INT. J. BIOL. BIOTECH., 7 (1-2): 39-42, 2010.

THE CANDIDATE GENE APPROACH IN PLANT ANAEROBIOSIS

S. Abdolhamid Angaji

Department of Biology, Tarbiat Moallem University, Tehran, Iran
Email address: angaji@tmu.ac.ir, ershad110@yahoo.com

ABSTRACT

There has been a tenfold increase in the number of Quantitative Trait Loci (QTL) studies published annually over the past 10 years. Once QTLs have been identified; the next challenge is to identify the genes. Identifying the genes behind these QTLs has been described as the greatest challenge for geneticists in this century. Three main approaches lead to the cloning of genes of interest. Classical methods such as positional cloning and insertional mutagenesis have been used with success to identify major genes, i.e. genes having a major effect on the phenotype. However, these methods are limited by genome size and/or by the lack of transposons in the species being studied. For those species that have been sequenced, there should be no need to generate the large insert clones because gene order is already known. All that is required is to locate a QTL on the sequence and then look for candidate genes (CG). In the present study, 36 CG were detected for seed germination under anaerobic condition in rice.

Keywords: Candidate gene, Anaerobic germination, QTL mapping

INTRODUCTION

There has been a tenfold increase in the number of studies published on quantitative trait loci (QTL) annually over the past 10 years. Once QTLs have been identified, the next challenge is to identify the genes (Salvi, and Tuberos, 2005). Identifying the genes behind these QTLs has been described as the greatest challenge for geneticists this century (Luo *et al.*, 2002). Three main approaches lead to the cloning of genes of interest. Classical methods such as positional cloning (Rommens *et al*., 1989; Xu *et al*., 2006) and insertional mutagenesis (Bechtold *et al*., 1993) have been used with success to identify major genes, i.e. genes having a major effect on the phenotype. However, these methods are limited by genome size and/or by the lack of transposons in the species being studied (Byrne, and Mcmullen, 1996).

The definition of a CG differs by discipline. Physiologists consider CG as all genes involved in the expression of a given trait, whereas geneticists consider as only polymorphic genes putatively involved in the trait variation. Thus, for geneticists, CGs differ based on the crosses being studied. For plant geneticists, a CG is any gene putatively involved in trait variation, based on its biological function and/or its map localization. In this study, the aim is to adapt the definition of plant geneticists: CGs are either genes with molecular polymorphisms genetically linked to major loci or QTLs, or genes with molecular polymorphisms statistically associated with variation of the trait being studied (Rothschild, and Soller, 1997). The CGs may be structural genes or genes involved in the regulation of a metabolic pathway. The working hypothesis assumes that a molecular polymorphism within the CG is related to phenotypic variation. The CG approach has been used with success in human and animal genetics (Price, 2006) and, since 1990s, in plant genetics (Byrneand Mcmullen, 1996; Angaji, 2009; Fray *et al.*, 2000; Rozmahel *et al*., 1989).

The CG approach consists of three chronological steps. First, CGs are proposed based on molecular and physiological studies (functional CGs) or based on linkage data of the locus being characterized (all closely linked genes may be positional CGs). Second, a molecular polymorphism must be revealed to localize the CGs on a genetic linkage map to look for genetic linkage between the CG markers and the loci being characterized, or to calculate statistical correlations between CG polymorphisms and phenotypic variation in a set of genealogically unrelated individuals. It is important to notice that these two strategies are fundamentally identical and can be conducted together or successively. Third, if map co-segregation and/or statistical correlation have been found, complementary experiments must be conducted to confirm the actual involvement of the CG in the trait variation. This is the validation step (Pflieger *et al*., 2001; Fridman *et al*., 2000).

For those species that have been sequenced, such as Rice and Arabidopsis, because gene order is already known, all that is required is to locate a QTL on the sequence and then look for CG (Price, 2006). Direct seeding of rice is increasingly being practiced in both rainfed and irrigated areas because of labor shortage for transplanting and opportunities for crop intensification. However, poor crop establishment remains a major obstacle facing its large-scale adoption in areas prone to flooding. To understand the genetic bases of tolerance to anaerobic conditions during germination and to identify relevant QTL(s) in rice, Khaiyan (*Aus* type), tolerant variety, was crossed with IR64, a semi-dwarf, modern variety moderately sensitive to the trait. Four putative QTLs were detected in BC2F2 population, one each on chromosomes 1 (*qAG-1*), 2 (*qAG-2-1*), 11 (*qAG-11*), and 12 (*qAG-12*). The LOD value of

these QTLs ranged from 3.66 to 5.71 with phenotypic variance in the range of 12 to 29.24%. Total phenotypic variation explained by the four QTLs was about 51.4% (Angaji, 2008).

