy of 89.1%.

# CHAPTER 4. STUDYING THE GENOTYPE-PHENOTYPE ASSOCIATION OF RICE GRAINS

## 4.1 Material and Methods

### 4.1.1 Genotyping

Genotypic data used in this research were acquired from Rice Diversity website (http://ricediversity.org/). The genotypic data were genotyped by applying GeneChip® Rice 44K SNP Genotyping Array (Zhao et al., 2011) on Affymetrix (San Diego, CA). The 44,100 SNPs are distributed among the 12 chromosomes with the certain density, one SNP per 10 kb. The map position of the SNPs was converted from the MSU version 6.0 Nipponbare rice reference genome to MSU version 7.0 Nipponbare rice reference genome by using Gramene database.

### 4.1.2 Genome-wide association study

Generalized linear model (GLM) and mixed linear model (MLM) were proposed to estimate the relationships among the 44,100 SNPs and the selected traits of the 255 varieties (Appendix 2) by using TASSEL (Bradbury et al., 2007). The traits were axis aspect ratio, surface area, perimeter, thinness ratio, arc ratio, ratio of L1 to the major arc, ratio of L2 to the minor arc, the three color parameters for both grain body and sterile lemmas in CIE L*a*b* color space. The models utilized phenotype data and population structure to predict the continuous phenotypical traits. The models were subsequently evaluated by conducting permutation test. The permutation test shuffles the phenotype data and randomly assigns them to each varieties. Each permutation test can simulate a possible experiment result. The null hypothesis of permutation test is that the phenotypic data are not associated with genotypic data. If the null hypothesis was rejected, it is reasonable to believe that the correlation exists between traits and markers (Bush and

Moore, 2012). In this study, PCs and kinship were applied to genotype data to represent the population structure. The kinship matrix $\boldsymbol{K}$ was defined as the covariance matrix of genotype data $\boldsymbol{X}$, i.e.,

$$\boldsymbol{K}=\boldsymbol{X}\boldsymbol{X}^T=\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T)^T=\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{S}\boldsymbol{U}^T=\boldsymbol{U}\boldsymbol{S}(\boldsymbol{U}\boldsymbol{S})^T=\boldsymbol{R}\boldsymbol{R}^T. \quad (3)$$

In GLM, the traits were estimate by the linear combination of genotype data, and PCs. The five subpopulations clustered based on the top four PCs of genotypic data, and were well differentiated from each other (Zhao et al., 2011). The GLM can expressed as

$$\boldsymbol{y}=\boldsymbol{\mu}+\boldsymbol{x}_i\boldsymbol{\beta}+\boldsymbol{\varepsilon}, \quad (4)$$

where $\boldsymbol{y}\in\Re^n$ is the value of the continuous phenotypical traits, $\boldsymbol{\mu}$ is the scalar mean, $\boldsymbol{x}_i\in\Re^n$ is the ith column of $\boldsymbol{X}$, and $\beta$ is the regression coefficient, $\boldsymbol{\varepsilon}$ are normally distributed residual errors with zero means

In MLM, the traits were estimate by the linear combination of genotype data, PCs, and kinship. The MLM can expressed as

$$\boldsymbol{y}=\boldsymbol{\mu}+\boldsymbol{x}_i\beta+\boldsymbol{R}\boldsymbol{\gamma}+\boldsymbol{\varepsilon}, \quad (5)$$

where $\boldsymbol{\gamma}$ is normally distributed residual errors with zero mean. While adding kinship in the model, the inflation of p-values was largely reduced.

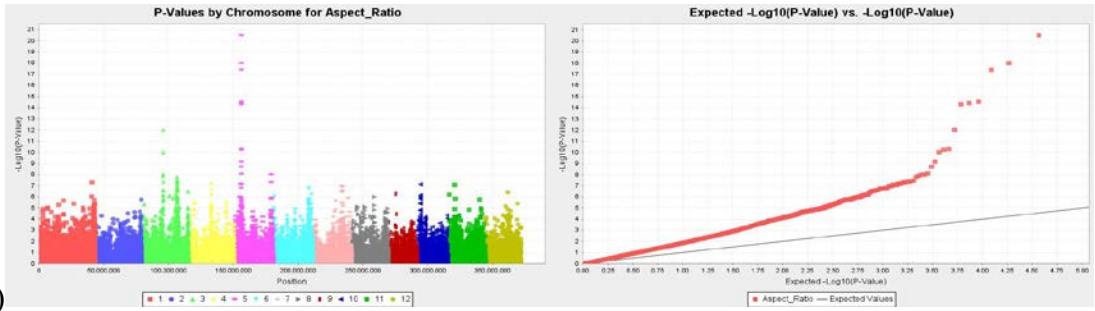
## 4.2 Results

### 4.2.1 Statistical models optimization

The GLM and MLM were constructed using the collected traits and 44K SNPs data The Manhattan plots and quantile-quantile (Q-Q) plots for both GLM and MLM of each trait were shown in Fig. 10 to Fig. 22. For each trait, the explanatory power of the models was

compared by visual inspection of Q-Q plots. In an ideal GWAS case where there are causal polymorphisms, the most of the observed p-values followed a uniform distribution, but the few observed p-values with a causal polymorphism will be extremely low. While the SNPs with low p-value shown in the Manhattan plots were located in the neighboring regions, the regions possibly contain the trait-related genes. According to the distribution of p-values in fig. 10, 12, and 18, the chromosomes contained SNPs with extreme low p-values were further observed in fig. 25 to 29. The fig. 25 to 29 showed that axis aspect ratio, perimeter, and parameter a* of grain body were possibly related with certain SNPs. For the rest of the traits, the GLM and MLM proposed in this study were not able to dissect the association between the phenotypic and genotypic variations of the varieties. The main limitations of the GLM and MLM used in this study may be that the traits ae controlled by many minor-effect genes and the environmental factors were not involved.

(a)




(b)

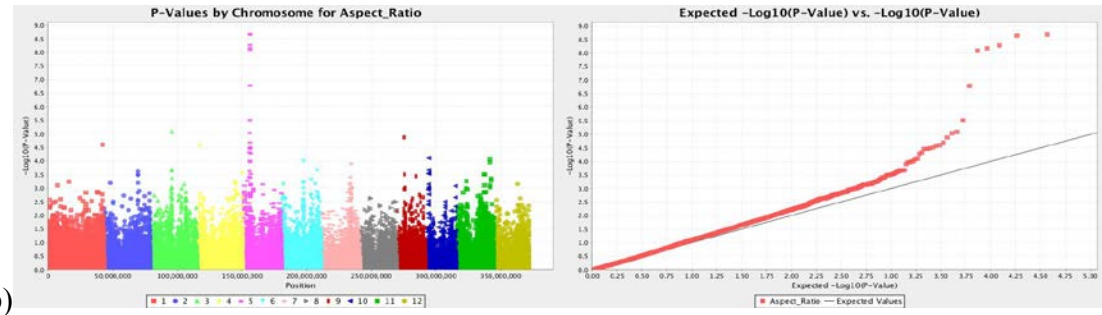



**Fig. 12.** Genome-wide association scan for axis aspect ratio to 44K SNPs. Manhattan plots and Q-Q plots were generated for (a) GLM and (b) MLM.

