Distribution and Associations of GATA Repeats in Rice Genome

Adari Prasad Babu1*, K Aruna Kumari2 and Neelamraju Sarla1

1Acharya N G Ranga Agriculture University, Hyderabad 500030, Andra Pradesh, India
2Rice Research, Rajendranagar, Hyderabad 500030, Andra Pradesh, India

Abstract

GATA repeats are associated with sex differentiation in man, buffalo, mouse and even in plants such as papaya. Human X-chromosome region Xp22 that escapes inactivation is ten-fold enriched in GATA repeats suggesting a role in preventing heterochromatinization. The close proximity of GATA repeats to matrix-associated regions (MARs) indicates a role in chromatin organization and function. Chromosome-wise distribution and density of GATA repeats, neighboring genes and Matrix associated regions were analyzed in rice. (GATA), and higher repeats were distributed non-randomly with the highest frequency on chromosome 11. About 60% of the repeats were found in intergenic regions flanked by regulatory genes involved in stress response or transposable elements. The GATA associated MAR sequences in rice had at least one or more of the consensus sequence to which GATA factors bind. The genomic milieu around GATA repeats suggests that their genomic context may determine their role in chromatin organization and gene regulation.

Keywords: Rice genome; GATA repeats; Matrix attachment regions; Genomics; Chromatin organization

Introduction

Genome sequences of eukaryotes reveal many non-coding sequences, a large proportion of which are repetitive in nature. Several sequenced genomes reveal absence of GATA repeats in prokaryotes and their accumulation in higher organisms during the course of evolution [1]. In animals, GATA repeats play an important role in differentiation of sex chromosomes. A striking ten-fold enrichment of GATA, was reported in the 10 Mb segment at Xp22 region of human X-chromosome [2], that escapes inactivation and a similar enrichment was found in other eutherian genomes [2] indicating their possible role in regulation and formation of facultative heterochromatin. GATA/GACA repeat sequences are transcribed exclusively in Sertoli cells in addition to somatic tissues of normal rats but not infertile rats suggesting their regulatory role in male gonad [3]. Binding of a factor (BBP) to enriched stretches of GATA repeats (Bkm) in the heterogametic sex-specific chromosome of snakes, birds, mouse and man results in germ cell specific decondensation and transcriptional activation of these otherwise highly condensed chromosomes in the somatic tissues [4]. In man, nine GATA repeats in a microsatellite D17S1303 were associated with hypertension while 14 GATA repeats were associated with normal tension [5]. Thus, GATA repeats appear to have a role both in chromatin organization and function. GATA repeat containing markers are routinely used in forensic science and paternity testing because of their high polymorphism [6].

In plants, GATA1 have been used in profiling rice germplasm [7,8], distinguishing various accessions within a single "Marzano" cv. of tomato and also individual plants of the same accession [9], fingerprinting wild and cultivated species of banana [10], Brassica juncea [11], cultivars of pearl millet [12], studying allelic diversity in sunflower [13]. GATA repeats reveal sex-specific differences even in plants. GATA containing sequences which were male specific were found useful in papaya where male and female plants do not show any specific morphological differences until flowering [14,15]. Earlier studies on genetic diversity in rice based on (GAGA)4, (AGAG) and (GATA), primers were very informative in grouping based on genetic relationships and also traits such as drought, flood or salt tolerance [16,17]. (GATA)4 containing sequences grouped bacterial leaf blight (BLB) susceptible and resistant rice varieties separately, indicating their association with BLB resistance [18]. However, the distribution and role of GATA repeats in plants has not been clearly demonstrated. MARs/SARs (Matrix attachment regions or Scaffold attachment regions) are DNA sequence elements that bind with some affinity to sites in the nuclear matrix. MAR from chicken lysozyme reduces variability in transgene expression and confers copy number dependence in transgenic rice plants [19]. Inclusion of MARs from soybean [20] in transgene cassettes reduces position effect variations and enhances the expression of transgenes in barley [21]. MARs can also act as boundary elements creating topologically isolated chromatin domains, which insulate genes located on the loop from cis-acting elements. Boundary elements were identified first in Drosophila and subsequently found to be present ubiquitously from yeast to humans. Predicted GATA-MAR regions on human Y-chromosome have been shown to function as enhancer blockers using transgenic assays in D. melanogaster [22]. Alternatively, the associated GATA repeats may be forming foci of transcription complex for the coordinated expression of spatially regulated genes [23].