In other study, Khao Hlan On (*Japonica* type), tolerant variety, was crossed with IR64, a semi-dwarf, modern variety moderately sensitive to the trait. Six QTLs were detected for tolerance of flooding during germination, located on chromosomes 1 (*qAG*-1-1 and *qAG*-1-2), 2 (*qAG*-2), 7 (*qAG*-7-2), and 9 (*qAG*- 9-1 and *qAG*-9-2). The contributions of the QTLs to the phenotypic variation ranged from 17.9 to 33.5%, and their LOD scores ranged from 5.69 to 20.34. The additive effects of all these QTLs were negative, suggesting that Khao Hlan On alleles contributed to increased tolerance of flooding conditions during germination (Angaji, 2008; Angaji *et al*., 2009). For the current study, the candidate gene approach was applied on the results of previous studies to identify genes involved in anaerobic germination.

MATERIALS AND METHODS

By means of a Contig view icon in the same site, TIGR Gene IDs of nearby putative candidate genes were obtained and the IDs were searched out in Rice Annotation (release 5) available in the TIGR Rice Genome Annotation (http://www.tigr.org/tdb/e2k1/osa1/) to find out the function of respective candidate genes. PubMed (http://www.ncbi.nlm.nih.gov/sites/entrez?db=PubMed) and Basic Query in rice mpss (http://mpss.udel.edu/rice/) were sought to investigate more on the function of candidate genes.

RESULTS

A large number of putative candidate genes were annotated in the intervals within which QTLs associated with tolerance to flooding during germination were detected. The putative candidate genes were short-listed based on their relevance to the intermediates and products of alcoholic fermentation and glycolysis as well as other enzymes and pathways known to be associated with anaerobic metabolism. The short list is presented in Table1 and Table 2.

Table 1. Putative candidate genes annotated near the QTL regions associated with tolerance to anaerobic conditions during germination for Khaiyan/IR64 population.

GENE FUNCTION	TIGR LOCUS ID	CHR.
Glycosyl transferase family 1 protein	LOC_Os01g46430	1
CTP synthase	LOC_Os01g46570	1
Acyltransferase family protein	LOC_Os01g19390	1
Patatin	LOC_Os01g55650	1
dnaK protein	LOC_Os01g08560	1
ABC transporter family protein	LOC_Os02g32690	2
Acetyl-coenzyme A synthetase	LOC_Os02g32490	2
Glutathione S-transferase, N-terminal domain containing protein	LOC_Os11g37730	11
ABC transporter	LOC_Os11g37700	11
Glycosyltransferase	LOC_Os11g36700	11
RING-H2 finger protein ATL2I	LOC_Os12g42540	12
Cysteine synthase	LOC_Os12g42980	12
Auxin responsive protein	LOC_Os12g43110	12
Actin-depolymerizing factor	LOC_Os12g43340	12
P21 protein	LOC_Os12g43450	12
Alpha-amylase/trypsin inhibitor	LOC_Os12g43490	12
Amine oxidase	LOC_Os12g43590	12

DISCUSSION

To establish marker- aided selection (MAS) to incorporate tolerance of anaerobic conditions during germination

in rice breeding programs, it is necessary to find additional and more diagnostic DNA markers that are tightly linked with the QTLs of interest. The annotation was done in the regions in near the identified QTLs and short-listed the genes that are more likely associated with tolerance to hypoxia or anoxia based on published functional evidences (Table 1 and Table2). This list will form the entry point for future studies to further shorten the list after fine-mapping of these QTLs and subsequently validate the functional roles of the genes involved. The isolation of the genes at the QTLs for tolerance to anaerobic conditions will help design more precise gene-specific diagnostic markers for marker- assisted backcrossing (MAB) and also advance our understanding of the genetic and physiological mechanisms of tolerance. Besides, this can also facilitate searching for new and stronger alleles for tolerance.

Table 2. Putative candidate genes annotated near the QTL regions associated with tolerance to anaerobic conditions during germination for Khao Hlan On /IR64 population.