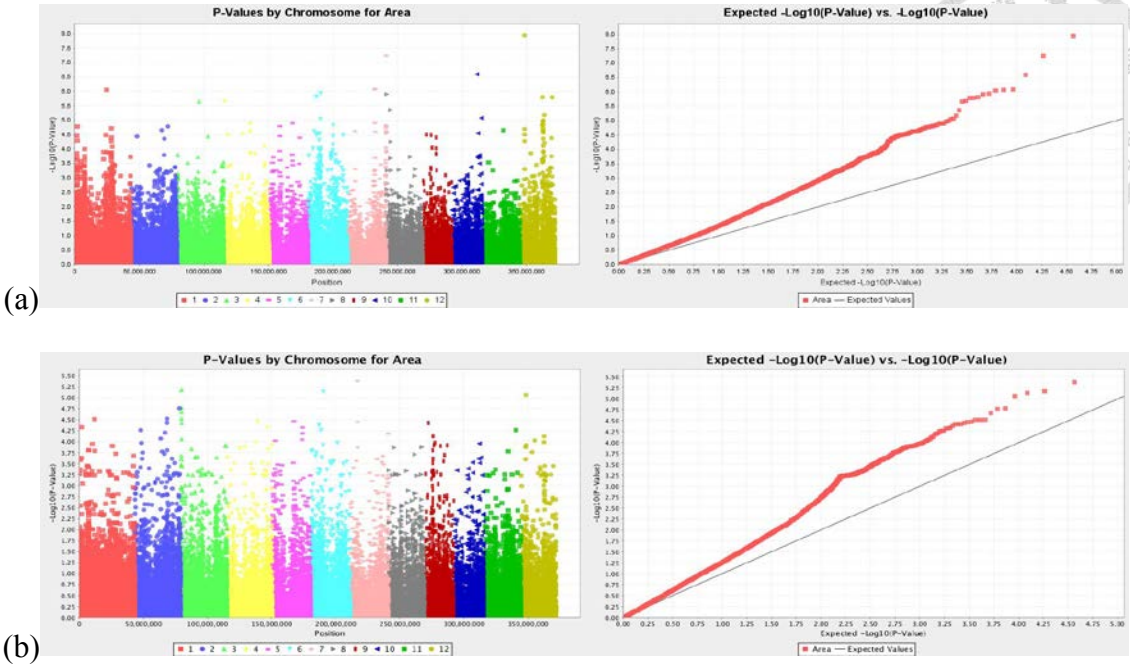



**Fig. 13.** Genome-wide association scan for surface area to 44K SNPs. Manhattan plots and quantile-quantile plots were generated for (a) GLM and (b) MLM.

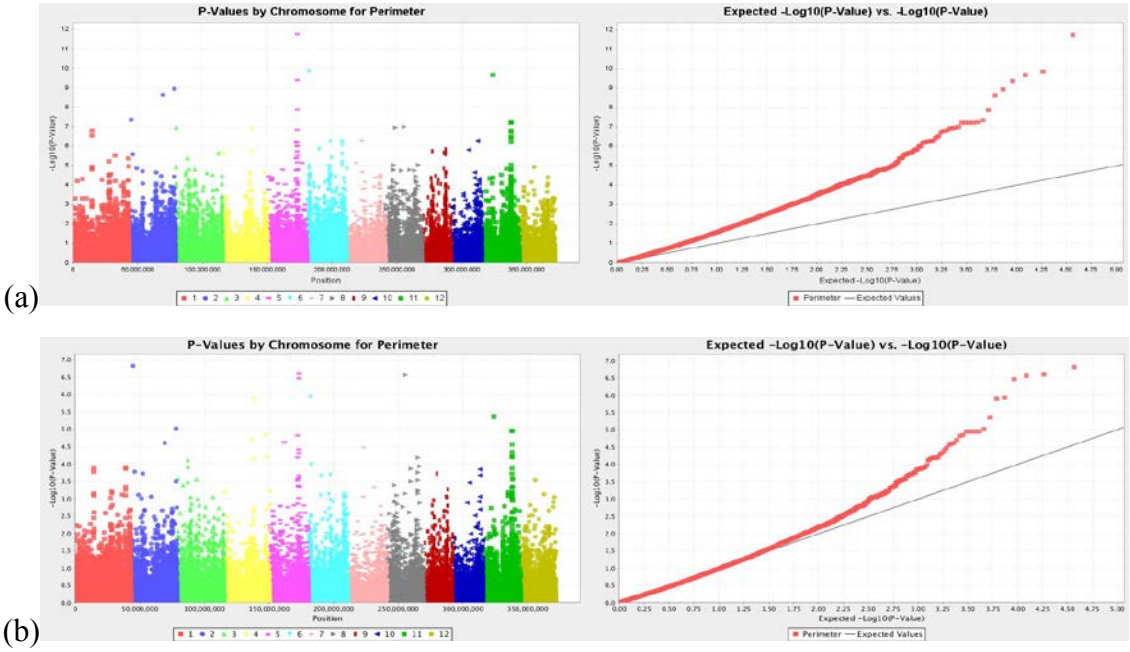



**Fig. 14.** Genome-wide association scan for perimeter to 44K SNPs. Manhattan plots and quantile-quantile plots were generated for (a) GLM and (b) MLM.

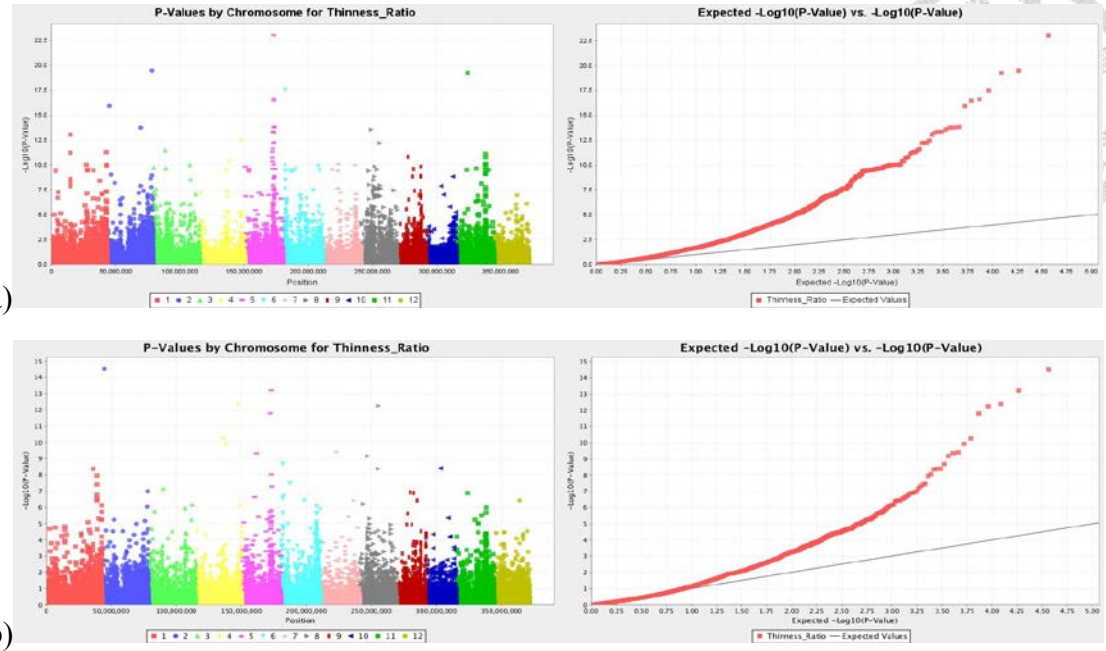



**Fig. 15.** Genome-wide association scan for thinness ratio to 44K SNPs. Manhattan plots and quantile-quantile plots were generated for (a) GLM and (b) MLM.