Rice has one of the smallest genome among plants with a relatively lower frequency of repeated sequences among monocots. There are about 5251 hyper variable SSRs per Mb or 3 SSRs per gene in the rice genome [24]. Cues to the function of GATA repeats can emerge from analyzing their distribution in rice genome, the kinds of genes associated with them and their association with known boundary elements such as MARs. Additionally, analyses of the genes adjacent to (GATA), or genes in which (GATA), occur will also provide insight into the possible function of these repeats in rice. This study was undertaken to analyze in silico, the frequency and distribution of GATA repeats, the kind of associated genes and the association with putative S/MARs in the rice genome.

Materials and Methods

Repeat analysis

The rice genome sequences were downloaded from ftp sites of IRGSP, http://rgp.dna.affrc.go.jp/e/IRGSP/Build4/build4.html. A java

*Corresponding author: K Aruna Kumari, Rice Research, Rajendranagar, Hyderabad 500030, Andra Pradesh, India, Tel: 919842222253; E-mail: arunaaqbsc@gmail.com

Received July 22, 2016; Accepted September 07, 2016; Published September 14, 2016

Citation: Babu AP, Kumari KA, Sarla N (2016) Distribution and Associations of GATA Repeats in Rice Genome. Mol Biol 5: 173. doi: 10.4172/2168-9547.1000173

Copyright: © 2016 Babu AP, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited
based program was written to analyze the distribution of the perfect tandem repeats of (GATA), and higher repeats in the whole rice genome. Sequences 10 kb on either side of selected GATA repeats were analyzed for Matrix associated region (MAR) potential using the MAR-Wiz program. The analysis was carried out with the window width of 1000, slide distance of 100, scan length of 3 and cut off threshold score for the MAR potential set to 0.60. The core MAR rules included origin of replication rule, TG-richness, curved DNA, Topoisomerase II recognition, AT-richness and also plant MAR consensus.

**Plant material and analysis of ISSRs**

The cultivars of rice used for PCR analysis using (GATA), primers were IR64, IR28, Swarna (control); Vandana, Rasi, Nagina 22 (drought tolerant); FR13A, Jalmagha, Sabita (flood tolerant); and CSR-30 (Yamini), Pokkali and Nonasail (salt tolerant). Selected (GATA), containing ISSRs from these 12 rice cultivars [17] were cloned into TOPO-TA cloning vector (Invitrogen, CA, USA) and the sequences obtained using vector specific primers were analyzed.

**Statistical analysis**

Blast similarity search is a standard method for assessing the statistical significance of molecular sequences to ascertain whether an unusual pattern (perfect GATA repeats in this case) could have arisen simply by chance [25]. This is done by assigning appropriate scoring values to individual residues of sequences employing the formula: \( E = K(n+e)^{-\lambda} \), where \( K=0.035 \) and \( \lambda=0.252 \) are known empirical values, \( E \) is the expectation value (=p value) and \( m \) and \( n \) are length of sequences.

The Wilcoxon signed-rank test on the normalized proportion of perfect GATA repeats and GATA repeats with one mismatch and two mismatches was performed to check if there was any preference for perfect GATA repeats over GATA repeats with one and two allowed mismatches. Percent of AT content was determined as percent of A and T in a window of 250 bases on 5' and 3' of each GATA repeat. The method given in [26] was followed to calculate the correlation between occurrence of GATA repeats and number of genes on each chromosome while correcting for local AT content.

**Results and Discussion**

**Analysis of GATA repeats in rice chromosomes**

There were a total of 787 GATA repeats in the rice genome, with (GATA)\(_5\) representing 50% and other higher repeats ranging from (GATA)\(_{36}\) constituting the other 50% of the total repeats (Table 1). The frequency of the repeats decreased from 395 (GATA)\(_2\), to 8 (GATA)\(_1\) repeats. Five types of repeats (15, 17, 19, 27 and 36) appeared only once in the genome (Table 1). Repeats of (GATA)\(_1\) were observed on all the chromosomes whereas GATA\(_{2,3}\) repeats were absent on chromosome 5.

