SSR markers in sampling a core collection and estimating the genetic diversity

Hong-liang Zhang, Zichao Li, Dongling Zhang, Junli Sun, Meixing Wang, Yongwen Qi and Xiang-kun Wang

Key Lab of Crop Genomics and Genetic Improvement, Ministry of Agriculture and Beijing Key Lab of Crop Genetic Improvement, China Agricultural University, Beijing, 100094, P.R. China.

Author for correspondence, Email: lizichao@cau.edu.cn

Abstracts

The selection strategies of SSR markers in sampling a core collection and estimating genetic diversity were studied, using a population with 358 rice landraces from China, and 72 SSR loci. Combining different locus numbers and four types of locus with different polymorphisms got forty-four sets of locus compositions. Two hundred and thirty core collections were sampled randomly and by clustering using different locus compositions, in different proportions to the whole population. Analysis was made on: (i) the correlation coefficient between the genetic distance/similarity matrix in the whole population estimated by different locus combinations, (ii) the correlation coefficient between the genetic distance of SSR markers should be focused on for different research purposes and objectives. The effects on estimating the genetic differences and sampling a core collection in fact related to the locus number rather than simply to the allele number. To precisely estimate the genetic diversity and sample a representative core collection, no less than 18 and no more than 36 SSR loci randomly selected could be recommended.

Media summary

The selection strategy of SSRs loci was determined for establishing the genetic diversity in a rice population and establishing its core collection.

Key words

Core collection, genetic structure, SSR, locus selection

Introduction

Microsatellites (SSRs) has been applied to many studies, such as variety identification, genetic patterns, phylogenesis and so on. The loci number varied from 4 (Kikuchi1, 2002) to 83 (Lu H, 2001). And many scholars have been studying how to select the makers in different aspects for different purposes (William, 1997; Le, 2001). All studies suggested that no optimal number suitable for all the species can be proposed. Using 358 rice landraces from the core collection of rice landraces in China and 72 SSR loci, the relationship of the selection of SSR locus with estimation of the genetic diversity of the population, and with sampling a representative core collection were studied. Strategies of locus selection were determined to estimate rice genetic diversity and to sample a rice core collection.

Materials and methods

Materials

The whole population includes 358 accessions from the core collection of rice landraces in China, of which 212 accessions originate in south-China's Guangxi province and the others originate in four north-China provinces. According to the subspecies (indica vs. japonica), geography (south-China vs. north-China) and ecotypes (lowland rice vs. upland rice) to which the accessions belong, this population could be stratified into 6 subpopulations.

Locus combination and core collection sampling

SSR primers at 72 loci on 12 rice chromosomes were randomly selected. Forty-four sets of SSR locus combinations were formed by combining different numbers and types of locus (Table 1). At the interval of 6 loci, eleven levels of locus number were set from 6 to 66. Four types of locus were set according to different criteria, i.e. the highest polymorphic type, the lowest polymorphic type, the most representative type (or called core type) and the randomly selected type (hereafter symbolized as "h", "l", "c" and "r" respectively). Two hundred and twenty five core collections were sampled by means of clustering according to all 72 SSR loci and each set of SSR locus combination, in the proportion of 10%, 30%, 50%, 70% and 90% to the whole population respectively. Another five core collections were sampled randomly in the same proportions.

Table 1 Allele number in different locus combinations

		6	12	18	24	30	36	42	48	54	60	66	
locus type	с	61	141	206	275	342	410	501	598	696	794	886	
	h	175	296	403	495	577	649	718	780	830	872	909	
	I	19	56	98	148	210	279	351	433	525	632	753	
	r	75	189	213	309	334	450	537	634	728	755	851	
Mean		82.5	170.5	230.0	306.8	365.8	447.0	526.8	611.3	694.8	763.3	849.8	
SD		66.1	100.1	126.8	143.4	153.3	153.2	150.8	142.5	126.8	100.1	68.8	

locus number

Data analysis

Using the Mantel test, the correlation coefficient between: i) genetic similarity matrix among 358 accessions in the whole population estimated by 72 SSR loci and that by each set of locus combinations (symbolized as CR-AL), and ii) that between genetic distance matrix among 6 subpopulations in the whole population estimated by 72 SSR loci and that by each set of locus combinations (CR-GL), and iii) that between genetic distance matrix among 6 subpopulations (CR-GL), and iii) that between genetic distance matrix among 6 subpopulations estimated by 72 SSR loci in each core collection and that in the initial population (CR-GC) was calculated.