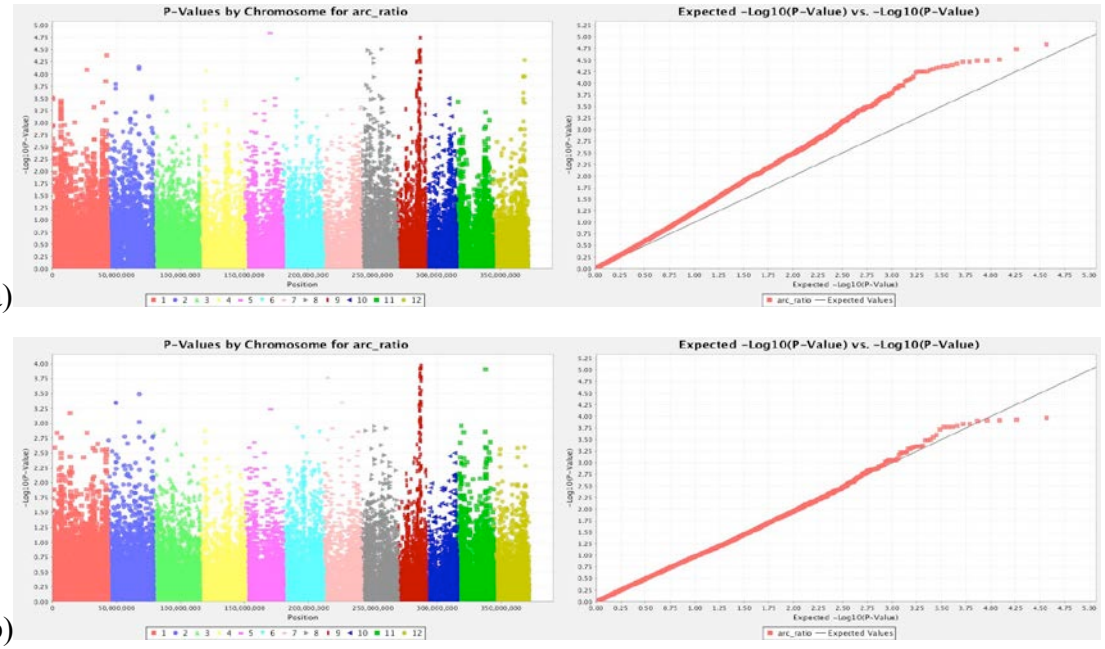



**Fig. 16.** Genome-wide association scan for arc ratio to 44K SNPs. Manhattan plots and quantile-quantile plots were generated for (a) GLM and (b) MLM.

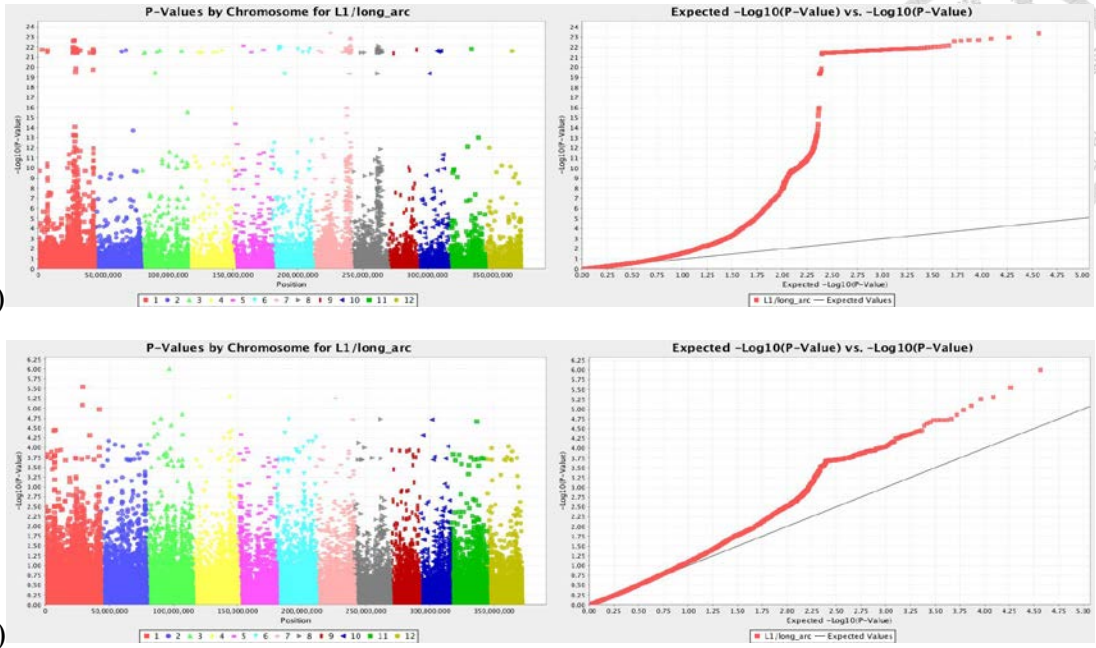



**Fig. 17.** Genome-wide association scan for ratio of L1 to the major arc to 44K SNPs. Manhattan plots and quantile-quantile plots were generated for (a) GLM and (b) MLM.

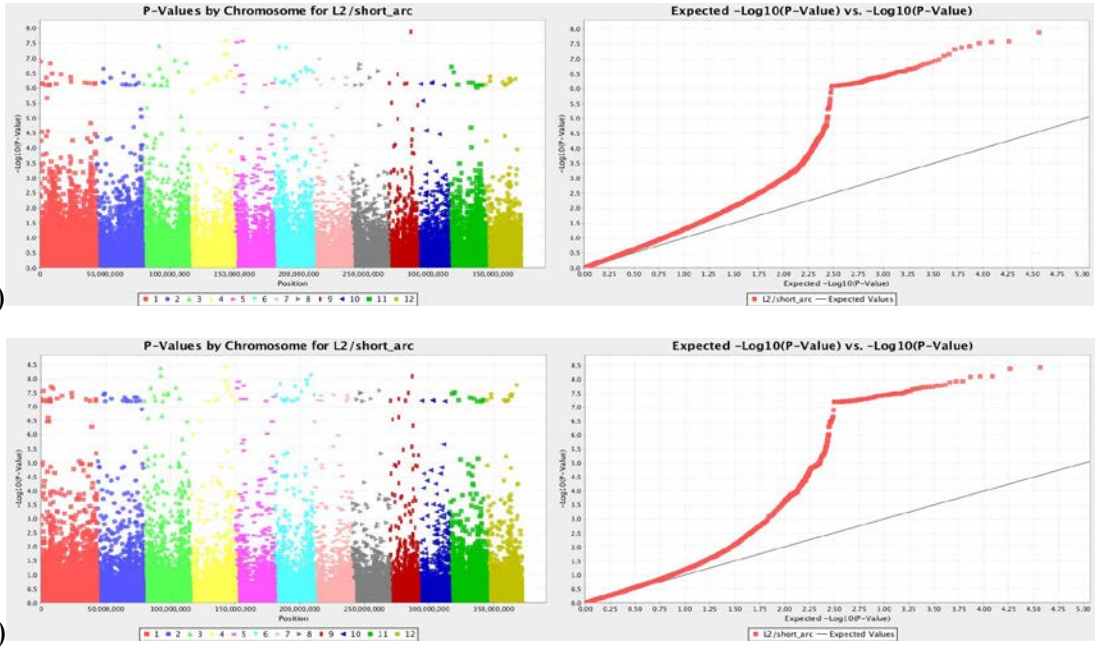



**Fig. 18.** Genome-wide association scan for ratio of L2 to the minor arc to 44K SNPs. Manhattan plots and quantile-quantile plots were generated for (a) GLM and (b) MLM.

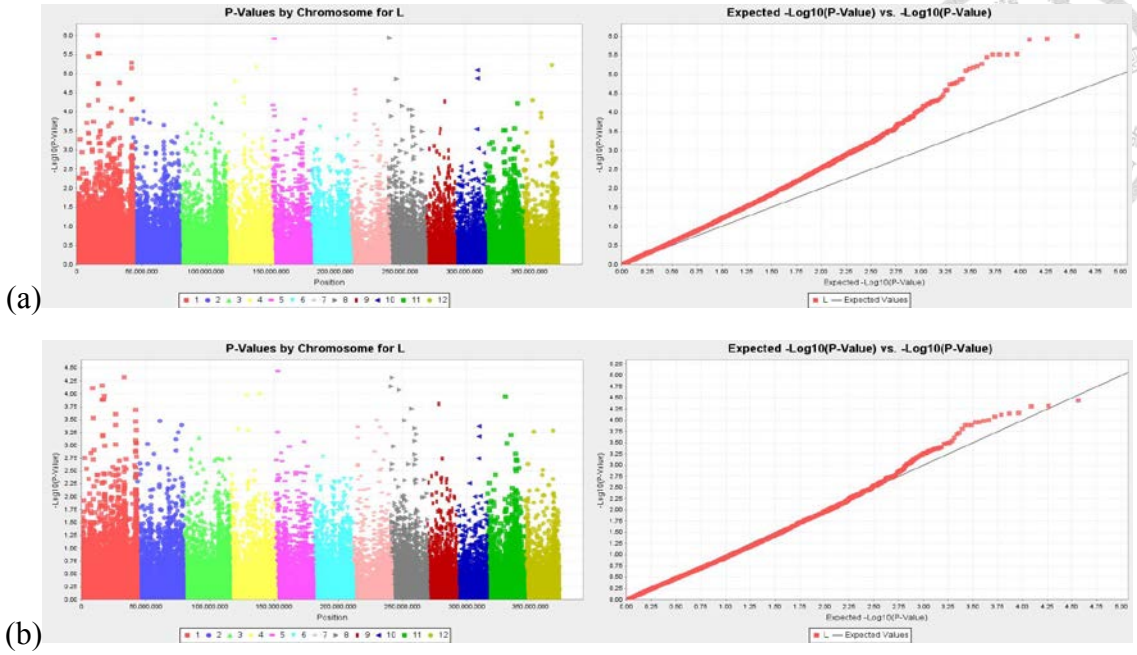



**Fig. 19.** Genome-wide association scan for L* of grain body to 44K SNPs. Manhattan plots and quantile-quantile plots were generated for (a) GLM and (b) MLM.