Score values calculated for various lengths of GATA repeats at E value of 0.01 indicated that no mismatches were tolerated up to 8 repeats of GATA and for repeats higher than (GATA), the allowed mismatches vary depending on sequence length. The value of 8 repeats would be higher for E=0.001, which is the more stringent value used for comparison of nucleotide sequences.

If we look for the distribution of GATA repeats by allowing mismatches, we found 3112 GATA repeats (GATA\(_{3,6}\)) with one mismatch, with (GATA)\(_{3,6}\) representing only 21% (659). When we allow 2 mismatches in the GATA repeats, we observe 15018 GATA repeats (GATA\(_{5,36}\)), with (GATA)\(_{5,36}\) representing only 6% (909) including the perfect GATA repeat (Table 2). Considering that no mismatches are allowed till GATA\(_1\), the representation of GATA\(_{3,6}\) will be only 0.48% which does not appear significant. We performed the Wilcoxon signed-rank test on the normalized proportion of perfect GATA repeats and GATA repeats with one mismatch and two mismatches to check if there was any preference for perfect GATA repeats compared to GATA repeats with one or two allowed mismatches. The p-value indicated that perfect GATA repeats was significantly more preferred than GATA repeats with two allowed mismatches (p-value 0.0011). The preference of perfect GATA repeats over GATA repeats with one mismatch was less significant (p value 0.022).

Apart from GATA, another tetranucleotide repeat GACA has been shown to be enriched in the transcripts from somatic tissues of rat and buffalo, whereas the germ line transcripts in these organisms were enriched for GATA repeats [3,27]. When GACA/GATA repeats were analyzed in *silico* in six species, human, dog and *Arabidopsis thaliana* genomes were GATA-rich and chicken showed similar occurrence of GATA/GACA [27]. Therefore, in addition to GATA repeats, we also analyzed the distribution of GACA repeats and found their numbers insignificant as compared to (GATA)\(_5\) in the rice genome (data not shown).

Chromosome-wise analysis of the GATA repeats from (GATA)\(_{35}\) revealed a maximum of 104 repeats on chromosome 11 and a minimum of 42 repeats on chromosome 10 (Table 1). The frequency of repeats was highest in chromosome 11 with 34 repeats per 10Mb followed by chromosome 12 and 9, with 23 and 18 repeats per 10Mb, respectively. In others it varied from 9 to 17 (GATA), per 10 Mb. Chromosome 11 had the highest number of (GATA), per 10Mb or per 1000 genes and chromosome 3 had the lowest. The GATA repeats were distributed along the entire length of chromosomes. Higher repeats were more and appeared to be clustered on Chromosomes 9, 11 and 12. Chromosomes 1 and 4 had more repeats around the centromere, whereas chromosome 5 showed more repeats at both the telomeric ends. Chromosomes 1 and 3 had the least number of higher repeats. The Pearson correlation coefficient between the perfect GATA repeats per chromosome and chromosome length/no. of genes per chromosome/ was not statistically significant indicating the genome-wide distribution of GATA repeats was non-random. Even when it was corrected for local AT content, i.e., 250 base pairs on either side of GATA repeats, the correlation (0.125) was not significant. The overall AT content per chromosome was 56.5% but in the 500 bp window around GATA repeats it was 63.7% (Table 3). The minimum AT content was 48.12 and the maximum was 74.26 around GATA repeats.

