African Journal of Biotechnology Vol. 8 (13), pp. 2963-2967, 6 July, 2009
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2009 Academic Journals

Full Length Research Paper

Rice genetic marker database: An identification of single nucleotide polymorphism (SNP) and quantitative trait loci (QTL) markers

ChangKug Kim[1], UngHan Yoon[1], GangSeob Lee[1], HwanKi Lee[2], YongHwan Kim[1] and JangHo Hahn[1]*

[1]Genomics Division, National Academy of Agricultural Science (NAAS), Suwon 441-707, Korea.
[2]Technology Information Division, Rural Development Administration (RDA), Korea.

Accepted 21 April, 2009

The National Academy of Agricultural Science (NAAS) has developed a web-based genetic marker system to provide information about SNP and QTL markers in rice. The SNP marker database provides 7,227 SNP markers including location information on chromosomes by using genetic map. It allows users to access a detailed characterization table of 12,829 potential SNPs in 3,356 genes. The QTL marker database provides 175 QTL markers information with 942 polymorphic markers on each of the 12 chromosomes in rice. Users are assisted in tracing any new structures of the chromosomes and gene positional functions through comparisons using specific SNP and QTL markers.

Key words: Genetic map, rice database, SNP marker, QTL marker.

INTRODUCTION

Rice (*Oryza sativa*) is one of the most important crops in Korea and in the Northeast Asia. To reveal the genomic relationship among genetic markers, genetic mapping has become a necessary step. A genetic map is a linear representation of the order of genetic markers. Constructing a genetic map with various markers positioned along a chromosome requires a mapping function for inferring map distances between genetic markers. In rice, several genetic maps based on various genetic markers have been constructed for the purpose of displaying positional relationships among mapped markers in each chromosome (Katsumi et al., 2000).

A Single Nucleotide Polymorphism (SNP) is a DNA sequence variation occurring when a single nucleotide in the genome differs between members of a species. SNP, which is molecularly defined, is widely used as a genetic marker in mapping experiments and Quantitative Trait Loci (QTL) analyses. QTL marker databases for complex traits have become a routine tool in functional genomic research. QTL mapping is widely used due to its simplicity and concept. QTL detection results are now available in various databases. To increase the importance of SNP and QTL markers, various databases have been constructed. The OryzaSNP (http://www.ory zasnp.org) has generated a large fraction of the SNPs present in cultivated rice through 20 rice genomes. The SNP Database Network in Japan (http://snpnet.jst.go.jp/) provides an integrated menu for SNP searching. The dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/) provides a function for SNP discovery across the entire plant genome. The BioMercator has been developed to automate map compilation and QTL meta-analysis, and to visualize colocations between genes and QTL through a graphical interface (Arcade et al., 2004). QTLNetwork is a software package for mapping the genetic architecture underlying complex traits for experimental populations derived from a cross between two inbred lines (Yang et al., 2008). The Gramene QTL database (http://www. gramene.org/qtl/) provides QTL identified for numerous agronomic traits in rice, maize, barley and other crops.

The National Academy of Agricultural Science (NAAS, http://www.niast.go.kr/), on the other hand, has developed a web-based database system to provide information about SNP and QTL markers in rice. The user is assisted in tracing any new structures of the chromosomes and gene positional functions through compari-

*Corresponding author. E-mail: jhhahn@rda.go.kr.