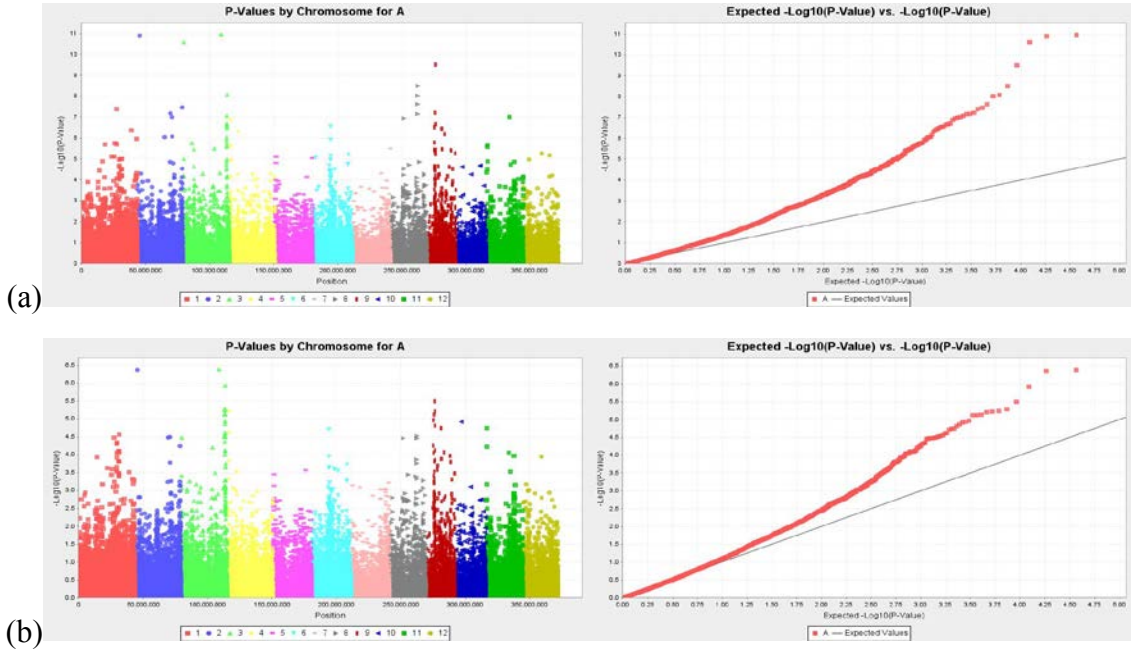



**Fig. 20.** Genome-wide association scan for a* of grain body to 44K SNPs. Manhattan plots and quantile-quantile plots were generated for (a) GLM and (b) MLM.

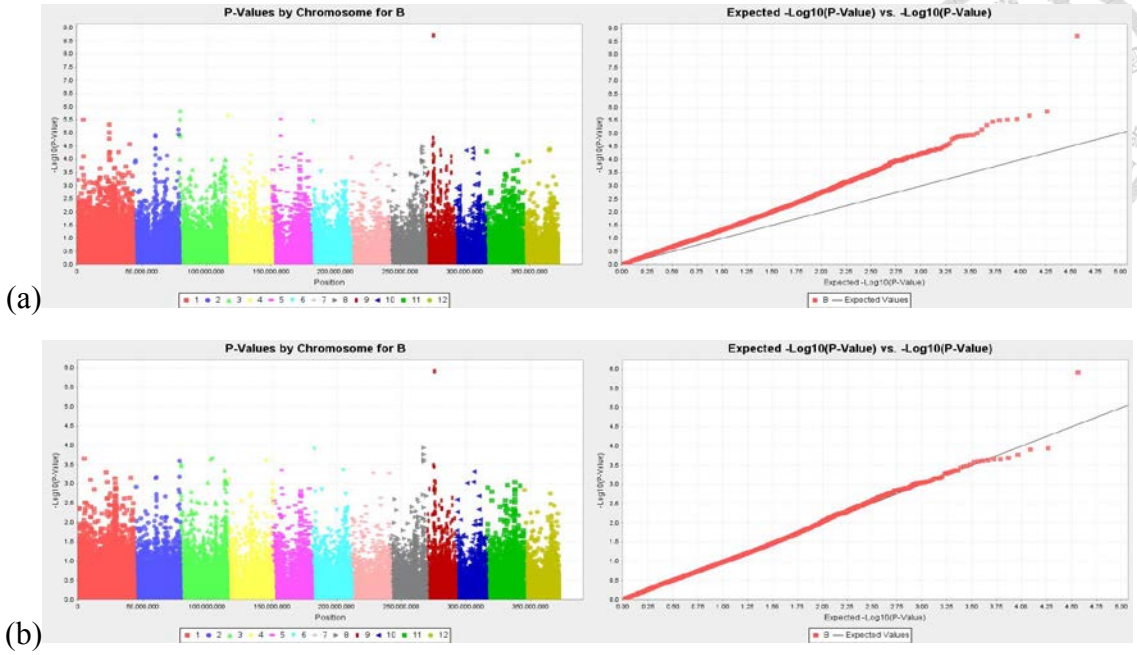



**Fig. 21.** Genome-wide association scan for b* of grain body of grain body to 44K SNPs. Manhattan plots and quantile-quantile plots were generated for (a) GLM and (b) MLM.