*In situ* hybridization studies showed a preferential localization of GATA repeats in the heterochromatic and/or centromeric chromosomal areas in sugar beet [28], chickpea [29] and tomato [30]. Clustering of GATA was highest in chromosomes 9, 11 and 12, which are also the prominent chromosomes mapped with genes for tolerance to abiotic and biotic stresses in rice. QTLs for submergence tolerance and other biotic stresses have been reported on chromosome 9 [31,32] in the regions where the (GATA), are found clustered. Chromosome 12 is also reported to have many significant QTL/s genomes for tolerance to abiotic [33] and biotic stresses. Thus, there appears to be a correlation between the chromosomal distribution of (GATA)\(_5\) clusters and that of genes/QTLs for tolerance to abiotic and biotic stresses [34]. It is also interesting to note a broad similarity of distribution of (GATA)\(_5\) clusters to the heterochromatin distribution in the 12 chromosomes [35,36] and the distribution of disease resistance gene clusters [34,37] with chromosome 11 showing the maximum number of GATA repeats, largest heterochromatin region among all rice chromosomes.
in Giemsa stained prometaphase mitotic chromosomes [35] and also the number maximum of genes involved in biotic and abiotic stress resistance [2] reported a striking ten-fold enrichment of (GATA), in the 10 Mb segment at Xp22 region of human X-chromosome that escapes inactivation. In this classic paper they clearly demonstrated that presence of (GATA) prevents heterochromatization. Similarity of (GATA), clusters distribution to the heterochromatin distribution in the 12 chromosomes at mitotic stage [35] and meiotic stage [36] suggests that (GATA), in association with MARs may be protecting or shielding these genes from the negative effect that heterochromatization may have on transcription of neighbouring genes [38].

### Distribution of GATA repeats in various genomic regions of rice

In the rice genome, 673 GATA repeats were intergenic and 114 were intragenic. The details of the genes flanking intergenic GATA repeats. The details of genes which had GATA repeats within them. Most of the GATA repeats in the X and Y human chromosomes also were intergenic [1]. Intergenic repeats were more at a distance greater than 5 kb downstream or upstream of genes compared to those within 5 kb of a gene. The frequency of repeats within 1 to 3 kb region of a gene than 5 kb downstream or upstream of genes compared to those within

| No of GATA repeats | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Total |
|--------------------|---|---|---|---|---|---|---|---|---|----|----|----|-------|
| 3                  | 23| 41| 27| 32| 35| 32| 39| 35| 25| 22| 46| 38| 395   |
| 4                  | 17| 9 | 8 | 16| 9 | 8 | 12| 8 | 16| 10| 14| 15| 142   |
| 5                  | 8 | 7 | 2 | 11| 4 | 7 | 7 | 6 | 9 | 4 | 14| 10| 89    |
| 6                  | 4 | 5 | 1 | 1 | 0 | 3 | 4 | 4 | 2 | 2 | 6 | 5 | 37    |
| 7                  | 4 | 2 | 2 | 2 | 0 | 2 | 0 | 3 | 2 | 2 | 8 | 0 | 27    |
| 8                  | 2 | 3 | 0 | 1 | 2 | 0 | 2 | 1 | 2 | 0 | 4 | 3 | 20    |
| 9                  | 1 | 2 | 0 | 0 | 2 | 1 | 1 | 2 | 1 | 1 | 3 | 2 | 16    |
| 10                 | 0 | 1 | 1 | 1 | 1 | 0 | 2 | 0 | 0 | 1 | 4 | 1 | 12    |
| 11                 | 1 | 0 | 1 | 4 | 0 | 0 | 1 | 1 | 2 | 0 | 0 | 2 | 12    |
| 12                 | 1 | 3 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 2 | 10   |
| 13                 | 1 | 1 | 0 | 2 | 1 | 0 | 0 | 1 | 1 | 0 | 2 | 11   |
| 14                 | 0 | 4 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 2 | 0 | 8     |
| 15                 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1     |
| 17                 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1     |
| 19                 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1     |
| 21                 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3     |
| 27                 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1     |
| 36                 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1     |
| Total              | 64| 80| 44| 72| 55| 54| 69| 62| 60| 42| 104| 81| 787   |