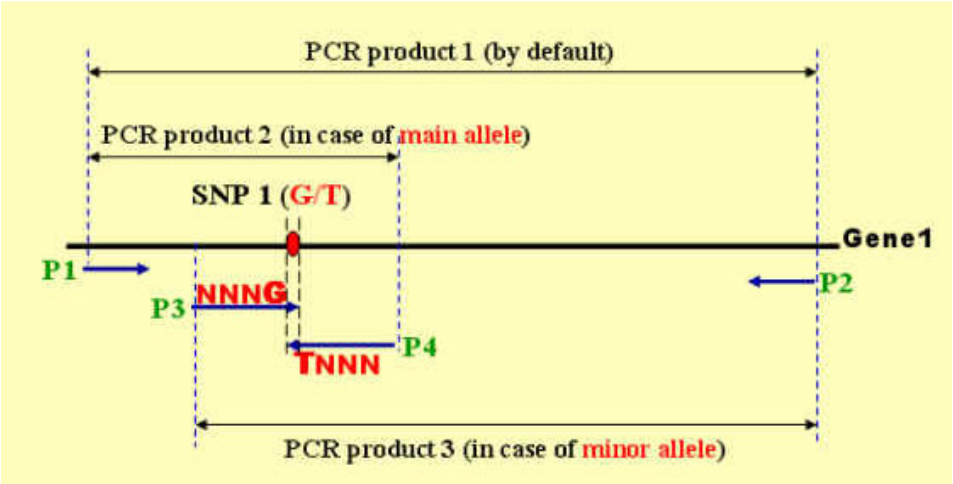

Figure 1. Polymerase Chain Reaction (PCR) primer design for SNP detection.

sons using specific SNP and QTL markers.

METHODOLOGY

Dataset

The SNP markers were collected from the SNP project (*Oryza sativa, Nipponbare: Japonica and China93-11: Indica*) of NAAS. The QTL markers, on the other hand, were collected from the Korea rice genome project (*Oryza sativa, Tongil: Indica/Japonica*) of NAAS and Seoul National University (*Oryza sativa, Seolak: Japonica*). QTL information (*Oryza sativa, Tongil*) was added to the experimental analysis including all released graph images. In order to develop a genetic map, additional genetic markers were collected from GRAMENE (ftp://ftp.gramene.org/pub/gramene/release16/data/).

Design for SNP markers

In order to construct a database of SNP markers, we analyzed the variation in gene sequences between japonica and indica rice using EST sequence data from the rice genome database and molecular markers. We analyzed the nucleotide sequences in the rice chromosome at an interval of 100 kb. Primers were designed for amplification and were utilized for the detection of SNPs. Each SNP can be diagnosed by allele-specific PCR (Polymerase Chain Reaction) primer provided from the localization of the gene (RSN) marker database (RSNMDB, http://nabic.niab.go.kr/SNP/). Figure 1 shows a design of PCR primer for SNP detection. The primer set (P1, P2, P3 and P4) is used in PCR reaction for SNP detection. The P1-2 primers are involved in PCR reaction providing a basis for deciding the success or failure of the reaction. The main allele has two PCR products: product 1 and 2. If the sequence of SNP position is 'G', P3 primer is aligned at SNP position on the genome sequence as main allele. If the sequence of SNP position is 'T', P4 is aligned at SNP position as minor allele. Therefore, the PCR products of minor allele are PCR product1 and PCR product 3.

Design for QTL markers

To construct a QTL markers database with genetic map, we analyzed 175 QTL markers using the Haldane mapping function method (Haldane, 1919). Five processes were followed in developing the map of QTL markers; (1) The recombination frequency was computed using maximum likelihood algorithm. (2) The LOD (Logarithm of the Odds) score was computed. The LOD score was computed using a range of recombination frequency values between markers (from 0 to 0.5). The LOD score was selected as the maximum value of the distance between markers. Therefore, the QTL position was decided by the maximum LOD value. LOD score less than 2 was excluded. Some markers had to be removed because the LOD threshold for a genome-wide error was beyond the 0.05 significance level. (3) The LOD score was converted. If θ corresponds to the recombination frequency between a pair of markers, then the Haldane mapping function is defined by

$$\theta = \frac{1-e^{-2m}}{2} \qquad m = -\frac{1}{2}\ln(1-2\theta)$$

where Θ is the recombination frequency; m is the map distance in centiMorgan (cM). (4) The map distance was calculated: In step 3, the value of 'm' is the distance in cM. With 'θ' as the recombination frequency, the map distance can be calculated. (5) A position with QTL marker length was derived: The red line length at LOD = 2.0 means inter-vals for the position of the QTL. The LOD score is on the Y-axis while the relative location of the marker in cM is on the X-axis (Figure 2).

