

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Assessment and Utilization of the Genetic Diversity in Rice (*Oryza sativa* L.)

Jin Quan Li and Peng Zhang

*State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University  
China*

## 1. Introduction

The basis for raising crop production and improving crop quality is to breed new varieties. The key to breed new varieties are largely depended on the breakthrough of mining the crop germplasm resources. Therefore, the research and utilization of crop genetic diversity plays an important role on crop improvement in the future. Previous researches have indicated that genetic bottleneck effects existed in the procedure of crop domestication and modern breeding, i.e. the allele variation within wild species and landrace would be lost and result in the reduction of gene diversity during domestication and breeding (Tanksly et al., 1997). The narrow genetic basis would lead to cultivars without resistance to new pests and virus and tolerance to bad environment as well as producing the platform effect of yield. These lost alleles in modern cultivars could only trace back to their original landrace and wild species and be recovered. The original landrace are close to cultivars and possess high genetic diversity and many exotic genes, therewith provide useful germplasm resources for crop breeding.

Identification, uses and conservation for the genetic diversity within crop germplasm resources are of importance for their sustainable use in plant breeding. The current rapidly development of bioinformatics, genomics, and molecular biology as well as conventional breeding methods provides useful means to mine the desirable genes in the resources.

Rice (*Oryza sativa* L.) feeds more than 50% of the world's population and is one of the most important crops in the world. Rice genetic resource is the primary material for rice breeding and makes a concrete contribution to global wealth creation and food security. Therefore, understanding its valuable genetic diversity and using it in rice genetic improvement is of importance for raising rice yield and the resistance to biotic and abiotic stress as well as improving rice quality to secure global food supplies. Furthermore, as a model plant of cereal family, two rice genome sequence map have been generated (Goff et al., 2002; Yu et al., 2002) and great progress has been made in gene mining with omics technology. Such researches helps to make use of rice genetic resources however in turn requires to make further insights of rice genetic diversity.

China is well known as an origin center of cultivated rice, with abundant rice genetic resources. As early as 1920-1964, Ying Ting, An academician of Chinese academy of science, collected more than 7128 rice landrace from all over China as well as some main rice

cultivated countries. As far as we know, the collection is one of the earliest collections for rice germplasm resources and therefore we named it as Ting's collection (Lu et al., 2006). Therefore, this chapter aims to explore effective methods on mining the exotic genes within these novel rice germplasm resources.

## 2. Ting's rice germplasm collection

The Ting's rice germplasm collection consists of 7128 accessions, which was collected and conserved by Academician and Professor Ying Ting during 1920-1964 from 20 different provinces of China as well as from North Korea, Japan, Philippines, Brazil, Celebes, Java, Oceania, and Vietnam (Fig. 1). Most of them are rice landraces and possess high genetic diversity. Due to that it is one of the earliest systematically rice collections in China and covered most of the Chinese rice cultivated regions, it could serve as an representative for the genetic diversity of Chinese rice germplasm resources.

Most accessions were characterized for taxonomical, geographical, morphological and agronomical descriptors, recorded by the previous laboratory of rice ecology of Chinese academy of agricultural sciences, South China Agricultural College and Guangdong academy of agricultural sciences, China (1961-1965). These recorded traits include 20 unordered qualitative traits, i.e. origin of variety, *indica* vs. *japonica*, paddy vs. upland, waxy vs. non-waxy, grain shape, rice color, grain quality, leaf color, leaf margin color, leaf cushion color, auricle color, inner sheath color, outer sheath color, stem color, leaf-green color, stigma color, glume-tip color, sterile lemma color and glume color; 14 ordered qualitative traits, i.e. early- or late-season, type of maturity, shattering habit, awn, awn length, leaf face pubescence, leaf base pubescence, flag-leaf angle, erect vs. bending leaf, compact vs. loose stem, panicle shape, compact vs. loose rachis-braches, sparse vs. dense glume hair, compact vs. loose glume hair; and 15 quantitative traits, i.e. culm length, culm size, thickness of culm wall, the second internode length, number of panicles per plant, panicle length, panicle size, number of seeds per panicle, grain length, grain length/width ratio, grain size, flag leaf length, flag leaf width, length of elongated uppermost internode, growth duration. These data provide a good basis for studying their phenotypic genetic diversity as well as core collection construction based on the phenotypes.

## 3. Genetic diversity of phenotypes of Chinese rice germplasm resources

About 6500 accessions of rice germplasm resources from the Ting's collection with well passport data were selected and studied for their genetic diversity of phenotypes. The origin, type and distribution of these accessions are listed in Table 1. All the studied rice accessions were classified as five regions, i.e. South China, Central China, Southwest China, Northwest China, and North China.

The genetic diversity index ( $I$ ) was calculated by: 
$$I = \frac{-\sum_i \sum_j P_{ij} \log P_{ij}}{N}$$
, where  $P_{ij}$  are the frequency of  $j$ th phenotypes for  $i$ th traits, and  $N$  is the total number of traits.

The average genetic diversity index for the five rice cultivated regions from minimum to maximum are: Southwest China > Central China > Northwest China > South China > North

China for Unordered qualitative traits, North China > Northwest China > South China > Central China > Southwest China for ordered qualitative traits, and South China > North China > Central China > Northwest China > Southwest China for quantitative traits (Fig.2).



Fig. 1. Rice germplasm resources in Ting’s collection and their regeneration. Upper left, preparing seeds for sowing; upper middle, sowing in the nursery field; upper right, transplanted in the field; bottom left, measuring the traits and harvest; bottom middle, a global view for the germplasm regenerating field and their genetic diversity; bottom right, seed cool room with air condition, where the rice seeds are conserved in the block jars.