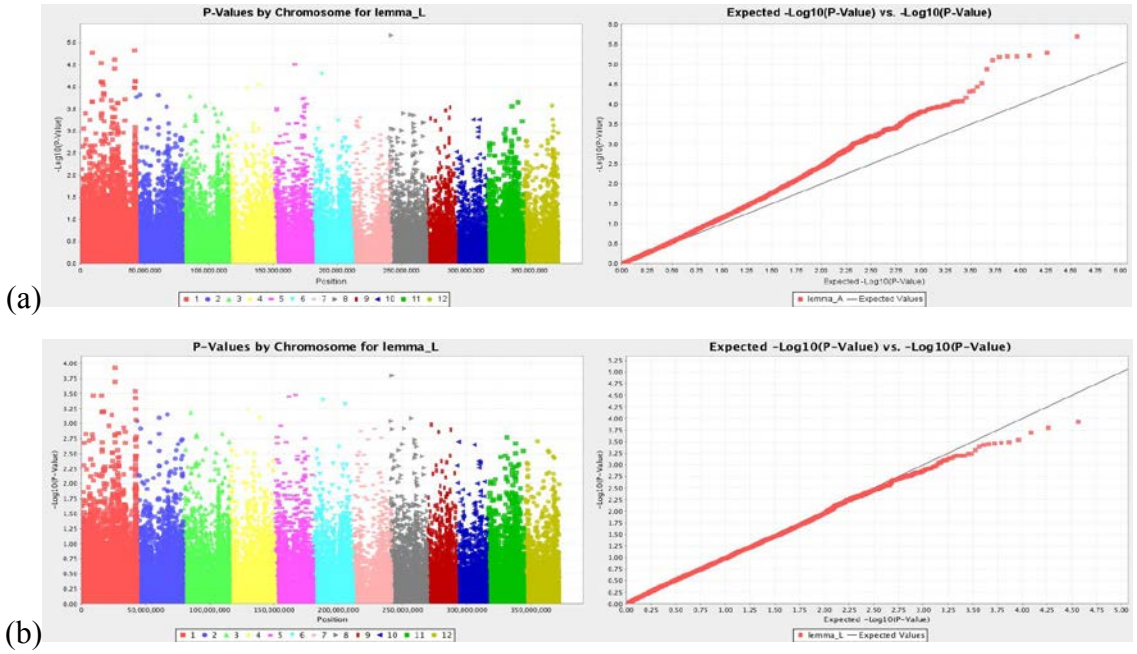
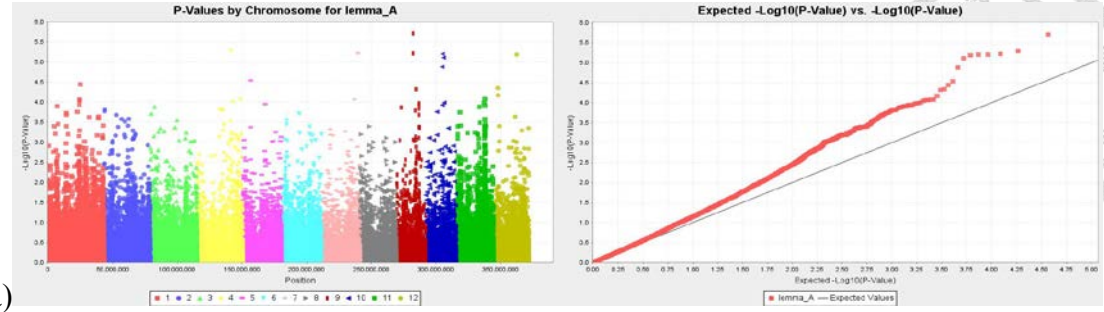



**Fig. 22.** Genome-wide association scan for L* of sterile lemmas to 44K SNPs. Manhattan plots and quantile-quantile plots were generated for (a) GLM and (b) MLM.

(a)




(b)

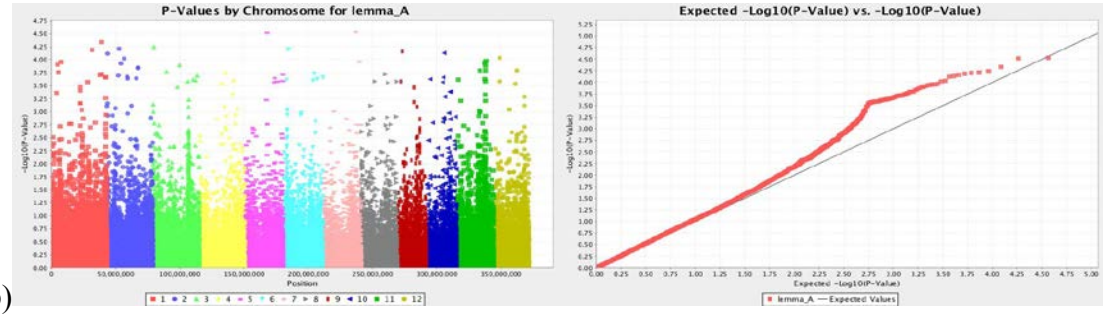
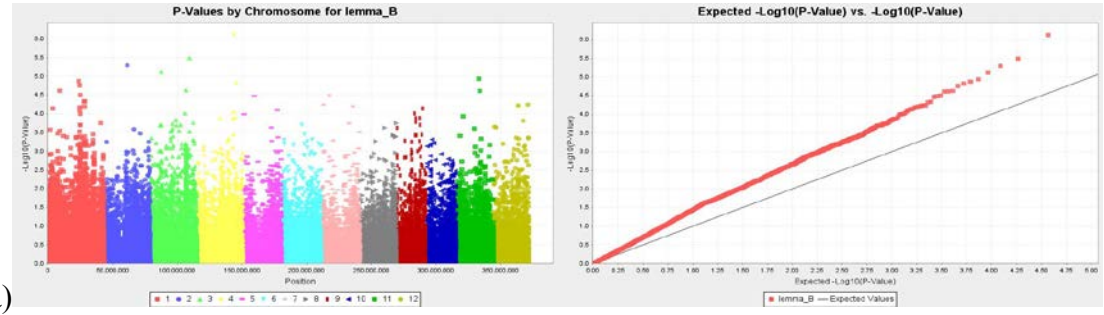



**Fig. 23.** Genome-wide association scan for a* of sterile lemmas to 44K SNPs. Manhattan plots and quantile-quantile plots were generated for (a) GLM and (b) MLM.

(a)




(b)

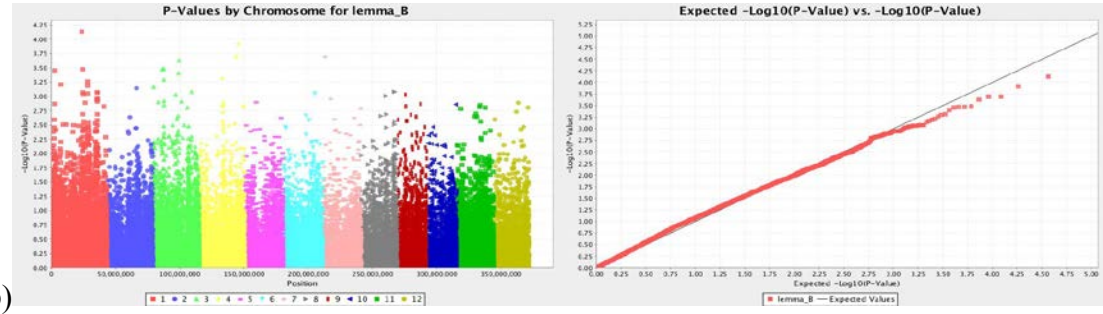



**Fig. 24.** Genome-wide association scan for b* of sterile lemmas to 44K SNPs. Manhattan plots and quantile-quantile plots were generated for (a) GLM and (b) MLM.

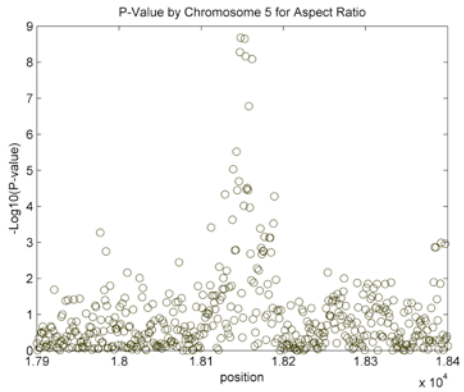



**Fig. 25.** Zoomed-in Manhattan plot generated for MLM of aspect ratio from position 17900 to 18400 on chromosome 5.

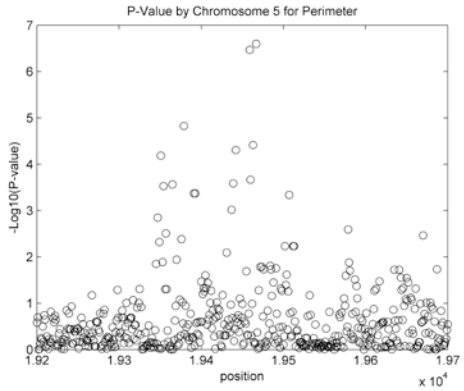



**Fig. 26.** Zoomed-in Manhattan plot generated for MLM of perimeter from position 19200 to 19700 on chromosome 5.

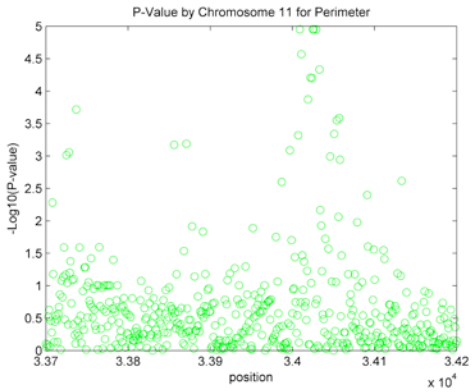



**Fig. 27.** Zoomed-in Manhattan plot generated for MLM of perimeter from position 33700 to 34200 on chromosome 11.

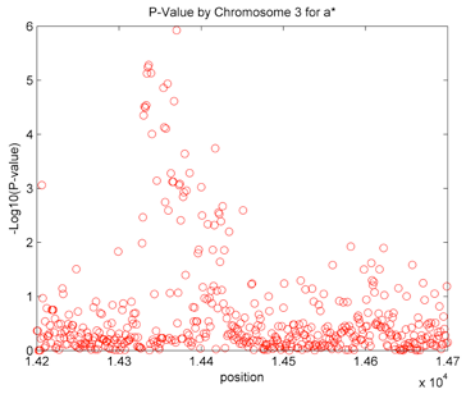



**Fig. 28.** Zoomed-in Manhattan plot generated for MLM of a* of grain body on from position 14200 to 14700 chromosome 3.