Two parameters of genetic diversity, i.e. the allele number and Nei's gene diversity index (Ha) at 72 SSR loci, were calculated in different core collections. The allelic retention (RT) in each core collection was calculated by the ratio between allele number in each core collection and that in the initial population.

Results

Genetic diversity and structure in the whole population

In the whole population, alleles detected at 72 SSR loci counted up to 928, averaged 12.9 per locus, and ranged from 2 to 33. The allele number and average gene diversity index in the whole population varied distinctly among different locus combinations (Table 1). The phylogenic tree was formed through UPGMA among six subpopulations in the whole population according to 72 SSR loci. This tree answered to the real genetic structure of rice germplasm in China.

It showed that CR-GL rapidly increased along with the locus number increasing from 6 to 24. After the locus number reached 24, along with the increase of locus number, CR-GL estimated by different locus combinations still increased, but very slowly. In this section of locus number, "I" locus combinations showed higher CR-GL than the other ones. However the differences between locus combinations were very small. It showed that, along with the locus number increasing, CR-AL took on similar trends as CR-GL. But to reach CR-AL equal to or higher than 0.9, only 12 loci was needed for "I" and "c" locus combination, 24 for the "r" one, but 36 for the "h" one.

Appropriate sampling proportion for this population

To sample a core collection with high quality, the appropriate sampling proportion should firstly be determined. By comparing CR-GC, the mean, maximum and minimum of gene diversity index and allelic retentions of all core collections in certain sampling proportions, it was found that in the core collection with a sampling proportion of 70%, the initial genetic structure could be held, at least 90% of the total alleles could be retained, and more genetic redundancy was removed than in the core collection with a sampling proportion of 90%. So it could be concluded that the appropriate sampling proportion is 70% for this population.

Genetic affinities and diversities in core collections with sampling proportion 70%

On the whole, CR-GC for the same type of locus combination increased and gradually stabilized to 1 along with the increase of locus number. But the increase and stabilization speed were different for different types of locus combination. The "h" locus combination could give stable CR-GC among different number of locus and higher than 0.999 after the locus number reached 12. CR-GC for "c" locus combinations increased stably along with the increase of locus number. But the difference was very small. CR-GC for "l" locus combinations fluctuated widely among different numbers of locus before the locus number reached 48, and then levelled out. CR-GC for "r" locus also took on evident fluctuation among different numbers of locus while the locus number was less than 30 but became similar to those for "h" locus combinations for higher locus numbers.

The gene diversity index in the core collection sampled according to all locus combinations increased rapidly along with the locus number increasing from 6 to 18. However it stopped increasing just below locus number 24 and became invariant after that. For the same number of locus that was more than 6, the core collection sampled according to "I" locus combination showed higher gene diversity indexes than those sampled according to other types of locus combinations. Those sampled according to "h" locus combination showed lower gene diversity indexes than those sampled according to other types of locus combinations. But the differences of gene diversity index among locus combinations after the locus number more than 12 were very small.

On the whole, the allelic retention increased along with the increase of locus number, and its change could be classified into three phases, i.e. the rapid increase phase, the wide fluctuation phase and the smooth increase phase. All types of locus went through the rapid increase phase from 6 to 18 loci. The wide fluctuation phase began with 24 loci. After 30 loci, the "I", "h" and "c" locus combination all started smoothly increasing along with the increase of locus number. None of four types of locus combinations showed a higher allelic retention than other types under all locus numbers.