**Table 1:** Frequency distribution of (GATA)3 and higher perfect tandem repeats of GATA in the 12 chromosomes of rice.

| chr01 | chr02 | chr03 | chr04 | chr05 | chr06 | chr07 | chr08 | chr09 | chr10 | chr11 | chr12 | Total |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1     | 2018511 | 1694211 | 1693947 | 1640112 | 1389079 | 1466234 | 1397908 | 1343062 | 1082722 | 1073974 | 1365727 | 1318551 | 17484038 |
| 2     | 17574 | 15266 | 14797 | 14184 | 11893 | 12919 | 12173 | 12229 | 9309 | 9482 | 12420 | 11800 | 154046 |
| 3     | 270 | 262 | 230 | 192 | 200 | 192 | 174 | 152 | 150 | 224 | 207 | 2453 |
| 4     | 23 | 29 | 18 | 28 | 27 | 23 | 26 | 31 | 17 | 19 | 30 | 289 |
| 5     | 14 | 7 | 10 | 15 | 6 | 7 | 14 | 6 | 15 | 6 | 16 | 13 |
| 6     | 7 | 7 | 2 | 10 | 3 | 6 | 9 | 7 | 6 | 3 | 11 | 80 |
| 7     | 5 | 2 | 2 | 1 | 4 | 1 | 1 | 1 | 4 | 9 | 3 | 33 |
| 8     | 3 | 4 | 2 | 1 | 1 | 2 | 2 | 1 | 7 | 23 |
| 9     | 3 | 2 | 1 | 3 | 1 | 3 | 1 | 1 | 4 | 3 | 22 |
| 10    | 2 | 1 | 1 | 3 | 1 | 1 | 1 | 2 | 3 | 16 |
| 11    | 1 | 2 | 3 | 1 | 3 | 1 | 1 | 1 | 3 | 13 |
| 12    | 1 | 1 | 3 | 1 | 1 | 2 | 1 | 2 | 12 |
| 13    | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 6 |
| 14    | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 11 |
| 15    | 3 | 1 | 1 | 1 | 2 | 7 |
| 16    | 1 | 1 |
| 18    | 1 | 1 |
| 20    | 1 |
| 21    | 1 | 1 |
| 22    | 1 |
| 28    | 1 |
| 36    | 1 |
| sum   | 330 | 326 | 267 | 258 | 243 | 244 | 249 | 228 | 198 | 185 | 311 | 273 | 3112 |

**Table 2:** GATE repeat distribution considering one mismatch.
distance; the repeats were more downstream of the gene indicating they may have a role in promoter regions of genes in plants.

It was found that the same types of genes (most of them coding for conserved hypothetical protein) flanked GATA repeats (present on both 5' and 3' of repeats) in 70 cases. Of these as many as 20 genes were in chromosome 11 and chromosome 10 had none. The details of such flanking genes are given in Table 4 also reported that the DNA flanking the GATA probe in tomato were highly homologous to each other [30].

Chromosome 11 was enriched in GATA, and also had the highest instances of the same gene present closest on either side of the repeats. In rice many disease and pest resistance genes map to chromosome 11, which has the highest frequency of resistance gene analogues [34]. On the other hand, the absence of instances of same gene present on both sides of (GATA), in chromosome 10 was striking. Wild species derived yield QTLs were also reported to be absent on chromosome 10 [39]. Also miRNA clusters were reported in all rice chromosomes except chromosome 10 [40]. It remains to be seen whether there exists a link between these observations or they are only anecdotal.

The presence of GATA repeats in rice transcriptome has not been reported but in papaya, a GATA repeat containing 0.8 kb sequence was transcribed only in the male plant indicating sex specific expression [15]. The specificity was maintained even in sex reversed (female to male) plants. The differential expression of GACA/GATA tagged transcripts has also been shown in buffalo, where all GATA-tagged transcripts were unique to testes or spermatozoa [27]. It is significant that all GATA-tagged transcripts showed highest expression and were conserved across species.