Region	<i>Indica</i>	<i>Japonica</i>	Paddy	Upland	Waxy	Non-Waxy	Early-seasonal	Late-seasonal
South China	2250	247	2311	58	2304	180	1094	1276
Central China	1450	343	1783	9	1720	73	1415	378
Southwest China	887	768	767	286	715	200	130	338
Northwest China	75	294	363	0	342	21	363	0
North China	70	125	189	6	177	16	195	0
Total	4732	1777	5413	359	5258	490	3197	1992

Table 1. The origin, type and distribution of rice germplasm resources

The results show that the most abundant genetic diversity for rice qualitative traits among the five rice cultivated regions are Southwest China, which is relevant to the geographic location and climate in Yunnan province and Guizhou province of China. The reasons for it might be the high variation in different sea levels (76~2700m) which results in several different climate zone. However, for the genetic diversity in quantitative traits, South and Central China are the highest ones. The reason might be that South and central China are

the main rice production area in our country traditionally and the quantitative traits were greatly improved by the farmers and breeders through thousands of years selection.

The results indicated that the rice germplasm resources from Southwest China might help to raise the rice genetic diversity for qualitative traits, such as grain colors, grain quality, etc. For the improvement of quantitative traits, the rice germplasm resources from South and Central China might contribute to the aim more than other regions.

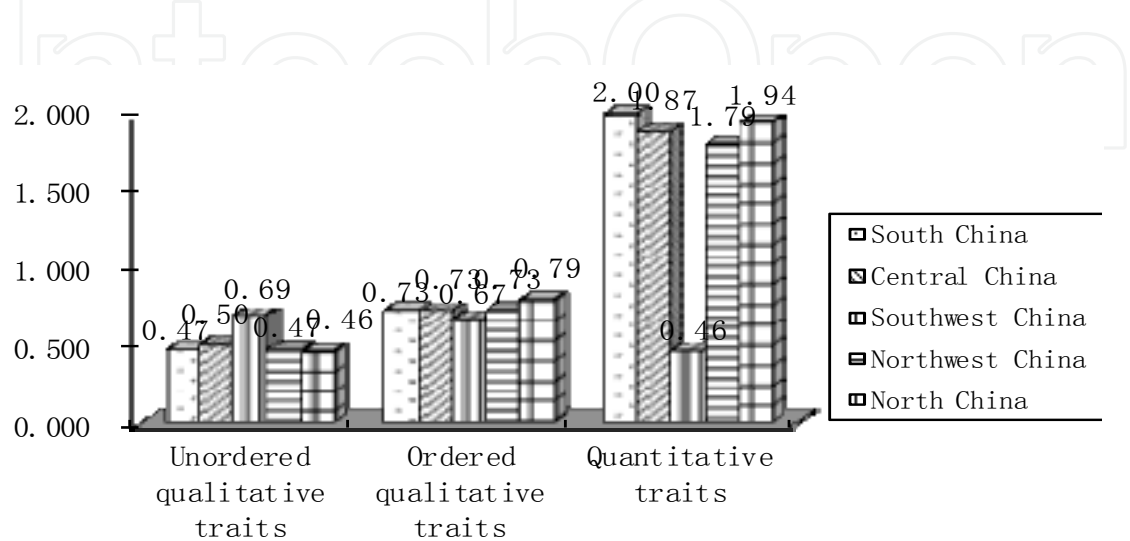


Fig. 2. The average genetic diversity index for rice germplasm resources from different cultivated regions

4. Genome-wide distribution of genetic diversity assessed with SSR markers

A subset containing 150 accessions were taken from the whole collection (described below) and were genotyped with 274 genome-wide distributed SSR markers. Gene diversity for the varieties in the subset is 0.544. Among them, *Indica* rice shows a higher gene diversity (0.484) than that of *Japonica* rice (0.454). Similarly, non-waxy rice shows a higher gene diversity (0.540) than that of waxy rice (0.515). However, early-seasonal rice shows a higher gene diversity (0.546) than that of late-seasonal rice (0.510) in our case.

Cultivated rice has been intensively selected during its domestication and breeding. Consequently, the genomic regions controlling traits of economic importance are expected to be shaped by this selection. Therefore, characterizing the genome-wide distribution of genetic diversity of cultivated rice germplasm which has been selected for different traits, such as waxy vs. non waxy might help to identify the genes controlling these traits. To do so, as one example, gene diversity was calculated for the waxy rice as well as non-waxy rice for each marker separately across the genome. Similarly, a measurement for genetic distance, modified Roger’s distance (MRD) between waxy and non-waxy rice was calculated on an individual marker basis.

Our results indicated that gene diversity for waxy and non-waxy rice varied across the genome (Fig.3). A different degree of divergence (as measured by MRD) between these two germplasm types was observed across the genome (Fig.4).



The unequal distribution of genetic diversity across the genome could be explained by the selection history of the different genome regions. Therewith, the genome-wide distribution maps of genetic diversity might be a first step to identify the target genes or regions selected during breeding history. For example, genes related to waxy and non-waxy rice might be present in the most divergent genomic regions between these two germplasm types. Common genes under selection in the breeding program of the both germplasm types (e.g. disease resistant genes) might be present in the genomic regions showing the same level of gene diversity and low MRD.

