

Bioinformatics approach through SNPs for drought tolerance in rice

PRASHANT VIKRAM, ALOK KUMAR SINGH AND S.P. SINGH

Accepted : December, 2009

SUMMARY

Rice is global staple food crop grown mainly in Asia. Drought hampers yield of rice drastically in rainfed areas (Khush 1997). Sequence of whole rice genome is available in database and Oryza SNPs (single nucleotide polymorphism) are already available for 20 varieties (oryzasnp.org). A small region on distal arm of chromosome 1 was annotated using the SNP (single nucleotide polymorphism) data to work out putative candidate genes. Interestingly there were some SNPs inside exons including DUF protein, F-box protein interaction domain, kinase domain protein, TPR-like domain containing protein, pectinacetyltransferase family protein and conserved hypothetical protein in exons. There was no strong clue from the *sd1* gene revealing the fact that other proteins in this region are equally important.

Key words : Rice, Swarna, N22, Single nucleotide polymorphism, Drought

Drought may simply be defined as reduction in yield due to shortage of water (Bernier *et al.*, 2008). Drought in rice crop can be classified mainly on basis of stage of crop falling under drought stress. There may be three kinds of drought (i) vegetative stage drought, (ii) intermittent drought and (iii) reproductive stage drought or terminal drought. Drought at reproductive stage is the most devastating one in terms of reduction of yield (Kamoshita *et al.*, 2008). So far breeding for drought tolerance is concerned; it has been concentrated on either drought avoidance or resistance. Various physio-molecular approaches have been employed for breeding of drought tolerance in rice. Numerous QTLs have been identified in rice for drought tolerance. So far more than 15 populations have been used for it (plantstress.com; Kamoshita *et al.*, 2008). Most of drought QTLs were found clustered on chromosome 1, 4 8 and 9 (Kamoshita *et al.*, 2008). Distal arm of chromosome 1 has been reported by many workers as most important locus for stress tolerance and yield related traits. Comparative analysis of SNPs in this locus could be quite useful in way to identify candidate genes for drought tolerance and yield as well.

MATERIALS AND METHODS

Comparative analysis of drought tolerance, height and yield:

Data pertaining to height, yield and drought tolerance

Correspondence to:

S.P. SINGH, Department of Crop Physiology, N.D. University of Agriculture and Technology, Kumarganj, FAIZABAD (U.P.) INDIA

Authors' affiliations:

PRASHANT VIKRAM AND ALOK KUMAR SINGH, Department of Genetics and Plant Breeding, T.D.P.G. College, JAUNPUR (U.P.) INDIA

related traits of released rice varieties of India was analyzed. More than 500 rice varieties have been released in India. Comparative analysis of varieties of different rice-ecosystems was done along with their respective plant height data. Similarly plant height of drought tolerant varieties was compared (<http://dacnet.nic.in/Rice/Rice%20Varieties%20-%202001.htm>).

Single nucleotide polymorphism data analysis on distal arm of chromosome 1 :

SNP data of N22, IR64 and Swarna were downloaded from the oryzasnp consortium website. All the QTLs identified so far on distal arm of chromosome 1 were compiled to know exactly the consensus marker associated with yield and drought related traits. There after sequences of rice microsatellite markers RM315 and RM431 were downloaded from Gramene database. Forward sequences were BLAST with whole rice genome to know the coordinates. Pseudomolecule build v4 was taken as reference. To intensify our study we annotated a small region only *i.e.* 39609051 bp and 40648577 bp. SNP polymorphism data of N22/ IR64/ Swarna/ Nipponbare was downloaded. Comparative analysis of SNP polymorphism was done. SNPs within genes were worked out.

Marker name	Position (Co ordinates)
RM315	38609051 bp
RM431	40648577 bp

Comparative analysis of SNPs within the genes:

Single nucleotide polymorphism survey among the genic-regions of N22 and other three varieties *e.g.* Swarna, IR64 and Nipponbare was done. Exons as well

