Asian J. of Bio Sci. (April, 2008) Vol. 3 No. 1 : (52-55)

Computational analysis of single-nucleotide polymorphisms in proximal promoter region of rice (*Oryza sativa* L.)

R.K. DESHMUKH[1]* N.K. DESHMUKH[2] G.B. PATIL[1], S.B. KADAM[1], H. SONAH[1] AND R.N. GACCHE[2]
[1] Department of Biotechnology, Indira Gandhi Krishi Vishwavidyalaya, RAIPUR (C.G.) INDIA
[2] Department of Biotechnology, Swami Ramanand Teerth Marathwada University, NANDED (M.S.), INDIA

(Accepted : November, 2007)

The term promoter is used to designate a region in the genome sequence upstream of a gene transcription start site (TSS). Most promoter elements regulating TSS are localized in the proximal promoter, is a region of several hundreds nucleotides around the TSS. If the SNP occurs within coding sequence, which may or may not alter amino acid sequence. Where as, SNPs occur within the proximal promoter region has more impact on gene expression. The publication of whole genome sequences for the japonica (Nipponbare) and indica (93-11) types of rice enables the determination of common SNPs occurring between them. In present study we used 31 non-redundant proximal promoter sequences in rice, from previous studies which are experimentally determined transcription start site for RNA polymerase II. When proximal promoters of japonica were compared with the corresponding proximal promoters of indica using BLAST alignment programme, a total of 69 SNPs were identified. Out of these 32 SNPs (46.3%) were transitions (A/G, T/C) and 37 SNPs (53.6 %) were transversions (A/C, A/T, G/C, G/T). Maximum frequency of SNPs was found in the region -50 to -150 (61.2 %). Minimum frequency found in region -151 to -201(9.6 %). It demonstrate that functional cis-regulatory polymorphisms segregate within sub species of rice and there is abundant SNP present in proximal promoter region. SNPs in proximal promoter may one of the causes of fictional variation. In addition, study indicates that SNPs are not evenly distributed. Selection pressure is always more for a region having more fictional impact and theoretically should have conserved, in contrast nonfunctional regions acting as reservoir of mutation.

Key words: Cis acting elements, Proximal promoter, SNP

INTRODUCTION

In most cases the genetic variation comprises single base changes in the DNA sequence, known as single nucleotide polymorphisms (SNPs), where this may occurs in coding DNA, or non coding regulatory sequences. The implications are readily apparent, if the changes occur in coding DNA in the form of modified amino acid sequence. Even where the polymorphism remains silent at the protein level, as in the case of synonymous mutations, the effect of the polymorphism can be assayed at the level of mRNA. This is based on the fact that when a polymorphism occurs in coding DNA it will be present in the transcribed mRNA, allowing the relative abundance of the two alleles in cells from an individual heterozygous for a given SNP to be assayed (Knight 2003). In contrast, assaying the functional effect of polymorphisms occurring in non-coding DNA is more problematic. Non-coding regions are interspersed throughout genome and most of them are junk. Promoter is non-coding regulatory sequence upstream of a gene transcription start site (TSS). Promoter elements decide the transcription initiation point, specificity and the rate. Promoters include sets of various elements participating in the complex process of cell, tissue, organ, developmental stage and environmental factors-specific regulation of transcription. Most promoter elements regulating TSS selection are localized in the proximal promoter **(**Shahmuradov**)**, is a region several hundreds nucleotides around the TSS. If the SNP occurs within coding sequence, that may alter amino acid sequence and if the SNP occurs within the proximal promoter region, it may affect the potential alteration in transcription factor binding.The cis-regulatory regions have been hypothesized to facilitate adaptive innovations, because subtle nucleotide changes may generate novel phenotypes while preserving existing functions (Wray et al., 2003). Promoters and other cis-regulatory regions form a protein/ DNA complex with trans-regulatory proteins (transcription factors), thereby promoting interactive and integrative control of the expression. The functional architecture of these regions consists of short and often redundant transcription factor binding sites interspersed within a background sequence of apparently nonfunctional regions. In some cases, binding site loss through point mutation may be easily complimented by remaining other binding sites (Piano *et*