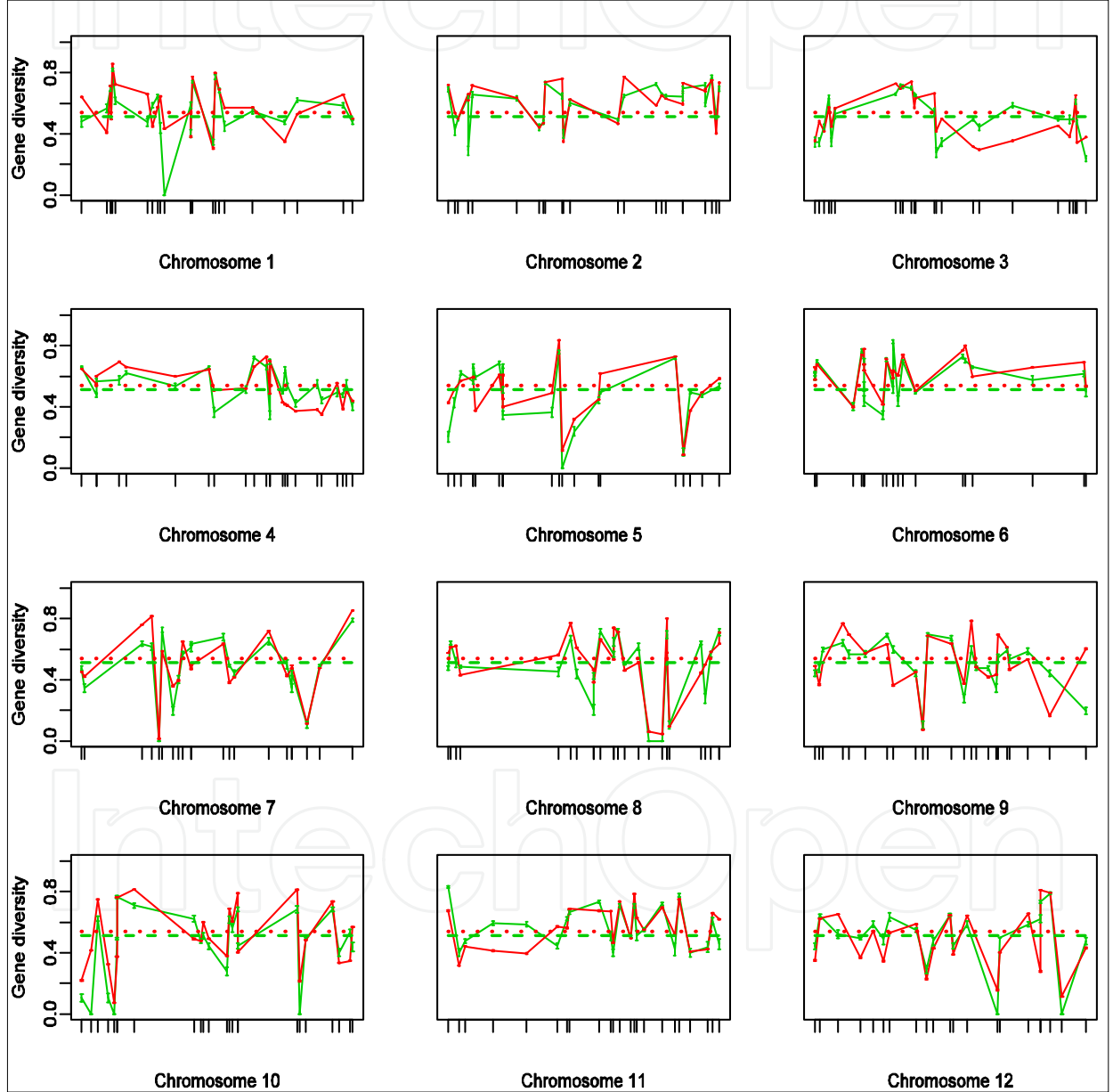


Fig. 3. Green and red lines indicate gene diversity of non-waxy and waxy rice, respectively. Dashed lines indicate the average gene diversity of the corresponding germplasm type. Vertical lines at each point indicate standard error multiplied by 100 which were calculated by bootstrapping across genotypes. Vertical lines at the x axis indicate genetic map positions of the SSR marker on the chromosome.

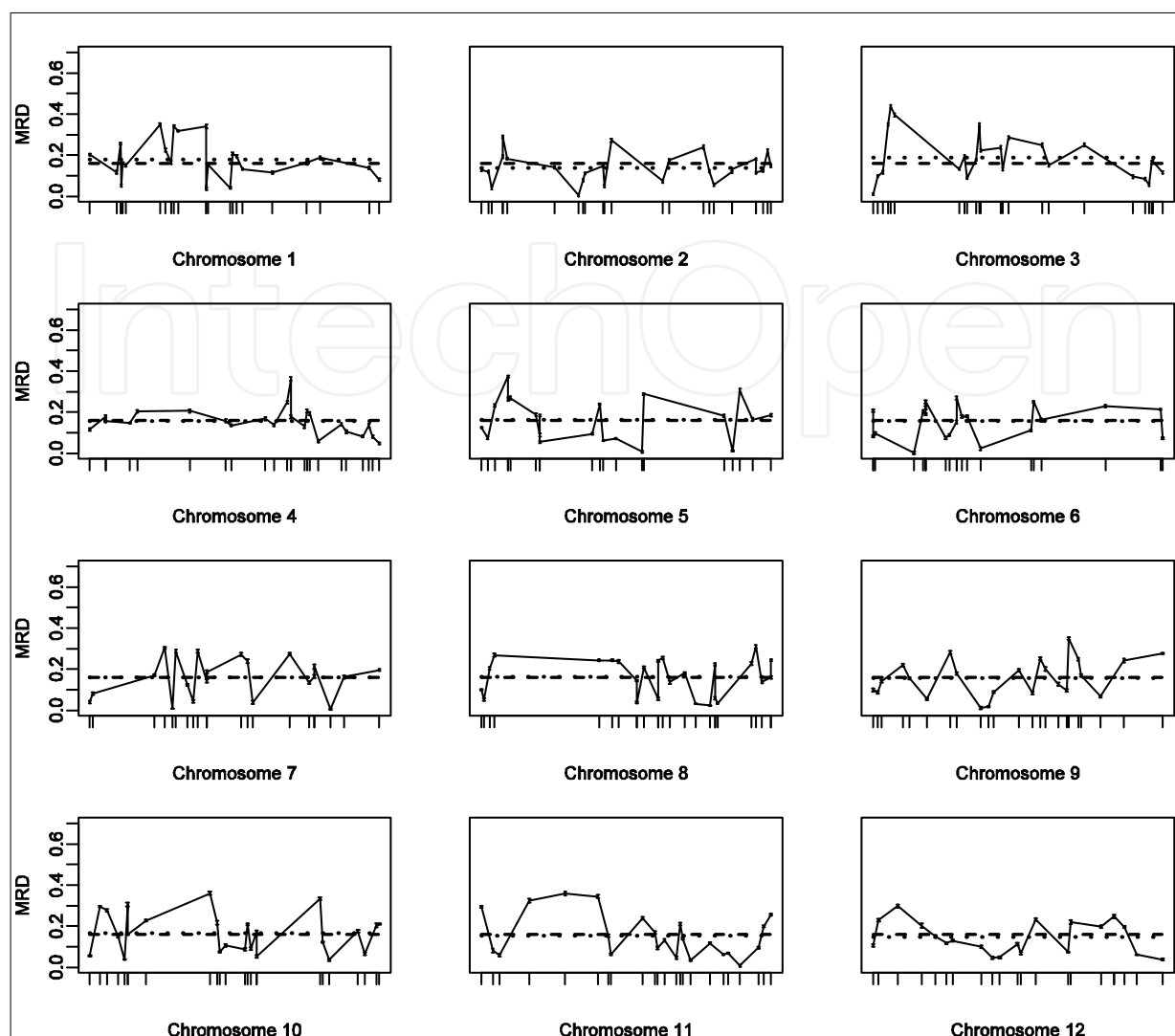


Fig. 4. Modified Roger's distance (MRD) between waxy and non-waxy rice across the genome. Dashed lines indicate average MRD across the genome and dotted lines average MRD for each chromosome. Vertical lines at the x axis indicate genetic map positions of the SSR markers on the chromosome.