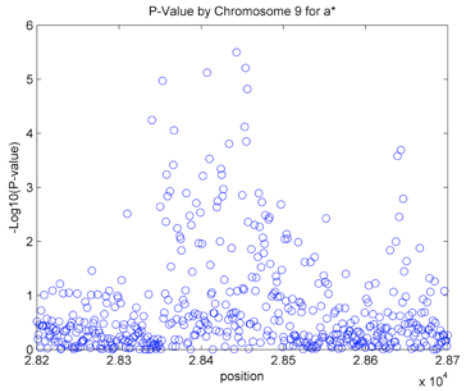



**Fig. 29.** Zoomed-in Manhattan plot generated for MLM of a* of grain body on from position 28200 to 28700 chromosome 9.

## 4.3 Concluding Remarks

This study fits the 13 traits into unified linear models for investigating the association between the phenotypic and genotypic variations of the 255 varieties. GLMs and MLMs were constructed for the traits to analyze the association. The results of Q-Q plots and Manhattan plots showed the potential to dissect the genetic basis of the traits.

# CHAPTER 5. CONCLUSION

This study aimed to differentiate rice grains of 30 varieties using locality constrained SRC and study the genotype-phenotype association of rice grains. In chapter 3, the rice grains of 30 varieties were nondestructively distinguished using image analysis and SRC techniques. In the proposed approach, images of the rice grains were acquired at a resolution of approximately 95 dots per millimeter. Morphological, textural, and color traits of the grains were quantified from the high resolution images. An SRC classifier was then developed to predict the varieties of the grains using the traits as the inputs. The classifier achieved an overall accuracy of 89.1%. In chapter 4, the association between the 13 phenotypic traits and genotypic variations of the 255 varieties were studied. GLMs and MLMs were constructed for the inspection of the association. The Q-Q plots and Manhattan plots of the results of permutation test showed the potential to dissect the genetic basis of the traits.

# REFERENCES

Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Statistics surveys 4, 40-79.

Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf: Speeded up robust features, Computer vision–ECCV 2006. Springer, pp. 404-417.

Becerra, V., Paredes, M., Gutiérrez, E., Rojo, C., 2015. Genetic diversity, identification, and certification of Chilean rice varieties using molecular markers. Chilean journal of agricultural research 75, 267-274.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23, 2633-2635.

Bush, W.S., Moore, J.H., 2012. Chapter 11: Genome-wide association studies. PLoS Comput Biol 8, e1002822.

Camelo-Méndez, G.A., Camacho-Díaz, B.H., del Villar-Martínez, A.A., Arenas-Ocampo, M.L., Bello-Pérez, L.A., Jiménez-Aparicio, A.R., 2012. Digital image analysis of diverse Mexican rice Varieties . J Sci Food Agric 92, 2709-2714.

Chen, H., Xie, W., He, H., Yu, H., Chen, W., Li, J., Yu, R., Yao, Y., Zhang, W., He, Y., 2014. A high-density SNP genotyping array for rice biology and molecular breeding. Molecular plant 7, 541-553.

Chuang, H.-Y., Lur, H., Hwu, K., Chang, M., 2011. Authentication of domestic Taiwan rice varieties based on fingerprinting analysis of microsatellite DNA markers. Botanical Studies 52, 393-405.

Cirillo, A., Del Gaudio, S., Di Bernardo, G., Galderisi, U., Cascino, A., Cipollaro, M., 2009. Molecular characterization of Italian rice Varieties . European Food Research and Technology 228, 875-881.

Costa, C., Menesatti, P., Paglia, G., Pallottino, F., Aguzzi, J., Rimatori, V., Russo, G., Recupero, S., Recupero, G.R., 2009. Quantitative evaluation of Tarocco sweet orange fruit shape using optoelectronic elliptic Fourier based analysis. Postharvest biology and Technology 54, 38-47.

Dillencourt, M.B., Samet, H., Tamminen, M., 1992. A general approach to connected-component labeling for arbitrary image representations. Journal of the ACM 39, 253-280.

Galloway, M.M., 1975. Texture analysis using gray level run lengths. Computer graphics and image processing 4, 172-179.

Garris, A.J., Tai, T.H., Coburn, J., Kresovich, S., McCOUCH, S., 2005. Genetic structure and diversity in *Oryza sativa* L. Genetics 169, 1631-1638.

Gonzalez, R.C., Woods, R.E., 2007. Digital image processing 3rd edition. Prentice Hall.

Haralick, R.M., Shanmugam, K., Dinstein, I.H., 1973. Textural features for image classification. Systems, Man and Cybernetics, IEEE Transactions (6), 610-621.

Harris, F.J., 1978. On the use of windows for harmonic analysis with the discrete Fourier transform. Proceedings of the IEEE 66, 51-83.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. Applied statistics 28 (1), 100-108.

Huang, Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., 2012. A map of rice genome variation reveals the origin of cultivated rice. Nature

490, 497-501.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nature genetics 42, 961-967.

Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., Li, W., Guo, Y., Deng, L., Zhu, C., 2012. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nature genetics 44, 32-39.

Hunter, A., 1975. The loss of community: An empirical test through replication. American Sociological Review 40 (5), 537-552.

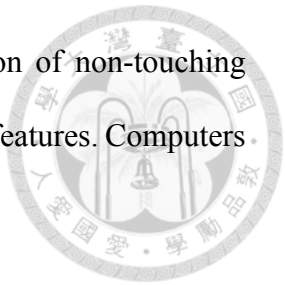
Jin, H.-y., Yang, X.-h., Jiao, L.-c., Liu, F., 2005. Image enhancement via fusion based on laplacian pyramid directional filter banks, Image Analysis and Recognition. Springer, pp. 239-246.

Kong, W., Zhang, C., Liu, F., Nie, P., He, Y., 2013. Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis. Sensors 13, 8916-8927.

Kovach, M.J., Sweeney, M.T., McCouch, S.R., 2007. New insights into the history of rice domestication. TRENDS in Genetics 23, 578-587.

Liu, M., Zhang, D., Shen, D., Initiative, The Alzheimer's Disease Neuroimaging Initiative, 2012. Ensemble sparse classification of Alzheimer's disease. NeuroImage 60, 1106-1116.

Majumdar, S., Jayas, D., 2000. Classification of cereal grains using machine vision: IV. Combined morphology, color, and texture models. Transactions of the ASAE 43, 1689-1694.



Mebatsion, H., Paliwal, J., Jayas, D., 2013. Automatic classification of non-touching cereal grains in digital images using limited morphological and color features. Computers and Electronics in Agriculture 90, 99-105.

Muja, M., Lowe, D.G., 2009. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. VISAPP (1) 2.

Pazoki, A., Farokhi, F., Pazoki, Z., 2014. Classification of rice grain varieties using two Artificial Neural Networks (MLP and Neuro-Fuzzy). Journal of Animal and Plant Sciences 24, 336-343.

Rohlf, F.J., Archie, J.W., 1984. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). Systematic Biology 33, 302-317.