Database development

Using the collected markers, a genetic marker database was constructed to distinguish SNP and QTL markers. The database platform was developed using MYSQL and Java languages. The data was stored in Oracle relational database management system (RDBMS). The logical and physical schema of the database followed the standard principles of relational database by ERWin Data Modeler software (http://www.ca.com). The genetic map construction utilized MapMaker/Exp program and the LOD values of the QTL markers were analyzed with MapMaker/QTL program (Lincoln et al., 1992). The genetic map in the QTL markers database was developed using the INE (INtegrated rice genome Explorer) method (Katsumi et al., 2000). The QTL map was developed in two stages

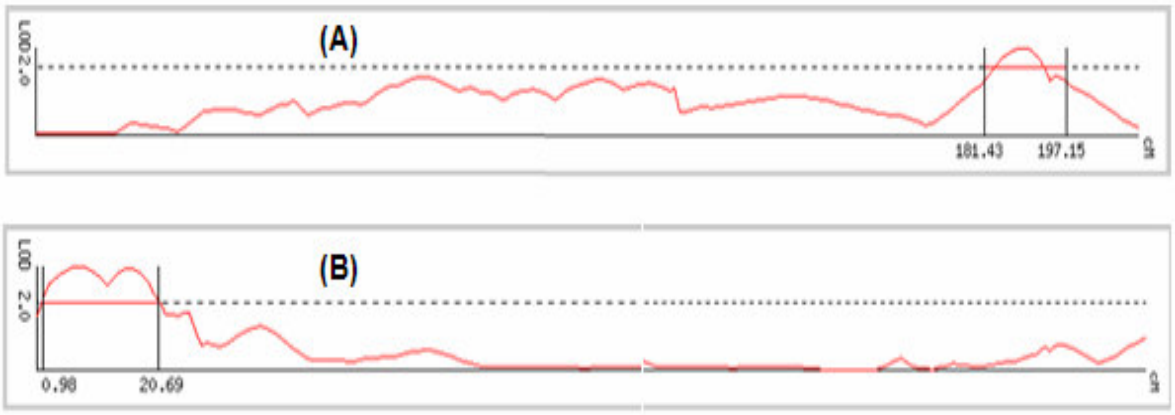

Figure 2. LOD score for a hypothetical quantitative trait. The dotted line at LOD=2.0 indicates the required significance level. Image (A) shows sample (Peak LOD: 2.57 Name: qAMY-8). Image (B) shows sample (Peak LOD: 3.06 Name: qYD-9) at the ninth chromosome.