## 5. Constructing a core collection to make use of genetic diversity

A large number of accessions in the germplasm collection (7128 accessions in our case) makes it difficult to choose the most promising ones for utilization. One feasible method is the development of core collections. A core collection is intended to contain, with a minimum repetitiveness, the maximum genetic diversity of a crop species and its wild relatives (Brown, 1989a, 1989b; Frankel and Brown, 1984). The development of a core collection could enhance the utilization of germplasm collections in crop improvement programs and simplify their management. Selection of an appropriate sampling strategy is an important prerequisite to construct a core collection with appropriate size in order to adequately represent the genetic spectrum and maximally capture the genetic diversity in available crop collections. Our studies were aimed to evaluate how sample size, clustering methods, sampling methods, and different data types affected the construction of core

collection and tried to find out an optimal strategy concerning the above factors for core collection construction.

By using three sampling strategies, three kinds of trait data, eight hierarchical clustering methods, and 15 kinds of different sampling proportions were applied to choose the optimal constructing strategies. Analysis of variance (ANOVA) and multiple comparisons were applied to compare different strategies. In order to choose the optimal constructing strategies, 12 evaluated parameters were applied to evaluate the validity of sampling.

The ANOVA analysis showed significant difference for different clustering methods, data types, sample size and sampling methods (Table 2). Furthermore, there were significant interaction effects between these factors except clustering method and sampling method, sampling size and sampling method. The results indicated that these factors as well as their interaction would affect the construction strategy and must be considered carefully.

For different sampling methods, preferred sampling plus multiple clustering and sampling on the degree of variation is better than preferred sampling plus multiple clustering and random sampling, and the completely random sampling is the worst; For the eight clustering methods, clustering analysis with shortest distances has the best of genetic diversity index, average Shano-Weaver index, phenotype retained percentage, and variance of phenotypic frequency; For the three different data types (qualitative trait data, quantitative trait data, intergrated qualitative and quantitative trait data), the core collections constructed by integrated qualitative and quantitative trait data retain the greatest genetic diversity and is the best one. For the sampling rate, the sampling rate of 3.4% ~ 24% is sufficient to retain the greatest genetic diversity of the initial population (Table 4-6).

Finally, a core collection was constructed by using preferred sampling plus multiple clustering and sampling on the degree of variation, clustering analysis with shortest distances, and based on the integrated qualitative and quantitative data. This core collection contains 150 accessions out of 2262 original collection with full recorded data from Ting’s collection, accounting for 6.6% of the initial collection.

Origin of variation	df	SS	MS	F value	P value
Clustering method	7	0.07	0.01	29.82**	<.0001
Data type	2	0.49	0.25	720.66**	<.0001
Sampling size	15	0.19	0.01	37.5**	<.0001
Sampling method	1	0.003	0.003	8.78**	0.003
Clustering method×data type	14	0.16	0.01	33.74**	<.0001
Clustering method×sample size	105	0.09	0.0008	2.43**	<.0001
Clustering method×sampling method	7	0.0006	0.0001	0.25	0.97
Data type×sample size	30	0.81	0.03	79.51**	<.0001
Data type×sampling method	2	0.01	0.0038	11.02**	<.0001
Sample size×sampling method	15	0.0035	0.0002	0.68	0.81
Error	615	0.21	0.0003		
Sum of variation	831	3.18			

Table 2. Analysis of variance for Shanno-weaver diversity index of different subsets



Clustering method*	Ratio of phenotype retained	Variance of phenotypic frequency	Average diversity index
1	0.9983A	0.1526C	2.0142A
2	0.9968BC	0.1619B	2.013A
3	0.9973BC	0.1630B	1.9958C
4	0.9977ABC	0.1659A	2.0032B
5	0.9978BA	0.1616B	2.0138A
6	0.9968C	0.1620B	2.0123A
7	0.9970BC	0.1621B	2.0129A
8	0.9972BC	0.1617B	2.0129A
9	0.9512D	0.1636B	1.974D

\* Clustering with 1. shortest distance, 2. complete linkage method, 3. median distance, 4.the centroid method, 5. average linkage, 6. the flexible-beta method, 7. the flexible method, 8. Ward’s minimum variance method, and 9. completely random method.  
\*\* the data with different alphabet show significance at 0.01 level.

Table 3. Multiple comparison for different clustering methods\*\*

Sampling method*	Diversity index For qualitative trait	Average diversity index
1	0.581A	2.005B
2	0.577A	2.013A
3	0.558B	1.974C

\* 1. preferred sampling plus multiple clustering and randomly sampling, 2. preferred sampling plus multiple clustering the clustering with sampling on degree of variation, 3. sampling with completely random method  
\*\* the data with different alphabet show significance at 0.01 level.

Table 4. Multiple comparison for different sampling methods\*\*

Data type*	Ratio of phenotype retained	Variance of phenotypic frequency	Average diversity index
1	0.996A	0.1509C	2.019A
2	0.9923B	0.18A	1.99B
3	0.9958A	0.1524B	2.017A

\*1. integrated qualitative and quantitative trait data, 2. qualitative trait data, and 3. quantitative trait data  
\*\* the data with different alphabet show significance at 0.01 level.