Association of GATA repeats with MARs

GATA-MAR sequences from human Y-chromosome were shown to function as boundary elements in enhancer blocking assays in D. melanogaster and human cells [41]. It is possible that intergenic GATA-MAR sequences could be acting as insulators thereby regulating genes which have to be expressed at different stages of development or in different tissues. Therefore, 97 intergenic (GATA), flanked by regulatory genes known to be involved in temporal or spatial expression were analyzed for MAR association (Supplementary File 3). All such GATA repeats were found to be associated with MARs with very high scores of 0.7 and above. Some GATA repeats were within the MAR sequences. The length of the MAR sequences ranged from 500 bp to 1 kb, a few was less than 200 bp and a few over 1 kb. Some MAR sequences also included the plant MAR consensus sequence. All the GATA associated MARs were found to have one or more of the GATA factor binding consensus sequences A/T GATA A/G. The genomes of Arabidopsis and rice show 29 and 28 loci respectively, that encode for putative GATA factors [42]. These proteins bind strongly to GATA motifs and regulate transcription of the neighboring genes. A 40 kb DNA (5a) containing 7 tandem repeats of GATA located at the 5'boundary area of Locus Control Region (LCR) of human β-globin gene exhibited a silencer activity in erythroid cells upon binding to GATA-1 protein [38]. GAGA-binding protein was shown to bind specifically to GAGA elements in the promoter of gene encoding chlorophyll and heme synthesis enzyme [43]. Duplication of 305 bp element containing GAGA repeat in the promoter of barley gene activates gene expression in tobacco by binding to BGC, a GAGA-binding factor [44]. Such examples for GATA repeats in promoters having activating or silencing effects are not reported in plants. It appears that GATA-MARs association ensures a certain degree of transcription of genes important for survival and adaptability and such genes are probably shielded from repressing influences.

Experimental validation of the presence of (GATA)₄ in rice cultivars

Analysis of fingerprinting patterns in 12 rice cultivars using GATA primers showed the presence of a 1111 bp unique band in cultivar Vandana, 630-664 bp band in Swarna, Rasi, FR13A, Jalmagna, Pokkali and Nonasail, and 285-320 bp band common to IR28, Swarna, Rasi, FR13A and Jalmagna. The sequences of all three polymorphic bands were intergenic and as expected showed 4 repeats of GATA and at the 5' and 3' ends (Supplementary File 4). BLAST analysis of the sequences identified the gene found upstream of 1110 kb sequence as NADH ubiquinone oxidoreductase chain4 on chromosome 12 and similar to cell cycle control protein cwf14 on chromosome 4. The genes downstream of the sequence were cyclin like F box domain containing protein and ubiquinone specific protease 22, respectively (Sup File 5). Genes coding for impedance induced protein and HD-Zip protein were upstream of the 664 bp and the 316 kb sequences, respectively while the downstream genes were plant lipid transfer/seed storage/trypsin-alpha amylase domain containing protein and inositol 1,3,4-triphosphate 5/6kinase family protein, respectively. All the three ISSR sequences were associated with MARs (the 316kb sequence was included within the MAR sequence) and appear to be associated with genes involved in one or more stress-responses [45-48]. The role of the various genes in stress responses. The PCR products specifically amplified in a given group of cultivars using (GATA)₄ primers could therefore, reflect loci

| Chromosome | GATA repeats | AT content % around GATA repeats 500bp window | AT content % for each chromosome |
|------------|--------------|---------------------------------------------|----------------------------------|
| chr01      | 64           | 63.19                                       | 56.23                            |
| chr02      | 80           | 62.14                                       | 56.67                            |
| chr03      | 44           | 63.48                                       | 56.31                            |
| chr04      | 72           | 64.65                                       | 55.75                            |
| chr05      | 55           | 63.44                                       | 56.05                            |
| chr06      | 54           | 63.55                                       | 56.39                            |
| chr07      | 69           | 63.84                                       | 56.50                            |
| chr08      | 62           | 64.00                                       | 56.62                            |
| chr09      | 60           | 63.00                                       | 56.50                            |
| chr10      | 42           | 62.94                                       | 56.44                            |
| chr11      | 104          | 65.29                                       | 57.10                            |
| chr12      | 81           | 65.05                                       | 56.99                            |
| mean       | 65.58        | 63.71                                       | 56.46                            |
| correlation| 0.27673781   |                                             |                                  |

Table 3: AT content around GATA repeats in each chromosome.
which are transcriptionally active and associated with stress adaptive functions [49].