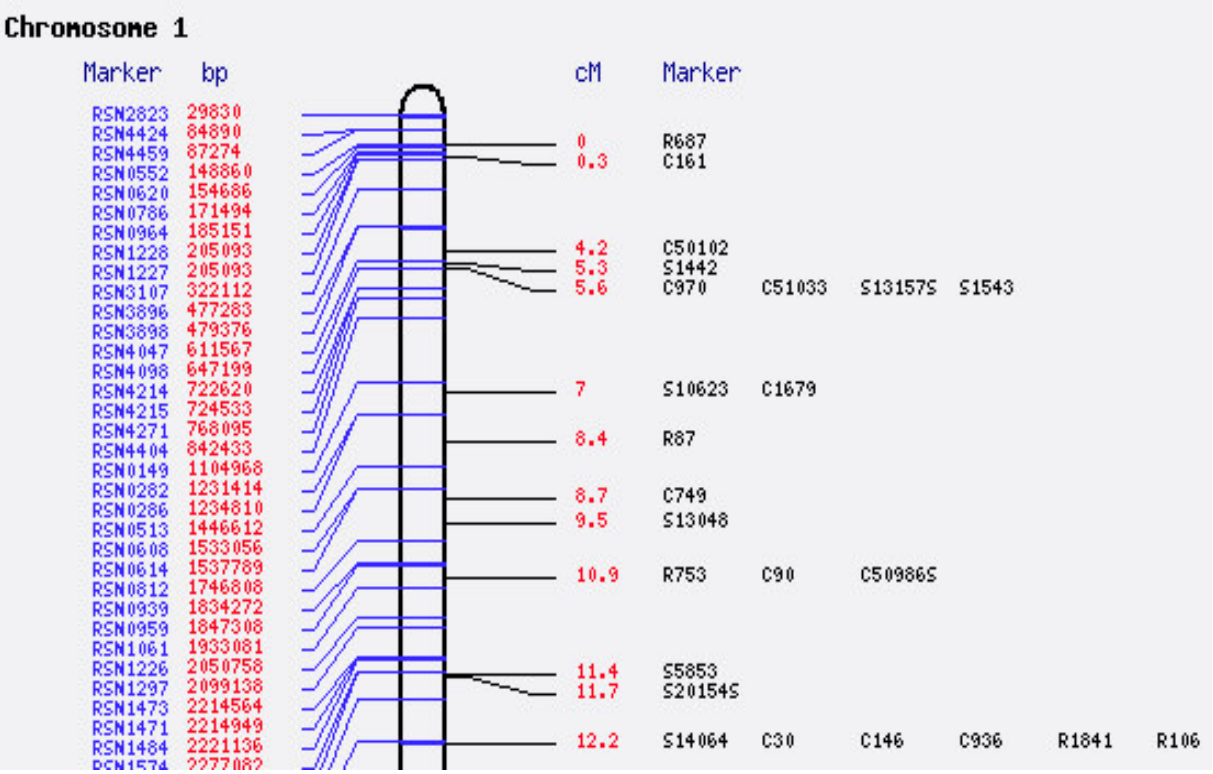

Figure 3. A snap shot of a genetic map in the SNP marker database. Each figure represents an accumulated distance in the chromosome.

as polymorphic markers and QTL markers. The QTL marker database provides information on 175 QTL markers with 942 polymorphic markers on each of the12 chromosomes in rice.

Implementation and features

The SNP marker database (http://nabic.niab.go.kr/SNP/) features three functions: map viewing, RSN marker searching and marker annotation. The SNP marker database provides 7,227 SNP markers including location information on chromosome using a genetic map (Figure 3). Users can access a detailed characterization table of 12,829 SNP markers. The 12,829 SNP markers were detected from 22,423 non-redundant genes in 3,356 genes by pairwise alignment between ESTs and genomic sequences of rice. When using keyword search or inputting identification number, users can view a table of detailed information. A specific table provides information such as marker name, region, gene definition,

Marker : RSN0002

1st Marker

Marker	Region	Gene Definition	Product				Primer			Include SNP
			Position		Size	GC(%)	Sequence		TM	
RSN0002	3'-UTR	Phosphatidylinositol 3-and 4-kinase family-like [Oryza sativa (japonica cultivar-group)]	Start	10013116	657	42.9	Forward (P1)	GCGATAGCAGTTGTTGCTCA	60.17	OSS07105 : G->T OSS07106 : G->T
			End	10013772			Reverse (P2)	AAATGCTGAAAGCACAACCA	59.32	

2nd Marker

Marker	snpID	Position	SNP desc	Primer		Detect Method	Remark (Primer interaction)	PCR Product size		
RSN0002.01	OSS07106	10013256	G->T	Forward (P3)	GGTCAGAAAAACTCGTGCTT	Sequencing	P3_vs_P1; P4_vs_P4	Main	P2-P3	535bp
				Reverse (P4)	TAGATGCATTTCTTAGCAAGAT			Minor	P1-P4	162bp
RSN0002.02	OSS07105	10013209	G->T	Forward (P3)	CCCCAAATCCAAGTTATCTTTAGT	Sequencing	P3_vs_P2; P4_vs_P2	Main	P2-P3	586bp
				Reverse (P4)	ACAGCGTTTTGCTCAAACAT			Minor	P1-P4	113bp

Figure 4. A snap shot of detail information table in the SNP marker database. A detail table provides a characterization of SNP marker to distinguish the two fields.