* Author for Correspondence

al., 1999; Ludwig *et al*., 2000; Dermitzakis *et al*., 2003). As the perfect binding sites are often detected in the sequence background of promoter regions, new binding sites can emerge randomly through point mutation, and which may permit the gain of function (Stone and Wray, 2001, Dermitzakis *et al*., 2003). Gene duplication has also been shown to promote gene expression variation, further supporting the idea that new functions can evolve readily from cis-regulatory changes (Gu *et al*., 2004). The evolutionary dynamics of cis- and trans-regulatory regions remain poorly understood. Several studies suggest that abundant neutral expression variation, most of which results from variation at the trans-regulatory level (Brem *et al*., 2002; Enard *et al*., 2002; Von Dassow and Odell, 2002; Khaitovich *et al.*, 2004). In addition, a comparative study of cis-regulatory activity between two closely related *Drosophila melanogaster* species reveals extensive divergence at the cis-regulatory level, with most genes have undergone more cis- than trans-regulatory changes (Wittkopp *et al*., 2004). Thus, cis-regulatory evolution may be the prime determinant of expression changes between species, although trans-regulatory polymorphism is more important within species (Wittkopp *et al*., 2004). Studies in human, fish, Drosophila, and maize document the role of specific cis-regulatory variants in adaptive evolution (Crawford *et al*., 1999; Wang *et al*., 1999; Schulte *et al*., 2000; Michalak *et al*., 2001; Bamshad *et al*., 2002; Lerman *et al*., 2003; Rockman *et al*., 2003). However, the fraction of naturally segregating cis-regulatory polymorphisms that are adaptive remains unknown. The fact that cis-regulatory function cannot be predicted from the basic nucleotide sequence, is a major hurdle to advancement of this field. To identify functional cis-regulatory regions, phylogenetic foot printing is currently the most widely used approach. This method assumes that sequence conservation in non-coding regions indicates function, and it has proven useful in identifying some functionally important elements in promoter sequences (Koch *et al*., 2001; Boffelli *et al*., 2003; Cliften *et al*., 2003). In this study we have identified of SNPs in Proximal promoter, estimated the SNP frequency in different promoter region, and identified the altered *cis*-acting element.

Materials and Methods

DNA sequence of 305 promoter regions [-200:+51] of monocots, with TSS on the fixed position +201 in the FASTA format, were retrieved from plant promoter database (www.softberry.com). A subset of 31 rice promoters of these was screened manually. The sequences of Nipponbare and indica 93-11 in the concerned promoter regions were downloaded from NCBI public domain (www.ncbi.nlm.nih.gov/). Proximal promoters of japonica were compared with the corresponding proximal promoters of indica using BLAST alignment,e. The sequences matched with each other except for few insertion/deletion events and SNPs. Identified SNPs were classified as transitions (A/G, T/C) and transversions (A/C, A/T, G/C, G/T). Each proximal promoter sequences further divided in to five divisions consisting 50 bp and SNP frequency for each division was calculated.

Results and Discussion

Total 69 SNPs were identified in 31 proximal promoter sequence analyzed, out of these 32 SNPs (46.3%) were transitions (A/G, T/C) and 37 SNPs (53.6 %) were transversions (A/C, A/T, G/C, G/T). Observed frequency of SNP was not uniform, maximum frequency of SNP found in region -50 to -150, 19 out of 31 promoter having SNP in these region where as minimum frequency found in region -151 to -201, only three promoter out of 31 having SNP in this region. Total five promoters are conserved among both sequences, showing 100 % similarity namely viz. *RGA*, *CatA*, *OsENOD40*, *RSs1* and *lip19*. Interesting four out of five conserved promoters are TATA-less in addition SNP frequency observed in TATA-less promoters was less than the TATA type promoters (Table. 1)

Transcription factore binding sites are the cis-regulatory regions where protein/DNA complexes are form with transcription factors, thereby promoting integrative control on expression. The functional architecture of these regions consists of short and often redundant transcription factor binding sites interspersed within a background sequence of apparently nonfunctional regions. In some cases, binding site loss through point mutation may be easily compensated by remaining binding sites (Piano *et al*., 1999; Ludwig *et al*., 2000; Dermitzakis *et al*., 2003). Because quasiperfect binding sites are often detected in the sequence background of promoter regions, new binding sites can emerge randomly through point mutation and permit a gain of function (Stone and Wray, 2001; Dermitzakis *et al*., 2003). Functional genetically based variation within species is a prerequisite for adaptive evolution. However, within species surveys of cis-regulatory diversity are scarce, and little is known about the adaptive importance and fate of this diversity. We analyzed nucleotide variation in the proximal promoter region of indica and japonica sub species to evaluate the reservoir of diversity in this region, which may contribute to evolution in rice. We demonstrate that functional cis-