Table 1 : Annotation of SNPs among N22, Swarna, IR64 and nipponbare

Chr.	SNP ID	Coordinate	N 22	Swarna	IR64	nipponbare	Location
1	TBGI063973	39609051	A	G	G	A	
1	TBGI063975	39609338	T	G	G	G	
1	TBGI063981	39651776	T	.	A	A	IR64
1	TBGI063994	39655040	G	T	T	G	
1	TBGI063996	39655377	A	G	G	G	
1	TBGI063998	39665818	A	.	G	A	IR64
1	TBGI064030	39687532	A	G	G	G	
1	TBGI064031	39687634	T	C	C	C	
1	TBGI064038	39689544	A	.	G	A	IR64
1	TBGI064064	39703746	A	G	G	G	
1	TBGI064080	39708455	T	.	C	T	IR64
1	TBGI064089	39709453	T	A	A	T	
1	TBGI064131	39722417	T	C	C	C	
1	TBGI064150	39726528	A	C	C	C	
1	TBGI064151	39726579	C	.	T	C	IR64
1	TBGI064158	39727590	G	.	A	G	IR64
1	TBGI064160	39728883	A	C	C	A	
1	TBGI064162	39729068	A	G	G	A	
1	TBGI064168	39746661	A	G	G	G	
1	TBGI064173	39747743	T	A	A	T	
1	TBGI064182	39747992	G	.	T	G	IR64
1	TBGI064183	39748023	T	A	A	T	
1	TBGI064184	39748048	T	C	C	C	
1	TBGI064190	39748824	C	.	T	C	IR64
1	TBGI064192	39749075	C	T	T	T	
1	TBGI064196	39749583	A	.	T	A	IR64
1	TBGI064204	39751831	A	G	G	G	
1	TBGI064206	39752362	T	C	C	C	Exon
1	TBGI064207	39752832	C	T	T	C	Exon
		39768242	A	T	T	T	?
1	TBGI064217						
1	TBGI064223	39769044	G	A	A	G	Exon
1	TBGI064226	39769954	C	T	T	C	Exon
1	TBGI064227	39770012	A	G	G	G	Exon
1	TBGI064228	39771126	G	T	T	G	?
1	TBGI064239	39775856	C	T	T	C	Intron
1	TBGI064245	39778087	A	T	T	A	?
1	TBGI064248	39778462	A	C	C	C	?
		39779831	G	C	C	C	Exon
1	TBGI064255						
		39780658	G	A	A	A	?
1	TBGI064260						
1	TBGI064262	39782652	C	T	.	T	Swarna
1	TBGI064263	39782726	A	G	.	G	Swarna
1	TBGI064274	39786070	T	G	.	G	Swarna
1	TBGI064280	39786892	A	T	.	T	Swarna
1	TBGI064282	39787091	A	G	.	G	Swarna
1	TBGI064285	39787547	A	C	.	C	Swarna
1	TBGI064287	39787932	T	C	.	C	Swarna

This SNP was between an hsp and DUF266 protein

Table 1 contd.....

Table 1 contd....

1	TBGI064288	39788045	C	.	T	T	IR64		
1	TBGI064299	39800284	T	.	C	T	IR64		
1	TBGI064304	39800960	C	.	A	C	IR64		
1	TBGI064331	39820809	A	.	G	A	IR64		
1	TBGI064346	39824532	T	.	C	T	IR64		
1	TBGI064366	39865939	G	T	T	G			
1	TBGI064373	39868178	T	G	G	T		Intron	
1	TBGI064374	39868396	T	C	C	C		Exon	Os01g0878300
		39868429	A	.	G	A	IR64	Intron	Kinase domain containing
1	TBGI064375			.					protein
1	TBGI064376	39868858	G	A	A	G		Intron	Start
1	TBGI064377	39868898	G	A	A	A		Intron	End
1	TBGI064378	39868966	A	G	G	A		Intron	
1	TBGI064383	39869998	A	G	G	A		Exon	
1	TBGI064384	39870004	T	C	C	C		Exon	
1	TBGI064388	39870190	A	.	T	A	IR64	Exon	
1	TBGI064410	39895659	G	.	A	G	IR64		
1	TBGI064412	39895880	C	T	T	C			
1	TBGI064413	39895969	G	T	T	T			
1	TBGI064418	39897184	T	A	A	A			
1	TBGI064436	39900550	G	C	C	C			
1	TBGI064439	39901026	T	.	G	T	IR64		
1	TBGI064450	39902683	C	.	G	C	IR64		
1	TBGI064465	39904215	G	.	A	G	IR64		
1	TBGI064468	39904631	G	.	A	G	IR64		
1	TBGI064470	39904850	C	.	G	C	IR64		
1	TBGI064492	39907861	T	.	C	T	IR64		
1	TBGI064740	40011443	T	C	C	T			
1	TBGI064750	40012681	G	.	A	G	IR64		
1	TBGI064754	40012936	G	.	A	G	IR64		
1	TBGI064761	40013630	A	C	.	A	Swarna		
1	TBGI064769	40014662	T	.	G	T	IR64		
1	TBGI064780	40017104	C	.	T	C	IR64		
1	TBGI064881	40170267	A	G	G	A			
1	TBGI064891	40172959	C	.	T	C	IR64		
1	TBGI064931	40197105	G	C	C	C			
1	TBGI064994	40230879	C	.	G	C	IR64		
1	TBGI064996	40230914	A	G	G	A			
1	TBGI065012	40233287	G	.	A	G	IR64		
1	TBGI065031	40238716	C	.	T	C	IR64		
1	TBGI065033	40239238	T	C	C	C			
1	TBGI065035	40239306	T	C	C	C			
1	TBGI065048	40268502	C	G	G	G			
1	TBGI065049	40268852	T	A	A	A			
1	TBGI065053	40269838	G	.	A	G	IR64		
1	TBGI065068	40273101	C	.	T	C	IR64		
1	TBGI065089	40296661	A	.	G	A	IR64		
1	TBGI065099	40297592	C	.	T	C	IR64		
1	TBGI065107	40298480	C	T	T	T			
1	TBGI065108	40298598	T	C	C	C			