Table 5. Multiple comparison for different data types\*\*

6. Association mapping with the rice core collection

Though a large number of exotic genes exist in crop germplasm resources, the rich genetic variations in crop germplasm resources haven’t been fully explored and utilized because of being lack of appropriate statistical methods.

In general, conventional method for mining gene from germplasm is linkage mapping. Identifying QTLs by linkage mapping needs to construct one or several segregating

Sample size	Ratio of phenotype retained	Variance of phenotypic frequency	Average diversity index
Original collection	1A	0.1581F	1.994I
1450	1A	0.1581F	1.993I
1350	0.9995A	0.1589EF	1.996HI
1250	0.9995A	0.1593EFD	1.999HG
1150	0.9995A	0.1604EFD	2GF
1050	0.9995A	0.161EFDC	2.003EF
950	0.999BA	0.1624BDC	2.005EDF
850	0.998B	0.1638BAC	2.006ED
750	0.998B	0.1649BA	2.006ED
650	0.998B	0.1651BA	2.008D
550	0.997C	0.1655BA	2.008D
450	0.996C	0.1648BA	2.012C
350	0.993D	0.1655BA	2.018B
250	0.99E	0.1659A	2.02A
150	0.983F	0.1615EDC	2.02A
50	0.996G	0.149G	2.02A

\*\* the data with different alphabet show significance at 0.01 level.

Table 6. Multiple comparison for different sample size\*\*

populations by crossing between parents (e.g., F<sub>2</sub>, Double haploid, Backcross population) and linkage mapping would be done in these segregation populations. The accuracy of QTL mapping is dependent largely on selected but limited parents and only two or a few alleles from the parents were detected. Moreover, abundant genetic variation stored in germplasm have not been developed and utilized due to lack of appropriate statistical methods. Provided using conventional QTL linkage mapping method for mining the abundant genetic variations in a large germplasm resources population, it is necessary to make diallel crossing with all studied accessions, which is hard to develop such mapping population and would take much more time, cost, space and analysis.

An alternative, association mapping based on linkage disequilibrium (LD) analysis might be an effective way to identify the function of the gene (or targeted high-resolution QTL), which has been successfully applied in human genetics to detect QTL coding for simple as well as complex diseases (Corder et al., 1994; Kerem et al., 1989). This method uses the LD between DNA polymorphisms and genes underlying traits. LD refers to the non-random combination among different genetic markers. The main mechanism for LD existing in a population is genetic linkage among different loci. Therefore, it is possible to detect QTLs by identifying LD between markers loci and potential QTLs. Through detecting abundant genetic markers loci locating in genome or those nearby candidate genes, the loci which link tightly with QTLs and show correlated to QTLs can be found. The application of association

mapping to plant breeding appears to be a promising approach to overcome the limitations of conventional linkage mapping (Kraakman et al., 2004).

Furthermore, choice of an appropriate germplasm to maximize the genetic diversity and the number of historical recombinations and mutation events (and thus reduce LD) within and around the gene of interest is critical for the success of association analysis (Yan et al., 2011). As described above, core collections are the core subset of the original collections with minimum samples while having the maximum genetic variability contained within the gene pool. Therefore, association mapping with a core collection population helps to catch as more phenotypic variation as possible and would make use of both the advantages of association mapping and core collection, thus could be an effective way to mine and utilize the abundant genetic diversity in the crop germplasm resources.

### 6.1 Population structure and LD pattern

Population structure is an important component in association mapping analysis because it can reduce both type I and II errors between molecular markers and traits of interest in an inbreeding specie. Moreover, low level of LD could lead to impractical whole-genome scanning because of the excessive number of markers required. Furthermore, the resolution of association studies in a test sample depends on the structure of LD across the genome. Therefore, information about the population structure and extent of LD within the population is of fundamental importance for association mapping.

The rice core collection consisting of 150 varieties were genotyped with 274 SSR markers. Based on the genotyping data, STRUCTURE software was run to detect the number of subgroups within the core collection population and assign the varieties into different subgroups with the membership probability of 0.80 as a threshold. To compare and confirm the STRUCTURE subgroups, a additional principal component analysis was done.

STRUCTURE indicated that the entire population could be divided into two subgroups (i.e. SG 1 and SG 2) (Fig. 5). With the membership probabilities of  $\geq 0.80$ , 111 varieties were assigned to SG 1, 21 varieties were assigned to SG 2 and 18 varieties were retained to the admixed group (AD) (Fig. 5). Principal component analysis confirmed the population structure, i.e. the varieties from SG 1 and SG 2 located in two distinct clusters, and those from AD located between the two subgroups (Fig. 5). The varieties in SG 1 are mainly *indica* rice, and those in SG 2 *japonica* rice, whereas those in AD intermediate (*indica*- or *japonica*-inclined) rice. Furthermore, the varieties from the same cultivated zone were clustered closely.

LD measured as squared correlation of allele frequencies ( $r^2$ ) between loci pairs in the core collection and different germplasm types were calculated (Table 7). The average  $r^2$  between linked loci (the loci at the same chromosome) varied between 0.0188 and 0.1. Using the 95% quantile of  $r^2$  between unlinked loci pairs as a threshold, 6.23% linked loci pairs were in significant LD. For different germplasm types (*indica*, *japonica*, early-seasonal, late-seasonal, waxy, non-waxy rice), the percentage of loci pairs in LD varied between 5.33 and 6.36%. LD ( $r^2$ ) against genetic map distance (cM) between linked loci pairs was plotted and a nonlinear regression of  $r^2$  vs. genetic map distance according to Heuertz et al., (2006) was performed (Fig. 6). The LD decays against the genetic distance, which indicated the linkage might be the main reason for the causes of LD. The LD decays to the threshold, i.e. the 95% quantile of  $r^2$  between unlinked loci pairs, at 1.03 cM in the entire collection (Table 7). The cut-off decay