Steele, K.A., Ogden, R., McEwing, R., Briggs, H., Gorham, J., 2008. InDel markers distinguish Basmatis from other fragrant rice varieties. Field Crops Research 105, 81-87.

Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y., 2010. Locality-constrained linear coding for image classification, Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference, San Francisco, CA pp. 3360-3367.

Wang, W., Chang, F., 2011. A multi-focus image fusion method based on Laplacian pyramid. Journal of Computers 6, 2559-2566.

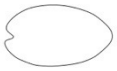
Wei, C.-P., Chao, Y.-W., Yeh, Y.-R., Wang, Y.-C.F., 2013. Locality-sensitive dictionary learning for sparse representation based classification. Pattern Recognition 46, 1277-1287.

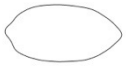
Xie, C., Zhang, J., Li, R., Li, J., Hong, P., Xia, J., Chen, P., 2015. Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning. Computers and Electronics in Agriculture 119, 123-132.

Yuan, X.-T., Liu, X., Yan, S., 2012. Visual classification with multitask joint sparse representation. Image Processing, IEEE Transactions on 21, 4349-4360.

Zhao, K., Tung, C.-W., Eizenga, G.C., Wright, M.H., Ali, M.L., Price, A.H., Norton, G.J., Islam, M.R., Reynolds, A., Mezey, J., 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nature communications 2, 467.

# APPENDIX 1

| No. | Name | NSFTV ID† | Subpopulation | Contour†† |
|---|---|---|---|---|
| 1 | NSF-TV 160 | 160 | *aromatic* |  |
| 2 | Dom Zard | 191 | *aromatic* |  |
| 3 | Dom-Sufid | 640 | *aromatic* |  |
| 4 | NSF-TV 13 | 13 | *aus* |  |
| 5 | Chuan 4 | 33 | *aus* |  |
| 6 | Kalamkati | 81 | *aus* |  |
| 7 | Phudugey | 131 | *aus* |  |
| 8 | T1 | 152 | *aus* |  |
| 9 | Aswina 330 | 312 | *aus* |  |
| 10 | DD 62 | 316 | *aus* |  |
| 11 | Dular | 651 | *aus* |  |
| 12 | Binulawan | 17 | *indica* |  |
| 13 | Guan-Yin-Tsan | 61 | *indica* |  |
| 14 | Kiang-Chou-Chiu | 90 | *indica* |  |
| 15 | Mudgo | 110 | *indica* |  |
| 16 | Pappaku | 126 | *indica* |  |

| | | | | |
|---|---|---|---|---|
| 17 | LD 24 | 298 | *indica* |  |
| 18 | Yodanya | 339 | *indica* |  |
| 19 | Agostano | 1 | *temperate japonica* |  |
| 20 | Chinese | 31 | *temperate japonica* |  |
| 21 | Oro | 118 | *temperate japonica* |  |
| 22 | Dosel | 296 | *temperate japonica* |  |
| 23 | Sitpwa | 338 | *temperate japonica* |  |
| 24 | Edomen Scented | 363 | *temperate japonica* |  |
| 25 | Tainung 67 | 641 | *temperate japonica* |  |
| 26 | Caawa | 22 | *tropical japonica* |  |
| 27 | NSF-TV 27 | 27 | *tropical japonica* |  |
| 28 | Dourado Agulha | 46 | *tropical japonica* |  |
| 29 | NSF-TV 107 | 107 | *tropical japonica* |  |
| 30 | R 101 | 310 | *tropical japonica* |  |

† Accession identification number of "Exploring the Genetic Basis of Transgressive Variation in Rice" project, National Science Foundation.

†† Mean contours of the 50 grains reconstructed using Fourier descriptors.

