Research Article
Evidence for Directed Evolution of Larger Size Motif in Arabidopsis thaliana Genome

Rajesh Mehrotra,1,2 Amit Yadav,¹ Purva Bhalothia,¹ Ratna Karan,2 and Sandhya Mehrotra ¹

1 Department of Biological Sciences, Birla Institute of Technology and Science, Pilani, Rajasthan 333031, India
2 Department of Plant Environment and Soil Sciences, Louisiana State University, Baton Rouge, LA 70894, USA

Correspondence should be addressed to Rajesh Mehrotra, rajmeh25@hotmail.com

Received 14 October 2011; Accepted 15 December 2011

Academic Editors: Z.-G. Han, Y. Y. Shugart, and W. A. Thompson

Copyright © 2012 Rajesh Mehrotra et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Transcription control of gene expression depends on a variety of interactions mediated by the core promoter region, sequence specific DNA-binding proteins, and their cognate promoter elements. The prominent group of cis acting elements in plants contains an ACGT core. The cis element with this core has been shown to be involved in abscisic acid, salicylic acid, and light response. In this study, genome-wide comparison of the frequency of occurrence of two ACGT elements without any spacers as well as those separated by spacers of different length was carried out. In the first step, the frequency of occurrence of the cis element sequences across the whole genome was determined by using BLAST tool. In another approach the spacer sequence was randomized before making the query. As expected, the sequence ACGTACGT had maximum occurrence in Arabidopsis thaliana genome. As we increased the spacer length, one nucleotide at a time, the probability of its occurrence in genome decreased. This trend continued until an unexpectedly sharp rise in frequency of (ACGT)N25(ACGT). The observation of higher probability of bigger size motif suggests its directed evolution in Arabidopsis thaliana genome.

1. Introduction

Gene expression in eukaryotic organisms has been a topic of great interest. Careful regulation and recruitment of transcription factors (TFs) to cis regulatory elements in promoter regions lead to generation of specificity and diversity [1] in genetic regulation. Promoters are arrays of cis regulatory elements present upstream of a gene arranged with other specific cis elements. At present 469 cis elements have been reported in the plant cis regulatory element (PLACE) database. The prominent group of cis acting elements in plants contains an ACGT core. Several cis elements with this core have been shown to be responding to abscisic acid [2–4], salicylic acid [5], and light signals [6]. It has been reported by Foster et al. [7] that bZIP class of transcription factors binds to this core motif. In an elegant study Krawczyk et al. [8] showed deletion of two base pairs between activator sequence-1 (as1) palindromes does not affect binding of activator sequence binding factor (ASF-1) and TGA factors (which binds to TGACG sequence), whereas insertion decreases factor binding in vitro. In their study the distance between palindromic centers was 12 base pairs. Mehrotra et al. [9, 10] have shown that this motif functions even when they are placed out of the native context. R. Mehrotra and S. Mehrotra [11] have shown that promoter activation by ACGT in response to salicylic and abscisic acids is differentially regulated by the spacing between these motifs. It contributes synergistically to gene expression by stabilising the transcription complex formed on minimal promoter [10]. The present study is an extension of aforementioned work. In this study, genome-wide comparison of the frequency of occurrence of two ACGT elements without any spacers and also separated by spacers of different length was done. Based on the data obtained we report that there is a directed evolution of bigger size of motif in the Arabidopsis thaliana genome.

2. Materials and Methods

The objective was to find out the frequency of the recurring sequences and then use these recurring sequences with
Table 1: Frequency of occurrence of the various promoter sequences in which spacer sequence length between two ACGT palindromes is gradually increased from 5 to 25 nucleotides.