### Conclusion

Thus, the genomic milieu around GATA repeats presented in our paper suggests that their genomic context may determine their role in chromatin organization and gene regulation. Based on the distribution of (GATA)$_{5,11}$ along the chromosome and their close proximity to Matrix Associated Regions (GATA-MAR) in man it was suggested that they may be delineating chromatin domains for their coordinated expression [1]. There is growing evidence that GATA repeat elements have chromatin domain boundary activity in Drosophila melanogaster as well as in human cells and play a role in packaging of genome and in regulatory mechanisms involving large regions of chromosomes [41].

The role of GATA-MARs associations may be more pronounced in plants as the need to ensure coordinated expression of genes when faced with an environmental stress is more as plants are sessile. The ISSR primers were associated along the chromosome and their close proximity to Matrix Associated Regions (GATA-MAR) in man it was suggested that they may be delineating chromatin domains for their coordinated expression [1]. There is growing evidence that GATA repeat elements have chromatin domain boundary activity in Drosophila melanogaster as well as in human cells and play a role in packaging of genome and in regulatory mechanisms involving large regions of chromosomes [41].

The role of GATA-MARs associations may be more pronounced in plants as the need to ensure coordinated expression of genes when faced with an environmental stress is more as plants are sessile. The ISSR sequences of three of the fragments obtained following PCR analysis of stress-tolerant cultivars using (GATA)$_5$ primers were associated with genes involved in one or more stress-responses. The positional information of perfect GATA repeats provided in this paper and the adjacent genes and MARs can serve as a framework for further analysis of their biological meaning. Eppelen noted that “the question of the functional meaning, if any, of simple, tandemly repeated sequences such as GATA/GACA DNA remains unanswered” [50]. Evidence from several organisms points to a definite role of these repeats in regulation of gene functions and that is not restricted to sex specific differences in man, rat, buffalo or papaya. An overlay of GATA repeat distribution with the distribution of heterochromatin, nucleosome positioning, whole genome methylation, acetylation, AT content and several such features involved in chromatin structure and function may give deeper insights into their function.

### Acknowledgement

APB and CSR were supported by the Department of Biotechnology, Government of India project BT/AB/03/FG-2/2003 as part of Network Project on Rice Functional Genomics and BMRs by CSIR-UGC fellowship. We thank Rashes Mishra for useful discussions. The help of Senthilkumar Ramamurthy in the initial analysis of repeats in rice genome is gratefully acknowledged.

### References

1. Apisitwanch W, Shishido R, Akiyama Y, Fukui K (2001) Quantitative chromosome map of representative indica rice. Euphytica 118:113-118.
2. Bernier J, Kumar A, Ramaiah V, Spaner D, Allin G (2007) A large-effect QTL for grain yield under reproductive-stage drought stress in upland rice. Crop Sci 47: 507-518.
3. Bhattacharaya B, Das S, Jain A, Lakshmikumaran M (1995) DNA fingerprinting of Brassica juncea cultivars using microsatellite probes. Electrophoresis 16: 1750-1754.
4. Blair MW, Panaud O, McCouch SR (1999) Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (Oryza sativa L.). Theor Appl Genet 98:760-792.
5. Breney P, Van Montagu M, depokker A, Gheysen G (1992) Characterization of a plant scaffold attachment region in a DNA fragment that normalizes transgene expression in tobacco. Plant Cell 4:463-471.
6. Buisson CM, Benbow RM (1994) Molecular analysis of transgenic plants generated by microprojectile bombardment, effect of petunia transformation booster sequence. Mol Gen Genet 243: 71-81.
7. Cai HN, Shen P (2001) Effects of cis arrangement of chromatin insulators on enhancer-blocking activity. Science 291: 493-495.
8. Cheng Z, Buell RC, Wing RA (2001) Toward a cytological characterization of the rice genome. Genome Res 11: 2133-2141.
9. Cheung BMY, Leung RYH, Man YB, Wong LYF, Lau CP (2005). Association of essential hypertension with a microsatellite marker on chromosome 17. Journal of Human Hypertension 19:407-411.