Contd.... Table 1

Table 1 contd....

1	TBGI065127	40329203	A	G	G	G			
1	TBGI065129	40329319	C	T	T	T			
1	TBGI065130	40329422	G	A	A	A			
1	TBGI065133	40330056	G	T	T	T			
1	TBGI065137	40332056	G	.	A	A	IR64		
1	TBGI065139	40332364	T	G	G	G			
1	TBGI065142	40332797	G	A	A	A			
1	TBGI065146	40333650	A	C	C	C			
1	TBGI065154	40334497	C	T	T	T			
1	TBGI065155	40334597	C	T	T	T			
1	TBGI065156	40334719	T	C	C	C			
1	TBGI065158	40334855	G	T	T	T			
1	TBGI065161	40335346	T	C	C	C			
1	TBGI065169	40373741	G	C	C	C			
1	TBGI065172	40374193	G	.	T	T	IR64		
1	TBGI065174	40374780	A	.	G	G	IR64		
1	TBGI065176	40375346	G	A	A	A		Exon	FGAM Synthase
1	TBGI065177	40376164	T	A	A	A		Exon	Os01g0888500
1	TBGI065179	40377592	C	T	T	T		Exon	
1	TBGI065183	40379257	C	A	A	A			
1	TBGI065185	40379963	C	T	T	T			
1	TBGI065187	40380603	G	A	A	A			
1	TBGI065191	40381647	A	C	C	C			
1	TBGI065198	40382042	A	G	G	G			
1	TBGI065204	40384069	G	T	T	T			
1	TBGI065208	40386622	T	C	C	C			Os01g0888600
1	TBGI065210	40387714	G	.	A	A	IR64		
1	TBGI065212	40388043	T	A	A	T			
1	TBGI065229	40406555	T	C	C	C			
1	TBGI065235	40408117	T	C	C	T			
1	TBGI065286	40418907	T	C	C	T			
1	TBGI065305	40441405	T	C	C	C			
1	TBGI065309	40443938	T	.	C	T	IR64		
1	TBGI065315	40446308	T	C	C	C			
1	TBGI065329	40460862	A	G	G	A			
1	TBGI065331	40460948	G	.	A	G	IR64		
1	TBGI065334	40461294	T	C	C	T			
1	TBGI065335	40461875	A	G	G	A			
1	TBGI065339	40464601	C	.	T	C	IR64		
1	TBGI065340	40464824	G	A	A	G			
1	TBGI065349	40466523	C	T	T	C			
1	TBGI065352	40467715	T	G	G	T			
1	TBGI065354	40489428	C	T	T	C		Exon	Kinase domain containing protein
1	TBGI065358	40490272	C	T	T	C		Exon	Os01g0890600
1	TBGI065360	40491100	T	C	C	C		Just at the tail	
1	TBGI065376	40493275	G	A	A	G		Exon	TPR-like domain containing protein
1	TBGI065377	40493506	T	G	G	G		Intron	