| Cis element                  | Chromosome 1 | Chromosome 2 | Chromosome 3 | Chromosome 4 | Chromosome 5 | Total |
|------------------------------|--------------|--------------|--------------|--------------|--------------|-------|
| (ACGT)$_2$                  | ACGTACGT     | 469          | 312          | 367          | 327          | 410   | 1885 |
| (ACGT)$_8$                  | ACGTACGT     | 70           | 31           | 12           | 28           | 59    | 200  |
| (ACGT)$_{N5}$ (ACGT)        | ACGTGGGCTAAAGC | 16     | 11           | 13           | 13           | 19    | 72   |
| (ACGT)$_{N10}$ (ACGT)       | ACGTGGGCTATGGGCGGAGC | 8     | 5            | 10           | 4            | 12    | 39   |
| (ACGT)$_{N25}$ (ACGT)       | ACGTGGGCTATGGGCGGAGC | 15     | 12           | 13           | 9            | 13    | 62   |
| (ACGT)$_{RN5}$ (ACGT)       | ACGTACGT     | 7            | 5            | 5            | 2            | 4     | 23   |
| (ACGT)$_{RN10}$ (ACGT)      | ACGT--GCTAG--ACGT | 2     | 2            | 4            | 3            | 3     | 14   |
| (ACGT)$_{RN25}$ (ACGT)      | ACGTAGACGGTTGGGGG--AAGCTTCAGTCGCA | 3     | 1            | 7            | 5            | 5     | 21   |
| (ACGT)$_{RN25}$ (ACGT)      | ACGT--ATATGAGATCGGGCGGAC--TCACGGGAC--ACGT | 4     | 14           | 6            | 4            | 4     | 32   |
| (ACGT)$_{N5}$ (ACGT)        | randomized   | GGAATCCCTTGGCA | 41     | 24           | 30           | 19    | 23   | 137  |
| (ACGT)$_{N10}$ (ACGT)       | randomized   | GCGGGCTATGGGCGGAC | 2     | 5            | 2            | 0     | 1    | 10   |
| (ACGT)$_{N25}$ (ACGT)       | randomized   | TAAGGCTTAGGCACGCC--AGGGTGTGACGCAC | 6     | 6            | 3            | 0     | 3    | 18   |
| (TGCA)$_{N25}$ (TGCA)       | randomized   | TGGACGCTATGGGCGGAGC--AAGGTTACCTTCGCA | 13     | 12           | 9            | 12    | 9    | 55   |

N5, N10, N25 denote sequence length between two ACGT palindromes. RN5, RN10, RN25—signify only spacer sequence being randomized. (ACGT)$_{N}$ (ACGT) randomized—signify complete sequence being randomized.

3. Results and Discussion

3.1. Promoters with Greater Length between ACGT Motifs Are More Frequent. It has been reported that ACGT cis elements function even when they are placed out of native sequence context [9, 10]. When the distance of separation between two ACGT elements are 5 base pairs, and 10 base pairs, they are induced in response to salicylic acid (SA) and abscisic acid (ABA), respectively. Interestingly, SA mimics biotic stress response and ABA mimics abiotic stress response in plants and thus is of great interest to plant biologists. Paixão and Azevedo [16] showed that multiplicity of cis element evolved through transitional forms showing redundant cis regulation. In this study, when the frequency of occurrence of two ACGT elements without any spacers and also separated by the spacer of different lengths was observed, we found that the total frequency of occurrence of two ACGT element in tandem is 1885 (Table 1), while the e value was same for all alignments obtained on a particular chromosome. When two ACGT elements were separated by spacer of 5, 10, and 25 nucleotides their frequency of occurrence was 72, 39, and 62, respectively. An unexpectedly high frequency of occurrence was observed when two ACGT elements were separated by 25 base pairs. According to the rule of probability the frequency of two ACGT elements separated by 25 base pairs should be less than when they are separated by 10 base pairs or lesser. Hobo et al. [17] have earlier reported that in ABA responsive promoters the distance between ACGT elements...
Table 2: Frequency of occurrence of nitrogenous bases when spacer sequence length between two ACGT palindromes is gradually increased from 5 to 25 nucleotides.