10. Chowdari KV, Venkatachalum SR, Davierwala AP, Gupta VS, Ranjekar PK, et al. (1998). Hybrid performance and genetic distance as revealed by the (GATA)n microsatellite and RAPD markers in pearl millet. Theor Appl Genet 97: 163-169.

11. Cui X, Xu SM, Mu DS, Yang ZM (2009). Genomic analysis of rice micro RNA promoters and clusters. Gene 431: 61-66.

12. Davierwala AP, Ramakrishna W, Ranjekar PK, Gupta VS (2000). Sequence variations at a complex microsatellite locus in rice and its conservation in cereals. Theor Appl Genet 101:1291-1298.

13. Davierwala AP, Ramakrishna W, Chowdari V, Ranjekar PK, Gupta VS (2001). Potential of (GATA)n microsatellites from rice for inter- and intra-specific variability studies. BMC Evol Biol 1: 7.

14. Diniz FM, Iyengar A, da Costa Lima PS, Maclean N, et al. (2007). Application of a double enrichment procedure for microsatellite isolation and the use of tail primers for high throughput genotyping. Genetics and Mol Biol 30: 390-394.

15. Esplen JT (1988). On simple repeated GATA sequences in animal genomes, a critical reappraisal. Journal of Heredity 79: 409-417.

16. Gangadharam S, Kapur V, Ali S (2001). GATA/GACA repeat sequences are transcribed in the normal fertile rat Rattus norvegicus but not in the infertile ones. Curr Sci 81: 1320-1324.

17. Gangopadhyay G, Roy SK, Ghose K, Poddar R, Bandyopadhyay T, et al. (2007). Sex determination in Carica papaya and Cucuris circinalis in pre-flowering stage by ISSR and RAPD. Curr Sci 92: 524-526.

18. Gortner G, Nenno M, Weising K, Zirk D, Nagit W, et al. (1998). Chromosomal localization and distribution of simple sequence repeats and the Arabidopsis-type telomere sequence in the genome of Cicer arietinum L. Chrom Res 6: 97-104.

19. Henriksson E, Olsson ASB, Johannesson H, Johannsson H, Hanson J, et al. (2005). Homeodomain leucine zipper class I genes in Arabidopsis expression patterns and phylogenetic relationships. Plant Physiol 139: 509-518.

20. International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome. Nature 436: 793-800.

21. Kaemmer D, Atza R, Weising K, Kahl G, Novak FJ (1992). Oligonucleotide and amplification fingerprinting of wild species and cultivars of banana (Musa spp.). Bio/Technology 10: 1030-1035.

22. Karlin S, Altschul SF (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. PNAS 87: 2264-2268.

23. Kim SH, Yi S (2007). Understanding relationship between sequence and functional evolution in yeast proteins. Genetics 131:151-156.

24. Kottappali KR, Sarla N, Kikuchi S (2006). In silico insight into two rice chromosomal regions associated with submergence tolerance and resistance to bacterial leaf blight and gull midge. Biotech Advances 24: 561-569.

25. Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, et al. (2002). Transcriptome changes for Arabidopsis in response to salt, osmotic and cold stress. Plant Physiol 130: 2129-2141.

26. Kumar RP (2007). Bkm (GATA-repeats) associated chromatin domain boundaries of human Y-chromosome. PhD Thesis Submitted to Jawaharlal Nehru University, New Delhi, India.

27. Lawson MJ, Zhang L (2006). Distinct patterns of SSR distribution in the Arabidopsis thaliana and rice genomes. Genome Biology 7:R14.

28. Leroy JX, Leon K, Branchard M (2000). Plant genomic instability detected by microsatellite-primers. Eur J Bot 3: 140-148.

29. McNeil JA, Smith KP, Hall LL, Lawrence JB (2006). Word frequency analysis reveals enrichment of dinucleotide repeat sequences in the human X chromosome and [GATA]n in the X escape. Genome Res 16: 477-484.

30. Michalak P (2008). Co-expression, co-regulation and co-functionality of neighboring genes in eukaryotic genomes. Genomics 91: 243-248.

31. Mishra R (2005). CCMB Annual Report 99.

32. Mishra R (2007). Genome chromatin and regulation of genetic information. CCMB Annual Report 83.