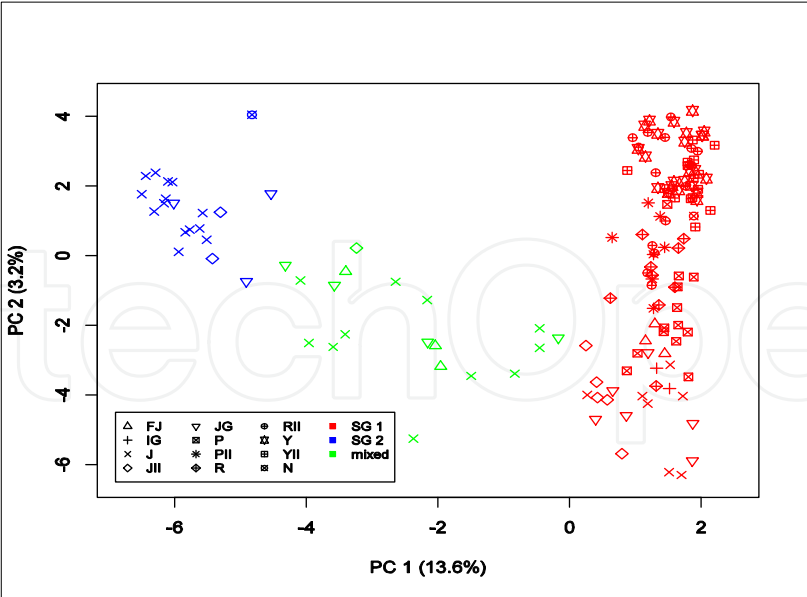


Fig. 5. Principal component analysis for the rice core collection combined with STRUcTURE subgroup assignment. PC 1 and PC 2 refer to the first and second principal components, respectively. The numbers in parentheses refer to the proportion of variance explained by the principal components. Symbols indicate different type of rice, and colors indicate different subgroups from STRUcTURE software. FJ- Foreign *japonica*, IG- glutinous *Indica*, J- early seasonal *Japonica*, J II- late seasonal *Japonica*, JG- glutinous *Japonica*, P- early seasonal *Indica* from Pearl river region or south China, P II- late seasonal *Indica* from Pearl river region, R- early seasonal Red grain rice, R II- late seasonal Red grain rice, Y- early seasonal *Indica* from Yangtze River region, Y II- late seasonal *Indica* from Yangtze River region, and N- Unknown origin.

Type	Sample size	Average R <sup>2</sup>	Percentage of loci pairs in LD	Cut-off (cM)
Entire collection	150	0.0330	6.23	1.03
Indica	90	0.0188	5.33	0.89
Japonica	60	0.0570	6.37	1.10
Early-seasonal	86	0.0500	6.22	1.00
Late-seasonal	64	0.0330	5.95	1.04
Waxy	18	0.1000	6.06	0.87
non-waxy	132	0.0350	6.36	1.01

Table 7. Linkage disequilibrium (measured as R<sup>2</sup> value) for linked loci, percentage of linked loci pairs in LD, and the cut-off decay distance for the core collection and different germplasm types.

distances for *indica*, *japonica*, early-seasonal, late-seasonal, waxy, non-waxy rice were 0.89, 1.10, 1.00, 1.04, 0.87, and 1.01cM, which were about 200-500kb physical distance. The results indicated that choice of the core collection could maximize the number of historical recombinations and mutation events and thus reduce LD within and around the gene of interest which is critical for the success of association analysis (Yan et al., 2011). Such short

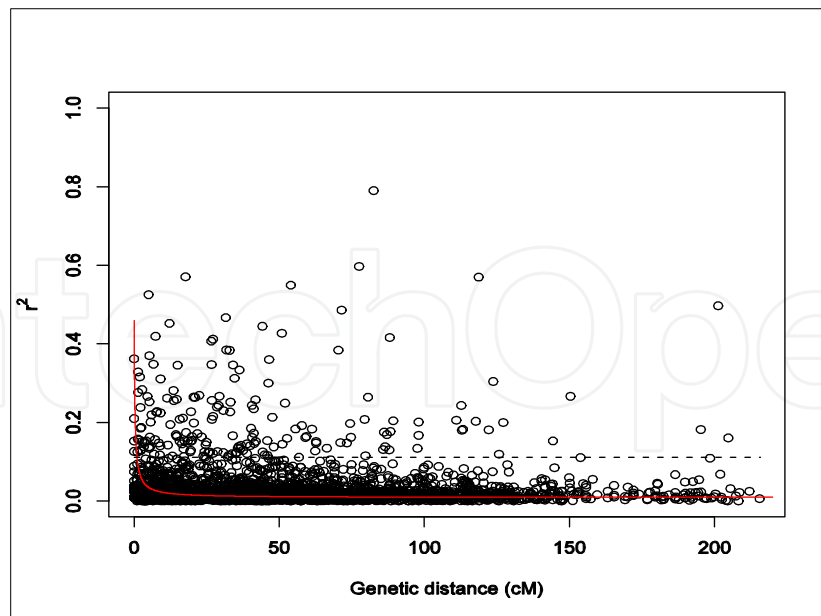


Fig. 6. Plot of linkage disequilibrium measured as squared correlation of allele frequencies ( $r^2$ ) against genetic map distance (cM) between linked loci pairs in the core collection. The red line is the nonlinear regression trend line of  $r^2$  vs. genetic map distance. The dashed line indicates the 95% quantile of  $r^2$  between unlinked loci pairs.

LD decay distance suggested that fine mapping with a core collection for desirable genes could be possible. However, due to low percentage of linked loci pairs in LD and the quick decay of LD, in turn it indicated that the density of markers for genome-wide association mapping should be greatly increased as compared to our study. Considered for the LD decay distance in the core collection and 1700cM map distance of rice genome, it might at least in theoretically require more than 1700 markers for a genome-wide association mapping with such a core collection. If higher power is needed, the number of required markers could be even more.

## 6.2 Association mapping

Mining the elite genes within rice germplasm is of importance to the improvement of cultivated rice. Therefore, genome-wide association mapping was applied with the rice core collection using 274 SSR markers.

All of the 150 rice varieties were cultivated at the farm of South China Agricultural University, Guangzhou (23°16'N, 113°8'E), during the late season (July–November) for two consecutive years (2008 and 2009). The yield related traits (such as grain weight, filled grains, tillers per plants, etc.) were measured for both years. As for an example, the trait yield per panicle (gram) was furthered used for association mapping.