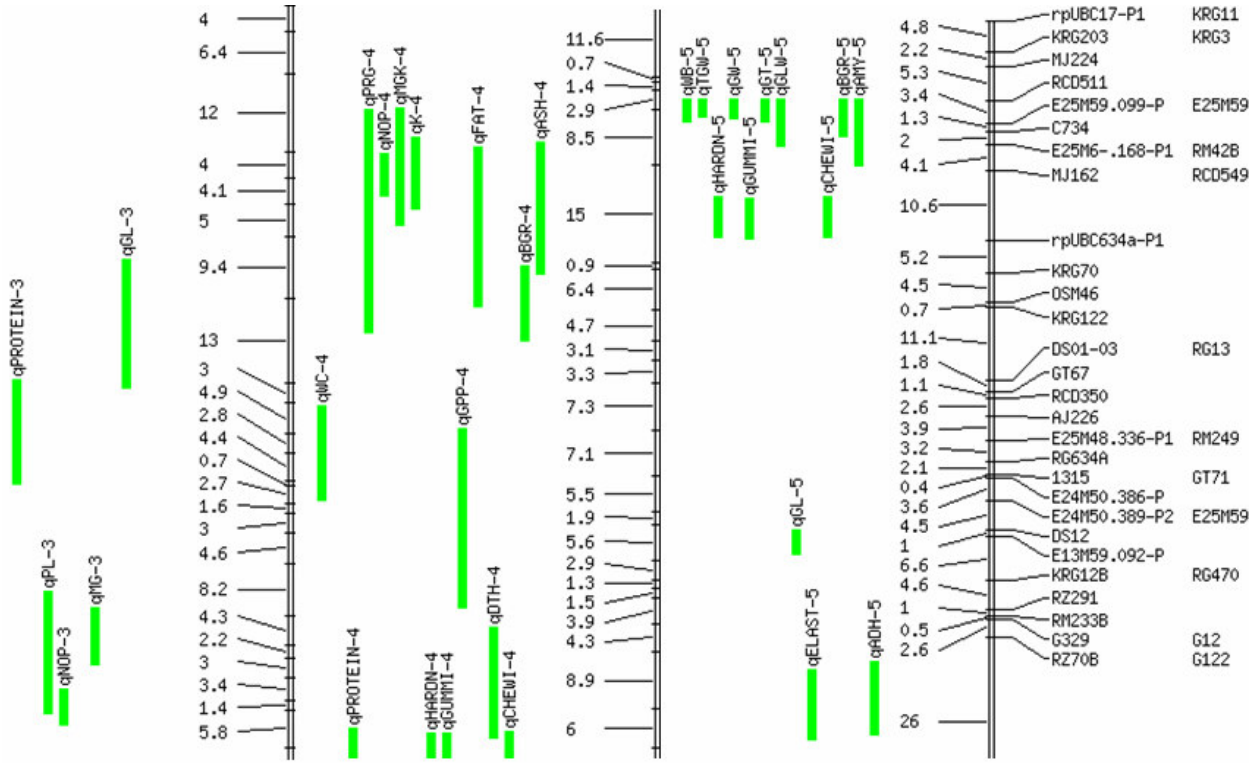

Figure 5. A snap shot of genetic linkage map. The view page shows the genetic map of 3, 4 and 5 chromosome in rice. The QTL marker locus is displayed on the left while the genetic markers as polymorphic marker are displayed on the right. The QTL names are shown next to each vertical green bar. Each figure represents a distance in the chromosome. In this figure, the images for 3 and 4 chromosomes have been omitted.

product, primer, etc (Figure 4).

The QTL marker database (http://nabic.niab.go.kr/nabic/geneticMap_Plant.php) provides information through a genetic map of about 175 QTL markers with 942 polymorphic markers on each of the 12 chromosomes in rice. The database has three major parts and functions such as map search, marker search and QTL search. Users can detect a mapped locus of specific QTL markers with associated trait. In addition, users can view a table of detailed information by clicking the linked QTL marker (Figure 5).

DISCUSSION AND FUTURE WORK

The SNP marker database provides 7,227 SNP markers

including location information on chromosome by using genetic map. Users can access a detailed characterization table of 12,829 SNP markers. The 12,829 SNP markers were detected from 22,423 non-redundant genes in 3,356 genes by pair-wise alignment between ESTs and genomic sequences of rice. The QTL marker database provides information through genetic map of about 175 QTL markers with 942 polymorphic markers on each of the 12 chromosomes in rice. The database has three major parts and functions such as map search, marker search and QTL search. In the future, NAAS plans to develop an integrated genetic linkage map that combines an SNP marker database and a QTL marker database for rice.

REFERENCES

Arcade A, Labourdette A, Falque M, Mangin B, Chardon F, Charcosset A and Joets J (2004) BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. Bioinformatics 20(14):2324-2326

Haldane JBS (1919). The combination of linkage values, and the calculation of distances between the loci of linked factors. J. Genet. 8: 299-309.

Katsumi S, Antonio BA, Mukai Y, Nagasaki H, Sakai Y, Makino K Sasaki T (2000) INE: a rice genome database with an integrated map view. Nucleic Acids Res. 28: 97-101.

Lincoln S, Daly M and Lander E (1992) Construction Genetic Maps with MAPMAKER/EXP 3.0. Whitehead Institute Technical Report, 3rd edition. Whitehead Insitute, Cambrige, MA.

Yang J, Hu C, Hu H, Yu R, Xia Z, Ye X and Zhu J (2008) QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. Bioinformatics 24(5):721-723