GENE FUNCTION	TIGR LOCUS ID	CHR.
Glycosyl transferase family 1 protein	LOC_Os01g46430	1
CTP synthase	LOC_Os01g46570	1
Acyltransferase family protein	LOC_Os01g19390	1
Patatin	LOC_Os01g55650	1
dnaK protein	LOC_Os01g08560	1
ABC transporter family protein	LOC_Os02g32690	2
Acetyl-coenzyme A synthetase	LOC_Os02g32490	2
Oxidoreductase	LOC_Os07g43370	7
Zinc finger family protein	LOC_Os07g43380	7
WRKY DNA binding domain containing protein	LOC_Os07g40570	7
Eukaryotic translation initiation factor 5A	LOC_Os07g40580	7
Strictosidine synthase family protein	LOC_Os09g20684	9
F-box protein interaction domain containing protein	LOC_Os09g20650	9
Zinc finger protein	LOC_Os09g20550	9
Pyruvate decarboxylase isozyme 3	LOC_Os09g20530	9
Major facilitator superfamily protein	LOC_Os09g20510	9
Succinate dehydrogenase and fumarate reductase iron-sulfur protein	LOC_Os09g20440	9
SCO1/SenC family protein	LOC_Os09g20430	9
Annexin family protein	LOC_Os09g20330	9
heavy-metal-associated domain-containing protein	LOC_Os09g20000	9
aminopeptidase	LOC_Os09g19820	9
Isochorismate synthases family protein	LOC_Os09g19734	9
Protein kinase domain containing protein	LOC_Os09g19350	9
Oxidoreductase, 2OG-Fe oxygenase family protein	LOC_Os09g18520	9
Flavonol synthase/flavanone 3-hydroxylase	LOC_Os09g18390	9
Histone acetyl transferase	LOC_Os09g17850	9
Oxidoreductase, short chain dehydrogenase/reductase family protein	LOC_Os09g17750	9

Conclusion

With the availability of large genome databases, it is now possible to predict putative function of a gene based on sequence information, thus enabling the identification of candidate genes involved in a particular biochemical pathway. A gene with a specific relevant function may be a good candidate if it maps to a region where a QTL has been identified. In the present study, 36 candidate genes were annotated with their TIGR locus number from rice genome annotation database of TIGR. The candidate genes should further be validated to identify the causal genes underlying these QTLs and to help develop more precise gene-specific markers for MAS breeding.

REFERENCE

Angaji, S. A. (2008). Mapping QTLs for submergence tolerance during germination in rice, *African Journal of Biotechnology*. 7 (15): 2551-2558.

Angaji, S. A. (2009). QTL mapping: A few critical points. *International Journal of Applied Research in Natural Products*, 2(2): 1-3.

Angaji, S.A., E. M. Septiningsih, D. J. Mackill and A.M. Ismail (2009). QTLs associated with tolerance of flooding during germination in rice (*Oryza sativa* L.). *Euphytica*, 14-15.

Bechtold, N., J. Ellis and G. Pelletier (1993). In planta *Agrobacterium* mediated gene transfer by infilteration of adult *Arabidopsis thaliana* plants. *Acad. Sci.* Paris, 3: 1194-1199.

Byrne, P. F. and M.D. Mcmullen (1996). Defining genes for agricultural traits: QTL analysis and the candidate gene approach. *Probe*, 7: 24-27.

Fray, A., T.C. Nesbitt, S. Grandillo, E. Knaap, B. Cong, J. Liu, J. Meller, R. Elber, K.B. Alpert and S.D. Tanksley (2000). *fw2.2:* a quantitative trait locus key to the evolution of tomato fruit size. *Science*, 289: 85-88.

Fridman, E., T. Pleban and D. Zamir (2000). A recombinant hotspot delimits a wild QTL for tomato sugar content to 484 bp within an invertase gene. *Proc. Natl. Acad. Sci*. USA, 97: 4718-4723.

Luo, Z. W., C. Wu and M.J. Kearsey (2002). Precision and high-resolution mapping of quantitative trait loci by use of recurrent selection backcross or intercross schemes. *Genetics*, 161: 915–929.

Pflieger, S., V. Lefebvre and M. Causse (2001). The candidate gene approach in plant Genetics: a review. *Mol. Breed*., 7: 275-291.

Price, A.H. (2006). Believe it or not, QTLs are accurate! *Trends in Plant Science*. 11(5): 213-216.

Rommens, J. M., M.S. Jannuzzi, B. S. Kerem, M.L. Drumm, G. Melmer, M. Dean, R. Rozmahel, J.L. Cale, D. Kenedy, N. Hideka, M. Zsiga, M. Buchwald, J.R. Riordan, L.C. Tsui and F.S. Cellins (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping- I, 245: 1059-1065.

Rothschild, M. F. and M. Soller (1997). Candidate gene analysis to detect genes controlling traits of economic importance in domestic livestock. *Probe*, 8: 13-20.

Salvi, S. and R. Tuberos (2005). To clone or not to clone plant QTLs: present and future challenges. *Trends Plant Sci*., 10: 297–304.

Xu, J. R., X. Zhao and R.A. Dean (2006). From genes to genomes; a new paradigm for studying fungal pathogenesis in *Magnaporthe oryzae* L. *Advances in genetics*, 57: 175-218.

(*Accepted for publication December, 2009*)