| (ACGT)$_{N5}$(ACGT) | ACGTGCT_ACGT | 72 42 33 34 72 5 690 |
|-----------------------|---------------|------------------------|
| (ACGT)$_{N6}$(ACGT) | ACGTGCTA_ACGT | 98 65 45 44 44 6 611 |
| (ACGT)$_{N7}$(ACGT) | ACGTGCTAT_ACGT | 92 91 77 80 77 7 824 |
| (ACGT)$_{N8}$(ACGT) | ACGTGCTATT_ACGT | 97 30 64 55 64 8 852 |
| (ACGT)$_{N9}$(ACGT) | ACGTGCTATTG_ACGT | 39 32 22 32 32 9 602 |
| (ACGT)$_{N10}$(ACGT) | ACGTGCTATTGG_ACGT | 34 36 39 66 39 10 600 |
| (ACGT)$_{N11}$(ACGT) | ACGTGCTATGGCG_ACGT | 36 23 38 29 38 11 681 |
| (ACGT)$_{N12}$(ACGT) | ACGTGCTATGGCGGA_ACGT | 56 54 65 45 56 12 638 |
| (ACGT)$_{N13}$(ACGT) | ACGTGCTATGGCGGAAG_ACGT | 78 50 77 59 77 13 652 |
| (ACGT)$_{N14}$(ACGT) | ACGTGCTATGGCGGAAGA_ACGT | 86 53 96 52 53 14 841 |
| (ACGT)$_{N15}$(ACGT) | ACGTGCTATGGCGGAAGAC_ACGT | 56 67 44 66 56 15 709 |
| (ACGT)$_{N16}$(ACGT) | ACGTGCTATGGCGGAACGA_ACGT | 60 34 52 34 60 16 843 |
| (ACGT)$_{N17}$(ACGT) | ACGTGCTATGGCGGAACGAAG_ACGT | 39 41 42 39 42 17 830 |
| (ACGT)$_{N18}$(ACGT) | ACGTGCTATGGCGGAACGAAGAG_ACGT | 49 47 58 48 49 18 719 |
| (ACGT)$_{N19}$(ACGT) | ACGTGCTATGGCGGAACGAAGAGA_ACGT | 50 38 49 44 44 19 695 |
| (ACGT)$_{N20}$(ACGT) | ACGTGCTATGGCGGAACGAAGAGAA_ACGT | 34 30 44 37 37 20 821 |
| (ACGT)$_{N21}$(ACGT) | ACGTGCTATGGCGGAACGAAGAGAAAG_ACGT | 36 40 42 43 40 21 717 |
| (ACGT)$_{N22}$(ACGT) | ACGTGCTATGGCGGAACGAAGAGAAAGAG_ACGT | 53 42 42 46 53 22 726 |
| (ACGT)$_{N23}$(ACGT) | ACGTGCTATGGCGGAACGAAGAGAAAGAGAG_ACGT | 91 55 60 61 55 23 771 |
| (ACGT)$_{N24}$(ACGT) | ACGTGCTATGGCGGAACGAAGAGAAAGAGAGA_ACGT | 77 64 57 53 53 24 1171 |
| (ACGT)$_{N25}$(ACGT) | ACGTGCTATGGCGGAACGAAGAGAAAGAGAGAAG_ACGT | 76 62 58 69 62 25 708 |

Table 3: Alterations in transcription factor binding sites when spacer sequence length between two ACGT palindromes is gradually increased from 5 to 25 nucleotides.

| Minimal promoter sequence (MPS) | (ACGT) | (ACGT)$_2$(MPS) | (ACGT)$_N5$(MPS) | (ACGT)$_N10$(MPS) | (ACGT)$_N25$(MPS) |
|--------------------------------|--------|-----------------|-----------------|------------------|------------------|
| ARR10                          | 0      | 0               | 0               | 0                | 0                |
| AGL3                           | 2      | 0               | 2               | 2                | 2                |
| ATHB-5                         | 1      | 0               | 1               | 1                | 1                |
| bZIP910                        | 0      | 0               | 0               | 1                | 1                |
| DoF3                           | 1      | 0               | 1               | 1                | 1                |
| EmBP-1                         | 2      | 0               | 2               | 2                | 2                |
| Gamyb                          | 5      | 0               | 5               | 5                | 5                |
| HAT5                           | 2      | 0               | 2               | 2                | 2                |
| HMG-1                          | 6      | 0               | 6               | 6                | 6                |
| HMG-I/Y                        | 6      | 0               | 6               | 6                | 6                |
| id1                            | 5      | 0               | 5               | 5                | 5                |
| myb.Ph3                        | 1      | 0               | 1               | 1                | 1                |
| PEND                           | 1      | 0               | 1               | 1                | 1                |
| squamosa                       | 2      | 0               | 3               | 3                | 3                |
| TGA1A                          | 1      | 0               | 1               | 2                | 2                |

Model name Frequency

| Model name | Frequency |
|------------|-----------|
| ARR10      | 1         |
| AGL3       | 2         |
| ATHB-5     | 1         |
| bZIP910    | 1         |
| DoF3       | 1         |
| EmBP-1     | 2         |
| Gamyb      | 5         |
| HAT5       | 2         |
| HMG-1      | 6         |
| HMG-I/Y    | 6         |
| id1        | 5         |
| myb.Ph3    | 1         |
| PEND       | 1         |
| squamosa   | 3         |
| TGA1A      | 2         |

35 0 36 36 39 38 40
A central question in promoter evolution is to know how does cis regulatory element multiplicity evolved. The promoter regions of many genes contains multiple binding sites for the same transcription factor. Multiplicity may have evolved through transitional forms showing redundant cis regulation. In this paper, we focused on multiplicity of ACGT cis element and the distances between them which occurs in natural promoters. We found that ACGT element separated by 25 base pairs is more frequent than those by 10 base pairs which is against the law of probability. It signifies that under some evolutionary forces this interval was favoured since this distance may cause changes in the level of gene expression or in its robustness against variation in transcription factor concentration. Selection for different levels of expression of certain genes in certain environment could, over time, generates a positive association between cis element multiplicity and expression level.