Table 1. Type and position of SNPs in proximal promoter of rice genes

Sr. No.	Gene Name or Product	Promoter type	Accession Number	Position of SNP (type of polymorphism)
1	phospholipase D	TATA-less	AB001920	100 (A/G)
2	RepA	TATA	AB004648	7 (G/T), 57 (C/A), 80 (C/A), 109 (T/C), 138 (C/T), 138(C/T)
3	lip19	TATA	D63955	None
4	ZB8	TATA-less	X87946	87 (-/G), 88 (-/G)
5	GluB-1	TATA	X54314	91 (G/-), 92 (T/-)
6	gns1	TATA	X58877 S76586	66 (T/G), 67 (G/T)
7	GOS2	TATA	X51910	65 (A/G)
8	GOS9	TATA	X51909 S76887	60(C/-)
9	RAmy3A	TATA	X56336	63 (T/C), 147 (G/T)
10	RAmy3C	TATA	X56338	13 (G/-), 14 (A/-), 20 (-/T), 52 (A/-), 89 (G/-), 90 (C/A), 116 (T/A)
11	RSs1	TATA-less	X64770 S41992	None
12	Pdc1	TATA	U26660	31 (A/T), 64 (A/T), 131 (T/A), 165 (C/T)
13	OSHSP18.0	TATA	U83670	18 (-/C), 68 (-/C), 69 (G/T)
14	Osg4B	TATA	D21159	39 (C/T), 49 (A/-), 84 (A/C)
15	RAmy3B	TATA	X56337	207 (G/C), 209 (G/C)
16	type II glutelin	TATA	Y00687	99 (A/-), 100 (A/-), 112 (C/T)
17	Cab1R	TATA-less	X13908 S47120	69 (T/A), 76 (A/C)
18	Osg6B	TATA	D21160	23 (G/A), 69 (T/A), 76 (A/C), 109 (C/A), 124 (T/A), 136 (-/A), 169 (G/C), 180 (-/A), 187 (G/A), 211 (G/C)
19	rep1	TATA	AB004819	118 (T/-)
20	RPA	TATA-less	AJ243829	95 (A/T), 117 (G/A)
21	PIB	TATA-less	AB013448	141 (G/A), 152 (G/A), 155 (G/A), 156 (G/A), 218 (T/C), 248 (G/A)
22	OsENOD40	TATA-less	AB024054	None
23	anther-specific protein	TATA-less	AF042275	92 (-/A),125 (C/G), 126 (G/C), 142 (-/C), 153 (-/A), 171 (C/G), 172 (G/C), 200 (G/-), 221 (C/G)
24	Reb	TATA-less	AF395819	143 (C/A), 83 (A/G)
25	SPS1	TATA-less	U33175	185 (-/G), 218 (-/G), 219(A/C)
26	CatA	TATA-less	D29966	None
27	RGA1	TATA-less	L35844	None
28	sps1	TATA-less	U33175	181 (-/G), 214 (-/G), 215 (A/C)
29	PCNA	TATA-less	X54046	26 (T/C)
30	GT3	TATA	M28158	51 (C/A), 59 (G/A), 75 (T/A), 195 (T/-)
31	Ltp1	TATA	???	45 (A/G), 90 (A/G), 101 (G/T), 106 (T/A), 149 (G/T), 163 (T/A)

regulatory polymorphisms segregate within sub species of rice and there is abundant SNP present in proximal promoter region. SNPs in proximal promoter may one of the causes of fictional variation. In addition, our study indicates that SNPs are not evenly distributed. Selection pressure is always more for a region having more fictional impact and theoretically should have conserved, in contrast nonfunctional regions acting as reservoir of mutation.

REFERENCES

Bamshad, M.J., Mummidi, S., Gonzalez, E., Ahuja, S.S., Dunn, D.M., Watkins, W.S., Wooding, S., Stone, A.C., Jorde, L.B., Weiss, R.B. and Ahuja, S.K. (2002). A strong signature of balancing selection in the 5' cis-regulatory region of CCR5. *Proc. Natl. Acad. Sci. USA*, **99**: 10539–10544.

Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K.D., Ovcharenko, I., Pachter, L. and Rubin, E.M. (2003). Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science*, **299**: 1391–1394.

Brem, R.B.; Yvert, G., Clinton, R. and Kruglyak, L. (2002). Genetic dissection of transcriptional regulation in budding yeast. *Science*, **296**: 752–755.

Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B.A. and Johnston, M. (2003). Finding functional features in Saccharomyces genomes by phylogenetic footprinting. *Science*, **301**: 71–76.

Crawford, D.L., Segal, J.A. and Barnett, J.L. (1999). Evolutionary analysis of TATA-less proximal promoter function. *Mol. Biol. Evol.*, **16**: 194–207.

Dermitzakis, E.T., Bergman, C.M. and Clark, A.G. (2003). Tracing the evolutionary history of Drosophila regulatory regions with models that identify transcription factor binding sites. *Mol. Biol. Evol.*, **20**: 703–714.

Knight, J.C. (2003). Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. *Clinical Sci.*, **104**: 493-501.

Gu, Z.L., Rifkin, S.A., White, K.P. and Li, W.H. (2004). Duplicate genes increase gene expression diversity within and between species. *Nat. Genet.*, **36**: 577–579.

Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansorge, W. and Paabo, S. (2004). A neutral model of transcriptome evolution. *PLoS Biology*, **2**: 682–689.

Koch, M.A., Weisshaar, B., Kroymann, J., Haubold, B. and Mitchell-Olds, T. (2001). Comparative genomics and regulatory evolution: Conservation and function of the Chs and Apetala3 promoters. *Mol. Biol. Evol.*, **18**: 1882–1891.

Lerman, D.N., Michalak, P., Helin, A.B., Bettencourt, B.R. and Feder, M.E. (2003). Modification of heat-shock gene expression in Drosophila melanogaster populations via transposable elements. *Mol. Biol. Evol.*, **20**: 135–144.

Ludwig, M.Z., Bergman, C., Patel, N.H. and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature*, **403**: 564–567.

Michalak, P., Minkov, I., Helin, A., Lerman, D.N., Bettencourt, B.R., Feder, M.E., Korol, A.B. and Nevo, E. (2001). Genetic evidence for adaptation-driven incipient speciation of Drosophila melanogaster along a microclimatic contrast in "Evolution Canyon," Israel. Proc. Natl. *Acad. Sci.* USA, **98**: 13195–13200.

Piano, F., Parisi, M.J., Karess, R. and Kambysellis, M.P. (1999). Evidence for redundancy but not trans factor-cis element coevolution in the regulation of Drosophila Yp genes. *Genetics*, **152**: 605–616.

Rockman, M.V., Hahn, M.W., Soranzo, N., Goldstein, D.B. and Wray, G.A. (2003). Positive selection on a human-specific transcription factor binding site regulating IL4 expression. *Curr. Biol.*, **13**: 2118–2123.

Schulte, P.M., Glemet, H.C., Fiebig, A.A. and Powers, D.A. (2000). Adaptive variation in lactate dehydrogenase-B gene expression: Role of a stress-responsive regulatory element. Proc. Natl. *Acad. Sci*. USA, **97**: 6597–6602.

Shahmuradov, I.A., Gammerman, A.J., Hancock, J.M., Bramley, P.M. and Solovyev, V.V. (2003). PlantProm: a database of plant promoter sequences. *Nucleic Acid Research*, **31**: 114-117.

Stone, J.R. and Wray, G.A. (2001). Rapid evolution of cis-regulatory sequences via local point mutations. *Mol. Biol. Evol.*, **18**: 1764–1770.

Von Dassow, G. and Odell, G.M. (2002.) Design and constraints of the Drosophila segment polarity module: Robust spatial patterning emerges from intertwined cell state switches. *J. Exp. Zool.*, **294**: 179–215.

Wang, R.L., Stec, A., Hey, J., Lukens, L. and Doebley, J. (1999). The limits of selection during maize domestication. *Nature*, **398**: 236–239.

Wittkopp, P.J., Haerum, B.K. and Clark, A.G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature*, **430**: 85–88.

Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V. (2003). The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol.*, **20**: 1377–1419.