Discussion

Determination on the marker polymorphism should depend on the research purpose and objects.

In this study the genetic affinity among individuals and subpopulations in the whole population could be better estimated by "I" locus combinations than by "h" locus combinations. It is apparent that negative correlation exists, to some extent, between the maintenance of original genetic structure and the removal of genetic redundancy in sampling a core collection. The "h" locus combination could achieve a core collection with more representative genetic structure in the population. But more genetic redundancy could be removed by sampling a core collection according to "I" locus combination than that according to "h" locus combinations. The core and randomly selected locus combinations behaved between the lowest and highest polymorphic locus combinations and similar to the former. Appropriate number of randomly selected locus could be enough to meet the above tasks. For balancing the time and costs of examining large numbers of markers with less magnitude of getting better results, both core and highest polymorphic locus compliane.

Number of locus, rather than the number of alleles in studying the genetic diversity or sampling a core collection.

Fanizza (1999) and Kalinowski (2002) thought that equivalent results could be achieved by using either a few loci with many alleles or many loci with a few alleles and preferred to use the number of alleles as the selected object. We do not agree with them. Firstly, we could not predict the number of alleles for each locus before detecting it. Secondly, these conclusions were based on the assumption that the alleles examined are independent each other. But associations between alleles within or between loci have been found existing widely in some organisms such as plants (Li YC, 2000). This association will make the selection of allele number no an effective criteria of marker selection for estimating genetic similarity and structure in a population. And the number of loci should be used for this purpose.

This study showed that the same results could be achieved with the same number of loci but different number of allele, and even better results could be achieved by the same number of loci with less number of alleles. As the locus number increased from 6 to 18, CR-GC, CR-GL, CR-AL, allele number and gene diversity index changed distinctly, but similar values could be gotten for these parameters measured by the same number of locus higher than 12, at which different numbers of alleles were detected for different type of locus. Thus it is obvious that these parameters are essentially correlated with the number of locus, rather than the number of alleles.

In summary, different aspects in selection of markers should be focused on for different research purposes and objectives. To precisely estimate the genetic diversity in the population and sample a representative core collection, the randomly selected locus combination with locus number no less than 18 could be recommended. Though more stable and better estimation and sampling could be achieved by using more loci, and a locus number than 36 is not recommended because the only a small gain is achieved for a large additional cost.

Acknowledgements:

This study is supported in part by Project of Cooperation between Yunnan Province and China Agricultural University (Y98ZEN07) and National Project of Important Fundamental Research ("973" Project, G1998010201)

References

Fanizza G, G Colonna, P Resta, G Ferrara. 1999. The effect of the number of RAPD markers on the evaluation of genotypic distances in *Vitis vinifera*. Euphytica, 107: 45-50

Kikuchi1 S, Y Isagi. 2002. Microsatellite genetic variation in small and isolated populations of *Magnolia* sieboldii ssp. japonica. Heredity, 88, 313-321

Le Clerc, V., Briard, M. and Peltier, D. 2001. Evaluation of Carrot Genetic Substructure: Comparison of the Efficiency of Mapped Molecular Markers with Randomly Chosen Markers. Acta Hort (ISHS), 546: 127-134, http://www.actahort.org/ books/546/546_10.htm

Li YC, MS R?der, T Fahima, VM Kirzhner, A Beiles, AB Korol, E Nevo. 2000. Natural selection causing microsatellite divergence in wild emmer wheat at the ecologically variable microsite at Ammiad, Israel. Theor Appl Genet, 100: 985-999

Lu H, R. Bernardo. 2001. Molecular marker diversity among current and historical maize inbreds. Theoretical and Applied Genetics, 103: 613-617

Nei M. 1973. Analysis of gene diversity in subdivided populations. P Nat Acad Sci USA, 70: 3321-3323

William J Boecklen, Daniel J Howard. 1997. Genetic analysis of hybrid zones: Numbers of markers and power of resolution. Ecology, 78 (8): 2611-2616