**Acknowledgments**

The authors are grateful to the Department of Science and Technology, New Delhi, India for Grant-in-Aid and financial support to carry out this work bearing the file no. SR/FT/LS-126/2008. The authors are grateful to the administration of Birla Institute of Technology and Sciences, Pilani, Rajasthan for providing logistic support. They are thankful to Professor C. Gatz for critically reading the paper.

**References**

[1] G. A. Wray, M. W. Hahn, E. Abouheif et al., “The evolution of transcriptional regulation in eukaryotes,” *Molecular Biology and Evolution*, vol. 20, no. 9, pp. 1377–1419, 2003.

[2] M. J. Guiltnan, W. R. Marcotte, and R. S. Quatrano, “A plant leucine zipper protein that recognizes an abscisic acid response element,” *Science*, vol. 250, no. 4978, pp. 267–271, 1990.

[3] Q. Shen and T. H. Ho, “Functional dissection of an abscisic acid (ABA)-inducible gene reveals two independent ABA-responsive complexes each containing a G-box and a novel cis-acting element,” *Plant Cell*, vol. 7, no. 3, pp. 295–307, 1995.

[4] P. K. Busk and M. Pagès, "Regulation of abscisic acid-induced transcription," *Plant Molecular Biology*, vol. 37, no. 3, pp. 425–435, 1998.

[5] I. Jupin and N. H. Chua, “Activation of the CaMV as-1 cis-element by salicylic acid: differential DNA-binding of a factor related to TGA1a,” *EMBO Journal*, vol. 15, no. 20, pp. 5679–5689, 1996.

[6] R. G. K. Donald and A. R. Cashmore, “Mutation of either G box or I box sequences profoundly affects expression from the Arabidopsis rbcS-1A promoter,” *EMBO Journal*, vol. 9, no. 6, pp. 1717–1726, 1990.

[7] R. Foster, T. Izawa, and N. H. Chua, “Plant basic leucine zipper proteins gather at ACGT elements,” *EASEB Journal*, vol. 8, pp. 192–200, 1994.

[8] S. Krawczyk, C. Thurow, R. Niggeweg, and C. Gatz, “Analysis of the spacing between the two palindromes of activation sequence-1 with respect to binding to different TGA factors and transcriptional activation potential,” *Nucleic Acids Research*, vol. 30, no. 3, pp. 775–781, 2002.

[9] R. Mehrrota, K. Kiran, C. P. Chaturvedi et al., “Effect of copy number and spacing of the ACGT and GT cis elements on transient expression of minimal promoter in plants,” *Journal of Genetics*, vol. 84, no. 2, pp. 183–187, 2005.
[10] S. V. Sawant, K. Kiran, R. Mehrotra et al., “A variety of synergistic and antagonistic interactions mediated by cis-acting DNA motifs regulate gene expression in plant cells and modulate stability of the transcription complex formed on a basal promoter,” *Journal of Experimental Botany*, vol. 56, no. 419, pp. 2345–2353, 2005.

[11] R. Mehrotra and S. Mehrotra, “Promoter activation by ACGT in response to salicylic and abscisic acids is differentially regulated by the spacing between two copies of the motif,” *Journal of Plant Physiology*, vol. 167, no. 14, pp. 1214–1218, 2010.

[12] L. Doelz, *BioCompanion*, Biocomputing Essentials Series, Dr. Ing. U. Doelz, Basel, Switzerland, 1990.

[13] S. Sawant, P. K. Singh, R. Madanala, and R. Tuli, “Designing of an artificial expression cassette for the high-level expression of transgenes in plants,” *Theoretical and Applied Genetics*, vol. 102, no. 4, pp. 635–644, 2001.

[14] J. C. Byrne, E. Valen, M. H. E. Tang et al., “JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update,” *Nucleic Acids Research*, vol. 36, no. 1, pp. D102–D106, 2008.

[15] A. Sandelin, W. W. Wasserman, and B. Lenhard, “ConSite: web-based prediction of regulatory elements using cross-species comparison,” *Nucleic Acids Research*, vol. 32, pp. W249–W252, 2004.

[16] T. Paixão and R. B. R. Azevedo, “Redundancy and the evolution of cis-regulatory element multiplicity,” *PLoS Computational Biology*, vol. 6, no. 7, Article ID e1000848, 2010.

[17] T. Hobo, M. Asada, Y. Kowyma, and T. Hattori, “ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent,” *Plant Journal*, vol. 19, no. 6, pp. 679–689, 1999.