The software STRUCTURE was applied to infer historical lineages that show clusters of similar genotypes and get the Q matrices (Pritchard et al., 2000). Kinship matrix (K) was calculated by software SPAGeDi (Hardy and Vekemans 2002). The quantile–quantile plots of estimated  $-\log_{10}(p)$  were displayed using the observed p values from marker-trait associations and the expected p values assuming that no associations happened between markers and any trait in the software SAS. Using the TASSEL software and the mixed linear regression model (MLM), association test was performed for the yield trait, incorporating K and Q matrices.



The QQ plot of observed vs. expected p values (Fig. 7) indicated that MLM model incorporating K and Q matrix was suitable for the association analysis for the yield trait. A total of 17 markers in 2008 and 15 markers in 2009 were detected to significantly ( $P < 0.05$ ) be associated with the yield traits (Table 8 and 9). 12 marker-phenotype associations were confirmed by previous researches (either using linkage mapping or association mapping) for 2008 year's results. And it was 7 for the 2009 year's results. Moreover, two marker-phenotype associations were located in the similar position (RM471 with RM218, PSM188 with RM235) for both years. The genetic variants explained by the markers varied between 3.49% and 24.86% in 2008, while it was between 3.02% and 13.87% in 2009. The genetic variants explained by the marker RM346 and PSM336 were more than 15%. It is worth to note that less common marker-phenotype associations were detected in both years, which indicated the yield trait might be easily influenced by the environment. Such problem might be overcome by using the yield data for multiple locations and years. The marker-phenotype associations could be further used in rice breeding by marker-assisted selection.

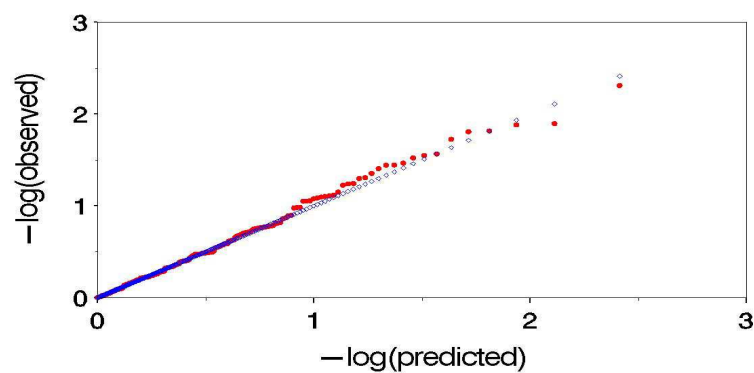


Fig. 7. Plots of observed vs. expected p values for MLM (Q+K) model for the yield trait.

No.	Markers	Chromosome	Map distance (cM)	P value	R <sup>2</sup> (%)
1	RM71	2	49.8	0.0377¥	3.49
2	PSM374	2	83.6	0.0448¥	5.05
3	RM293	3	193.4	0.0268	5.90
4	RM471	4	53.8	0.0202¥	6.39
5	RM559	4	155.8	0.0350	7.02
6	RM469	6	2.2	0.0150¥	10.33
7	RM549	6	42.7	0.0270	7.51
8	RM170	6	2.2	0.0346¥	7.05
9	RM432	7	43.5	0.0021¥§	12.26
10	RM429	7	96.9	0.0119¥	5.16
11	PSM142	7	26	0.0385¥	13.45
12	PSM336	7	80.5	0.0002¥§	16.43
13	RM346	7	47	0.0006¥§	24.86
14	RM201	9	81.2	0.0011§	21.34
15	RM167	11	20.3	0.0133	7.07
16	PSM366	11	89	0.0018¥§	14.44
17	PSM188	12	86.5	0.0059¥	10.34

§The Bonferroni threshold ( $< 0.0036$ ); ¥supported by previous researches; R<sup>2</sup> represents the genetic variants explained by the marker.

Table 8. Association mapping results for yield per panicle in 2008 using MLM models.

No.	Markers	Chromosome	Map distance (cM)	P value	R <sup>2</sup> (%)
1	RM220	1	28.4	0.0392¥	3.02
2	PSM369	1	170.4	0.0130¥	4.41
3	RM450	2	150.8	0.0270¥	5.17
4	RM218	3	67.8	0.0491	5.62
5	RM153	5	3	0.0341	4.83
6	RM334	5	141.8	0.0188	8.74
7	RM340	6	133.5	0.0152	13.87
8	RM182	7	61	0.0155¥	5.99
9	RM10	7	63.5	0.0358¥	3.14
10	RM257	9	66.1	0.0049	5.71
11	RM242	9	73.3	0.0360	3.13
12	RM304	10	73	0.0126	4.45
13	RM228	10	96.3	0.0282	3.43
14	RM309	12	74.5	0.0299¥	3.36
15	RM235	12	91.3	0.0442¥	13.00

§The Bonferroni threshold (< 0.0036); ¥supported by previous researches; R<sup>2</sup> represents the genetic variants explained by the marker.

Table 9. Association mapping results for yield per panicle in 2009 using MLM models.

7. Discussion and prospect

Rice feeds more than 50% of the world’s population and is one of the most important crops in the world. Rice germplasm resource is the primary material for rice breeding and makes a concrete contribution to global wealth creation and food security. Therefore, understanding its valuable genetic diversity and using it in rice genetic improvement is of importance for raising rice yield and the resistance to biotic and abiotic stress as well as improving rice quality to secure global food supplies.

To mine the wide genetic diversity in plant germplasm populations, identification of phenotypic traits might the first and an important step. Besides the agronomic traits, physiological traits, stress-related traits, quality traits, resistant to virus and pets traits, etc. should be furthered studied in details. Based on the full evaluation of phenotypes, a dynamic core collection could be constructed either on a specific target trait or on all the traits so that the core collection could retain as much as genetic diversity with the minimum accessions. The core collection could be furthered studied with high density markers as well as exact measurement of its phenotype.