# APPENDIX 2

| No. | NSFTV ID | Subpopulation | Name |
|---|---|---|---|
| 52 | 85 | *aus* | Kasalath |
| 53 | 86 | *temperate japonica* | Kaw Luyoeng |
| 54 | 87 | *admixture* | Keriting Tingii |
| 55 | 88 | *aus* | Khao Gaew |
| 56 | 89 | *tropical japonica* | NSF-TV 89 |
| 57 | 90 | *indica* | Kiang-Chou-Chiu |
| 58 | 91 | *temperate japonica* | Kibi |
| 59 | 92 | *tropical japonica* | Kinastano |
| 60 | 94 | *temperate japonica* | Koshihikari |
| 61 | 98 | *tropical japonica* | L-202 |
| 62 | 99 | *tropical japonica* | LAC 23 |
| 63 | 100 | *admixture* | Lacrosse |
| 64 | 101 | *tropical japonica* | Lemont |
| 65 | 103 | *temperate japonica* | Luk Takhar |
| 66 | 104 | *temperate japonica* | Mansaku |
| 67 | 105 | *aus* | Mehr |
| 68 | 107 | *tropical japonica* | NSF-TV 107 |
| 69 | 109 | *indica* | MTU9 |
| 70 | 110 | *indica* | Mudgo |
| 71 | 113 | *temperate japonica* | Norin 20 |
| 72 | 114 | *admixture* | Nova |
| 73 | 116 | *tropical japonica* | NSF-TV 116 |
| 74 | 118 | *temperate japonica* | Oro |
| 75 | 120 | *tropical japonica* | OS6 |
| 76 | 121 | *temperate japonica* | Ostiglia |
| 77 | 126 | *indica* | Pappaku |
| 78 | 128 | *admixture* | Pato De Gallinazo |
| 79 | 131 | *aus* | Phudugey |
| 80 | 133 | *temperate japonica* | Rikuto Kemochi |
| 81 | 134 | *temperate japonica* | Romeo |
| 82 | 135 | *tropical japonica* | RT 1031-69 |
| 83 | 137 | *indica* | RTS14 |
| 84 | 138 | *indica* | RTS4 |
| 85 | 139 | *tropical japonica* | S4542A3-49B-2B12 |
| 86 | 140 | *admixture* | Saturn |
| 87 | 143 | *temperate japonica* | Shinriki |
| 88 | 144 | *temperate japonica* | Shoemed |
| 89 | 145 | *indica* | Short Grain |
| 90 | 146 | *indica* | Shuang-Chiang |
| 91 | 148 | *indica* | Sintane Diofor |
| 92 | 150 | *tropical japonica* | Sultani |
| 93 | 151 | *temperate japonica* | Suweon |
| 94 | 152 | *aus* | T1 |
| 95 | 153 | *aus* | T26 |
| 96 | 154 | *temperate japonica* | Ta Hung Ku |
| 97 | 156 | *indica* | Taichung Native 1 |
| 98 | 157 | *temperate japonica* | Taina Iku 487 |
| 99 | 160 | *aromatic* | NSF-TV 160 |
| 100 | 161 | *indica* | Te Qing |
| 101 | 162 | *indica* | TKM 6 |
| 102 | 165 | *tropical japonica* | Trembese |
| 103 | 166 | *admixture* | Tsipala 421 |
| 104 | 167 | *tropical japonica* | B6616A4-22-Bk-5-4 |
| 105 | 171 | *indica* | ZHE 733 |
| 106 | 172 | *indica* | Zhenshan 2 |
| 107 | 173 | *temperate japonica* | Nipponbare |
| 108 | 174 | *tropical japonica* | Azucena |
| 109 | 178 | *aus* | ARC 6578 |
| 110 | 179 | *temperate japonica* | Bellardone |
| 111 | 180 | *temperate japonica* | Benllok |
| 112 | 181 | *temperate japonica* | Bergreis |
| 113 | 182 | *admixture* | Blue Rose Supreme |
| 114 | 183 | *tropical japonica* | Boa Vista |
| 115 | 185 | *tropical japonica* | British Honduras Creole |
| 116 | 186 | *temperate japonica* | Bul Zo |
| 117 | 187 | *tropical japonica* | C57-5043 |
| 118 | 189 | *indica* | Criollo La Fria |
| 119 | 191 | *aromatic* | Dom Zard |
| 120 | 195 | *tropical japonica* | IRAT 13 |
| 121 | 259 | *admixture* | Sadri Tor Misri |
| 122 | 261 | *aus* | Shim Balte |
| 123 | 262 | *aus* | Halwa Gose Red |
| 124 | 263 | *temperate japonica* | Maratelli |
| 125 | 264 | *admixture* | Baldo |
| 126 | 265 | *temperate japonica* | Vialone |
| 127 | 267 | *temperate japonica* | Hatsunishiki |
| 128 | 269 | *indica* | Sundensis |
| 129 | 270 | *admixture* | Osogovka |
| 130 | 271 | *admixture* | M. Blatec |
| 131 | 274 | *tropical japonica* | Padi Pagalong |
| 132 | 276 | *aus* | Kaukau |
| 133 | 277 | *temperate japonica* | Gambiaka Sebela |
| 134 | 278 | *admixture* | C1-6-5-3 |
| 135 | 279 | *temperate japonica* | Kon Suito |
| 136 | 280 | *admixture* | Saku |
| 137 | 281 | *temperate japonica* | Patna |
| 138 | 282 | *temperate japonica* | Triomphe Du Maroc |
| 139 | 283 | *temperate japonica* | Chibica |
| 140 | 284 | *indica* | IR-44595 |
| 141 | 286 | *tropical japonica* | IITA 135 |
| 142 | 287 | *temperate japonica* | Zerawchanica karatals |
| 143 | 288 | *temperate japonica* | Italica Carolina |
| 144 | 289 | *temperate japonica* | Lusitano |
| 145 | 290 | *temperate japonica* | Amposta |
| 146 | 291 | *temperate japonica* | Toploea 70/76 |
| 147 | 292 | *temperate japonica* | Stegaru 65 |
| 148 | 293 | *admixture* | Tog 7178 |
| 149 | 295 | *temperate japonica* | Bombilla |
| 150 | 296 | *temperate japonica* | Dosel |
| 151 | 297 | *temperate japonica* | Bahia |
| 152 | 298 | *indica* | LD 24 |
| 153 | 299 | *indica* | SML 242 |
| 154 | 300 | *temperate japonica* | Sml Kapuri |
| 155 | 302 | *temperate japonica* | WIR 3039 |
| 156 | 303 | *temperate japonica* | Kihogo |
| 157 | 304 | *indica* | 519 |
| 158 | 305 | *admixture* | Doble Carolina Rinaldo Ba |
| 159 | 306 | *temperate japonica* | WIR 3764 |
| 160 | 307 | *temperate japonica* | Uzbekskij2 |
| 161 | 308 | *tropical japonica* | Llanero 501 |
| 162 | 309 | *tropical japonica* | Manzano |
| 163 | 310 | *tropical japonica* | R 101 |
| 164 | 311 | *temperate japonica* | 56-122-23 |
| 165 | 312 | *aus* | Aswina 330 |
| 166 | 313 | *indica* | BR24 |
| 167 | 314 | *aus* | CTG 1516 |
| 168 | 315 | *indica* | Dawebyan |
| 169 | 316 | *aus* | DD 62 |
| 170 | 317 | *aus* | DJ 123 |
| 171 | 318 | *aus* | DJ 24 |
| 172 | 319 | *aus* | DK 12 |
| 173 | 321 | *aus* | DM 56 |
| 174 | 324 | *aus* | DV 123 |
| 175 | 325 | *indica* | EMATA A 16-34 |
| 176 | 328 | *aus* | Jamir |
| 177 | 330 | *aus* | Khao Pahk Maw |
| 178 | 331 | *aus* | Khao Tot Long 227 |
| 179 | 333 | *temperate japonica* | Leuang Hawn |
| 180 | 334 | *temperate japonica* | Lomello |
| 181 | 335 | *admixture* | Okshitmayin |
| 182 | 336 | *aus* | Paung Malaung |
| 183 | 337 | *indica* | Sabharaj |
| 184 | 338 | *temperate japonica* | Sitpwa |
| 185 | 339 | *indica* | Yodanya |
| 186 | 340 | *admixture* | Berenj |
| 187 | 341 | *aus* | Shirkati |
| 188 | 342 | *tropical japonica* | Cenit |
| 189 | 343 | *admixture* | Victoria F.A. |
| 190 | 344 | *admixture* | Habiganj Boro 6 |
| 191 | 345 | *aus* | DZ 193 |
| 192 | 346 | *aus* | Karkati 87 |
| 193 | 350 | *tropical japonica* | Ligerito |
| 194 | 352 | *tropical japonica* | Guatemala 1021 |
| 195 | 354 | *admixtureture* | BALA |
| 196 | 355 | *temperate japonica* | ASD 1 |
| 197 | 356 | *indica* | JC 117 |
| 198 | 357 | *aus* | 9524 |
| 199 | 359 | *aus* | Surjamkuhi |
| 200 | 360 | *aus* | PTB 30 |
| 201 | 363 | *temperate japonica* | Edomen Scented |
| 202 | 366 | *temperate japonica* | Kiuki No. 46 |
| 203 | 367 | *admixture* | Sanbyang-Daeme |
| 204 | 369 | *aus* | Sathi |
| 205 | 371 | *aus* | Santhi Sufaid |
| 206 | 372 | *aus* | Sufaid |
| 207 | 375 | *tropical japonica* | Upland |
| 208 | 376 | *admixture* | Breviaristata |
| 209 | 379 | *tropical japonica* | Wanica |
| 210 | 380 | *temperate japonica* | Tainan-Lku No. 512 |
| 211 | 381 | *tropical japonica* | 325 |
| 212 | 384 | *tropical japonica* | 318 |
| 213 | 386 | *admixture* | Palmyra |
| 214 | 387 | *admixture* | M-202 |
| 215 | 390 | *admixture* | Cl 11026 |
| 216 | 391 | *tropical japonica* | Della |
| 217 | 392 | *tropical japonica* | Edith |
| 218 | 393 | *admixtureture* | La 110 |
| 219 | 394 | *tropical japonica* | Lady Wright Seln |
| 220 | 395 | *tropical japonica* | OS 6 |
| 221 | 396 | *tropical japonica* | Cocodrie |
| 222 | 397 | *tropical japonica* | Cybonnet |
| 223 | 612 | *admixtureture* | IR 64 |
| 224 | 616 | *indica* | RT0034 |
| 225 | 617 | *admixtureture* | MCr010277 |
| 226 | 618 | *admixture* | Pecos |
| 227 | 619 | *admixtureture* | Rosemont |
| 228 | 619 | *admixtureture* | Rosemont |
| 229 | 620 | *indica* | Jasmine 85 |
| 230 | 621 | *tropical japonica* | LaGrue |
| 231 | 622 | *admixture* | Bengal |
| 232 | 623 | *indica* | Shufeng 121-1655 |
| 233 | 624 | *tropical japonica* | Kaybonnet |
| 234 | 625 | *tropical japonica* | Katy |
| 235 | 626 | *indica* | C101A51 |
| 236 | 627 | *admixture* | Early |
| 237 | 628 | *tropical japonica* | Jefferson |
| 238 | 629 | *admixture* | Panda |
| 239 | 630 | *tropical japonica* | Saber |
| 240 | 633 | *indica* | Jing 185-7 |
| 241 | 635 | *tropical japonica* | Azucena |
| 242 | 636 | *indica* | Sadu Cho |
| 243 | 638 | *tropical japonica* | Moroberekan |
| 244 | 639 | *temperate japonica* | Nipponbare |
| 245 | 640 | *aromatic* | Dom-Sufid |
| 246 | 641 | *temperate japonica* | Tainung 67 |
| 247 | 642 | *indica* | Zhenshan 97B |
| 248 | 643 | *indica* | Minghui 63 |
| 249 | 644 | *indica* | IR 64 |
| 250 | 645 | *admixture* | M-202 |
| 251 | 646 | *admixture* | Swama |
| 252 | 647 | *tropical japonica* | Cypress |
| 253 | 648 | *indica* | Shan-Huang-Zhan-2 |
| 254 | 649 | *admixtureture* | FR13A |
| 255 | 650 | *admixtureture* | Aswina |