Contd.... Table 1

Contd.... Table 1

1	TBGI065380	40494077	G	.	A	G	IR64			
1	TBGI065462	40510794	G	A	A	G				
1	TBGI065471	40513602	C	.	T	C				
1	TBGI065472	40513655	G	T	T	G				
1	TBGI065525	40570536	T	C	C	T			Exon	Pectinacetylerase family protein
1	TBGI065528	40571006	C	T	T	C			Exon	Os01g0892600
1	TBGI065595	40640532	G	A	A	A			?	
1	TBGI065599	40641693	G	A	A	G			Intron	Fructokinase
1	TBGI065600	40641856	C	A	A	C			Exon	Os01g0894300
1	TBGI065602	40642142	T	A	A	T			Intron	
1	TBGI065628	40648577	A	C	C	A	Exon	Conserved Hypothetical protein		

as introns were studied for the exact location of SNPs.

RESULTS AND DISCUSSION

The results obtained from the present investigation are presented in Table 1.

Drought, height and yield:

After comparing the dwarf, semi-dwarf and tall varieties we can come to conclusion that all the rainfed ecosystems have tall, dwarf and semi-dwarf cultivars suitable for their respective ecosystems. Also the drought tolerant cultivars in all ecosystems were either dwarf or tall or semi-dwarf as well.

Annotation at *sd1* (semi-dwarf 1 gene) locus on chromosome 1:

It has been reported that on the distal end of chromosome 1 most of genes for yield attributes and grain characteristics have been mapped. Quantitative trait loci for different drought tolerance traits have been mapped in this region. Markers RM212, RM315 and RM431 were found correlated with drought tolerance traits in different populations. Some workers suspect it may be due to *sd1* gene (Laffite *et al.*, 2002). Sequences of *sd1* gene were downloaded from Gramene database. Nucleotide sequences of *sd1* as well as related sequences were subjected to BLAST with whole rice genome to locate its position. Position of *sd1* gene was found around 1 Mb

apart from RM315 and RM431. It resides between them. To validate the effect of *sd1* as well as other genes SNPs were critically analyzed in this region.

Single nucleotide polymorphism was worked out among N22 (the drought tolerant variety) Swarna and IR64 (susceptible varieties) between coordinates (RM315 and RM431). Total of 153 SNPs between N22/IR64/Swarna were identified between two coordinates. Nipponbare was also taken in to account and found that there was a 90 Kb region between (40298480 bp to 40387714 bp) unique to N22 only. SNPs were identified within exons of DUF protein, F-box protein interaction domain, kinase domain protein, TPR-like domain containing protein, pectinacetylerase family protein and conserved hypothetical protein in exons. Role of protein kinases in abiotic stresses is well known particularly in cold, drought and salinity stress (Saijo *et al.*, 2000). These genes could be equally important as *sd1* gene if analyzed.

Future prospects:

The annotated proteins as well as the 90 Kb region unique to N22 could be an important clue for further study in way to identify stress tolerance genes from “*Aus*” cultivars like N22, Dular, Vandana (*aus* x *indica*) etc. These genes particularly protein kinases could be confirmed by the RT-PCR (reverse transcriptase-polymerase chain reaction) analysis. Transgenics approaches are also advisable.

REFERENCES

- Bernier Jérôme, Atlin, Gary N., Serraj, Rachid, Kumar, Arvind and Spaner, Dean (2008). Review: Breeding upland rice for drought resistance, *J. Sci. Food & Agric.*, **88**(6):927-939.
- <http://dacnet.nic.in/Rice/Rice%20Varieties%20-%202001.htm>
- Kamoshita, Akihiko, Babu, R., Chandra, Boopathi, N., Manikanda and Shu Fukai (2008). Phenotypic and genotypic analysis of drought-resistance traits for development of rice cultivars adapted to rainfed environments. *Field Crops Res.*, **109**: 1–23.

Khush, G.S. (1997). Origin, dispersal, cultivation and variation of rice. *Plant Molecular Biology*, 35:25-34

Lafitte H.R., Courtois, B., Arraudeau, M. (2002). Genetic improvement of rice in aerobic systems: progress from yield to genes. *Field Crops Res.*, 75:171–190.

Oryzasnp.com

Plantstress.com

Saijo, Y., S. Hata, J. Kyojuka, K. Shimamoto, and K.Izui. (2000). Over-expression of a single Ca²⁺-dependent protein kinase confers both cold and salt/drought tolerance on rice plants. *Plant J.*, 23:319-327.

www.gramene.org