Association mapping has become a promising approach to mine the elite genes within germplasm population compared to traditional linkage mapping. Association mapping based on a core collection would help to catch as more phenotypic variation as possible. Compared to a natural population or a breeding population with narrow genetic basis, the LD level in a core collection might be low due to its diverse origin. Therefore, more markers might be needed for genome-wide association mapping. However, due to the quick LD decay, fine mapping might be possible with a core collection. To perform a precisely association mapping, multiple replications either locations or years for the phenotypic

identification, exact measurement of the population structure and the kinship should be considered. Furthermore, though a rapid progress has been made in genotyping, a quick, automated, economic genotyping technology (such as SNP array) for a large number of germplasm resources are desirable for association mapping with germplasm resources population. How to effectively combine the linkage and association mapping in plants (such as the nested association mapping in maize, NAM) might be another question which should be concerned. Due to such associations could be further applied in rice breeding with molecular assisted selection, it provide a bright future to make use of the elite genes in the diverse germplasm resources.

A strategy is proposed for exploring and utilization of the wide genetic diversity in plant germplasm populations, i.e. firstly, evaluation of the genetic diversity for germplasm populations at phenotypic and genotypic level; secondly, constructing core collection to achieve the maximum diversity with minimum accessions; thirdly, combining linkage mapping and association mapping to map desire QTL in a large scale and with high resolution; fourthly, developing near isogenic lines to verify and fine map QTLs; finally, cloning desirable genes and make use of them in cultivated plant breeding.

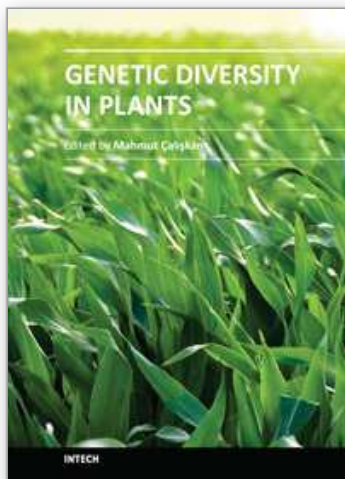
## 8. Acknowledgment

The authors thank very much to Academician of Chinese academy of science, professor Lu, Yong Gen for his supervision and support for the research. Many thanks are given to Dr. Li, Xiao Ling, Dr. Feng, Jiu Huan, Ms. Zhao, Xing Juan, Mr. Zeng Kai Long, Mr. Deng, Yong Hong, Dr. Xu, Hai Ming for their helps in the experiment and data analysis. The research was supported by Fund of the National Natural Science Foundation of China grant (30700494).

## 9. References

- Brown A H D. 1989a. Core collections: A practical approach to genetic resources management. *Genome*, 31: 818–824.
- Brown A H D. 1989b. The case for core collections. In: Brown A H D, Marshall D R, Frankel O H, Williams J T. *The Use of Plant Genetic Resources*. Cambridge: Cambridge University Press: 136–156.
- Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, Rimmer JB, Locke PA, Conneally PM, Schmechel KE, Small GW, Roses AD, Haines JL, Pericak- Vance MA (1994) Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet* 7:180–184
- Frankel O H, Brown A H D. 1984. Plant genetic resources today: A critical appraisal. In: Holden J H W, Williams J T. *Crop Genetic Resources: Conservation and Evaluation*. London, UK: George Allen and Unwin Ltd: 249–257.
- Goff SA, Ricke D, Lan TH, et al., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296: 92–100.
- Hardy, O. J., Vekemans, X. (2002) SPAGeDi: a versatile computer program to analysis spatial genetic structure at the individual or population levels. *Molecular ecology notes*, 2: 618–620

- Heuertz M, Emanuele D P, Kallman T, Larsson H, Jurman I, Morgante M, Lascoux M, & Gyllenstrand N (2006), Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce (*Picea abies* (L.) Karst). *Genetics* 174:2095 – 105.
- Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox DK, Chakravarti A, Buchwald M, Tsui LC (1989) Identification of the cystic fibrosis gene: genetic analysis. *Science* 245:1073–1080
- Kraakman ATW, Niks RE, Van den Berg PMMM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* 168:435–446
- Lu, Y. G., Li, X. L., Li, J. Q. (2006). Construction of a rice core collection. In: Seminar for the developing strategy of Chinese super rice and 10th annual memory. Shenyang Agricultural University, 124-133
- Pritchard JK, S. M., DONNELLY P 2000a Inference of population structure using multilocus genotype data. *Genetics*: 945-959.
- Tanksly, S.B., Mc Couch, S. R. (1997) Seed bank and molecular maps: unlocking genetic potential from the wild. *Science*, 277(22): 1063-1066
- Yan, J. B., M. Warburton and J. Crouch, 2011 Association Mapping for Enhancing Maize (*Zea mays* L.) Genetic Improvement. *Crop Science* 51: 433-449.
- Yu J, Hu S, Wang J, et al., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296: 79–92.



## **Genetic Diversity in Plants**

Edited by Prof. Mahmut Caliskan

ISBN 978-953-51-0185-7

Hard cover, 498 pages

**Publisher** InTech

**Published online** 14, March, 2012

**Published in print edition** March, 2012

Genetic diversity is of fundamental importance in the continuity of a species as it provides the necessary adaptation to the prevailing biotic and abiotic environmental conditions, and enables change in the genetic composition to cope with changes in the environment. Genetic Diversity in Plants presents chapters revealing the magnitude of genetic variation existing in plant populations. The increasing availability of PCR-based molecular markers allows the detailed analyses and evaluation of genetic diversity in plants and also, the detection of genes influencing economically important traits. The purpose of the book is to provide a glimpse into the dynamic process of genetic variation by presenting the thoughts of scientists who are engaged in the generation of new ideas and techniques employed for the assessment of genetic diversity, often from very different perspectives. The book should prove useful to students, researchers, and experts in the area of conservation biology, genetic diversity, and molecular biology.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Jin Quan Li and Peng Zhang (2012). Assessment and Utilization of the Genetic Diversity in Rice (*Oryza sativa* L.), Genetic Diversity in Plants, Prof. Mahmut Caliskan (Ed.), ISBN: 978-953-51-0185-7, InTech, Available from: <http://www.intechopen.com/books/genetic-diversity-in-plants/assessment-and-utilization-of-the-genetic-diversity-in-rice-oryza-sativa-l->

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821



© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen