Analysis of regulatory elements in GA2ox, GA3ox and GA20ox gene families in *Arabidopsis thaliana*: an important trait

Shiferaw Teshome\(^a,b\) and Mulugeta Kebede\(^a\)

\(^a\)Department of Applied Biology, School of Applied Natural Science, Adama Science and Technology University, Adama, Ethiopia; 
\(^b\)Department of Biotechnology, College of Natural and Computational Science, Wolaita Sodo University, Sodo, Ethiopia

**ABSTRACT**

The enzymes that are independently encoded by GA2ox, GA3ox and GA20ox genes control gibberellins (GA) biosynthesis and catabolism. This study analysed the promoter regions, transcription start site (TSS), motifs and CpG islands (CGIs) in 15 GA oxidase genes of *Arabidopsis thaliana* using 1-kb sequences. Excluding ATGA2ox5 and ATGA3ox4, promoter regions and TSS analysis showed that 60.00% of the genes have more than one TSS in 1000 bp. Relative to the ATG, 80.00% of TSSs with highest score were in the range of −100 bp. Sixty percent of investigated GA genes have TSS in their 5′ untranslated regions (UTRs) with highest prediction score at 0.8 cut-off values. A total of 5 motifs were discovered in 15 GA oxidase genes. More than half (53.33%) of the motifs were distributed throughout promoter regions on both strands. Positional clustering of more than half (55.31%) of motifs was found between +1 and −500 bp. MfIV was detected (100%) in all promoter regions of GA genes. Using the Plant CARE database, various similar motifs were identified in all promoter regions of GA genes. Including two gibberellin-responsive elements (GARE-motif, P-box), numerous motifs such as CAAT-box, TATA-box, ATC-motif, Gap-box and others were distinguished in the promoter regions of GA genes. CGIs were not detected in the promoter sequences of all GA oxidase genes. However, 86.66% of CGIs were discovered in the gene bodies of GA oxidase genes. Analysing the regulatory elements of GA genes is the first step to control crop height, increase yield and resistance to lodging.

**Introduction**

Plant hormones are a group of organic compounds that influence growth, differentiation and other physiological processes at very low concentrations. Gibberellins (GAs) are endogenous plant hormones that act throughout the life cycle of plants and play a role in seed germination, anther development, shoot growth, stem elongation, flower induction and development, and fruit development [1]. In addition to their intrinsic effect, GAs could also have correlation with environmental conditions which play a role in plant development. GAs are a natural large family of tetracyclic diterpenoid phytohormones; 136 GAs have been identified in higher plants and fungi [2]. Three major oxidase gene families of GA, namely GA 2-oxidase (GA2ox), GA 3-oxidase (GA3ox) and GA 20-oxidase (GA20ox), are involved in GA catabolism and biosynthesis. The biosynthesis of bioactive GAs, such as GA\(_1\) and GA\(_3\), is performed from trans-geranylgeranyl diphosphate by the serial action of membrane-associated mono-oxygenases in plastids and soluble dioxygenases in the cytosol [1,2].

GA2oxs are the key genes that cause dwarfism or semi-dwarfism in plants by reducing the availability of endogenous bioactive GA (GA\(_1\) and GA\(_3\)) content in plants [3–5]. The GA2ox gene, which belongs to a small multigene family, encodes GA 2-oxidase, the major catabolic enzyme in plants that converts the active GA to an inactive one. So far, eight GA2ox genes have been identified in *Arabidopsis*: AtGA2ox1, AtGA2ox2 and AtGA2ox3 [6], AtGA2ox4, AtGA2ox5 and AtGA2ox6 [7] and AtGA2ox7 and AtGA2ox8 [8]. Five GA2ox members that catalyze the conversion of GA\(_{12}\) to GA\(_9\) or GA\(_{53}\) to GA\(_{20}\) are encoded by a multi-member gene family in *Arabidopsis* [9]. The expression of GA20ox gene members is tissue-specific or developmental stage-specific. The conversion of precursor GA\(_{20}\) to bioactive forms (GA\(_1\)) or GA\(_9\) to bioactive forms (GA\(_3\))

**CONTACT** Shiferaw Teshome \(\text{tesheshife@gmail.com}\) Department of Applied Biology, School of Applied Natural Science, Adama Science and Technology University, P.O. Box 1888, Adama, Ethiopia

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
is catalyzed by GA3ox. Four GA3ox (designated as AtGA3ox1 to AtGA3ox4) gene members that play a direct role in determining the levels of bioactive GAs have been identified in Arabidopsis [10]. The GA3ox gene member in the GA biosynthetic pathway catalyzes the final step, which leads to the production of bioactive GAs. In the final stages of GA biosynthesis, GA20ox and GA3ox convert GA12 to bioactive GA in the cytosol.

The regulatory element in plant genes usually defined as 1000bp 5′ upstream of the transcription start site (TSS) is known as a promoter. The promoter contains sequences that serve as binding sites for RNA polymerase II and several transcription factors (TFs), which play a significant role in transcription. The TFs that regulate gene expression bind to short DNA sequences called motifs in the promoter regions of particular genes. These cis-motifs, also known as cis-acting regulatory elements, or the transcription factor-binding sites (TFBs), in promoter regions are approximately 5–25 bp long [11]. Regulation of gene expression via TFs is critical for plant survival and adaptation, so the characterization of TF function and determination of TF binding sites in promoter regions of target genes is very crucial [12]. To date, there are several computational and experimental approaches as well as some databases that have been employed to identify TFs and TF-binding sites. However, there is still poor evidence about relationship between most TFs and cis-acting regulatory elements. For instance, in Arabidopsis among 1717 TFs, only 64 TFs and 3 TF complexes with their target genes have been characterized [12].

The agronomic importance of GAs has been reported in various plants. The stems of tall cereal crops such as wheat, maize, and rice are susceptible to lodging with subsequent large yield losses. Lodging resistance, increased yield and improved harvest index in wheat and rice were achieved during the Green Revolution by introduction of dwarfing traits into the crops [13]. GA is one of the most important plant growth hormones that determines plant height, even though various reasons cause dwarf phenotype in plants [14]. The height of plants is regulated either by lowering the availability of bioactive GA content or by inhibiting GA signalling pathway. Inducing dwarfism or semi-dwarfism in different plant species is of agronomic importance and these futures are widely regulated by GA genes. Semi-dwarfism is a valuable trait in many crop species and semi-dwarf plants possess short and strong stalks [13], which are more resistant to lodging (to wind and rain damage) and increase yield. Semi-dwarfism in trees and horticulture has significant economic and environmental benefits including increased yield, reduced lodging, altered fibre and fruit production, greater ease of management, for short-rotation forestry or energy plantations, reclamation, phytoremediation, and reduced risk of spread in wild populations [5,15]. Plant transformation and generation of transgenic dwarf or semi-dwarf phenotypes through manipulation of GA genes has been reported in several plant species [3,4,15–18].

Even though there are many works on the biosynthesis and metabolism of GAs, there are limited reports on the TFs, cis-acting regulatory elements and CpG content that regulate the expression of GA2ox, GA3ox and GA20ox genes. Thus, the aim of this study was to predict the regulatory elements present in the 5′ upstream region of GA2ox, GA3ox and GA20ox gene families in A. thaliana using bioinformatics tools.

Materials and methods
Identification of transcription start sites and promoter regions of ATGA2OXs, ATGA20OXs and ATGA3OXs genes

The nucleotide sequences of GA2ox, GA3ox and GA20ox gene families in A. thaliana were obtained from the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nih.nim.gov) and TAIR database (https://www.arabidopsis.org/index.jsp). The genes were registered in Genbank under different accession numbers. Each gene was retrieved as a FASTA file and coding sequences and ATG (translational start site) were identified for analysis. The TSSs were determined by excising 1-kb sequences upstream from the ATG for each gene family. In this analysis, TSS of all genes were scanned using the promoter prediction tool set called Neural Network Promoter Prediction (NNPP version 2.2) with the minimum standard predictive score (between 0 and 1) at a cut-off value of 0.8 [19].

Identification of motifs and CpG islands in ATGA2OXs, ATGA20OXs and ATGA3OXs genes

MEME (Multiple EM for Motif Elicitation) is one of the most popular tools used for discovering novel signals called motifs in DNA sequences. MEME tools are applicable in searching for new TF-binding sites and protein domains, searching for repeated and ungapped sequence patterns in DNA or protein sequences [20]. Based on the previously predicted promoter region, the conserved motifs for all genes were discovered...
using MEME software Version 5.1.1 (http://meme-suite.org/tools/meme) [21]. The sequences in the FASTA format were used within minimum and maximum motif widths between 6 and 50 and a maximum number of 5 motifs, by keeping the rest of the parameters at the default settings. In the MEME output, the HTML page displayed the motifs as local multiple alignments and sequence LOGOS for each of the input sequences. To analyse the function of motifs of the query sequence, the MEME results were aligned to other motifs of known biological functions in the database. Among those which have TSSs in 1 kb (for uniformity) from ATG were used in the analysis of regulatory elements. Except for ATGA2OX5 and ATGA3OX4, TSSs were identified in 1-kb regions in all genes (Table 2). ATGA2OX5 from ATGA2OX gene family was excluded from further analysis because of annotation as a pseudogene in the A. thaliana database (https://www.arabidopsis.org/index.jsp). In ATGA3OX gene family, there was no TSS found in 1 kb of ATGA3OX4, so that it was not used in the analysis. The TSSs prediction through NNPP tools included the 5′-UTRs in some of the GA oxidase gene families. Interestingly, 57.14% of ATGA2OX (ATGA2OX1, ATGA2OX2, ATGA2OX3, ATGA2OX6), 60.00% of ATGA20OX (ATGA20OX1, ATGA20OX2, ATGA20OX3), and 66.66% of ATGA3OX (ATGA3OX1, ATGA3OX2) genes in this study have TSS in the 5′-UTRs with highest prediction score at 0.8 cut-off values. In total, 60.00% of investigated GA genes have TSS in their 5′-UTR regions which is very close to ATG. Relative to the ATG, 80.00% of TSSs with highest score were located in the range of −100 bp. Most of the genes (60%) have more than one TSS in 1-kb regions, but only one TSS with highest score was considered for further analysis.

**Results**

**Identification of TSS and promoter regions of ATGA2OXs, ATGA20OXs and ATGA3OXs genes**

From NCBI and TAIR databases, the gene name, accession number, gene ID, number of base pairs, chromosomal location, gene coordinates, open reading frame (ORF), 5′ untranslated regions (5′-UTR) and 3′ untranslated regions (3′-UTR) of each sequence were retrieved (Table 1). Annotated ORFs, 5′-UTRs and 3′-UTRs of GA oxidase genes were accessed from the TAIR database (https://www.arabidopsis.org/index.jsp). In ATGA20OX5, the 5′-UTR was unspecified in ATGA20OX5.

In the present work, only the functional genes and those which have TSSs in 1 kb (for uniformity) from ATG were used in the analysis of regulatory elements.

**Table 1. Information retrieved from NCBI and TAIR database about A. thaliana GA oxidase gene families.**

| Gene name     | Accession number | Gene ID   | Number of base pairs | Exon count | Location on chromosome | Coordinates | Orientation | ORF | 5′-UTR | 3′-UTR |
|---------------|------------------|-----------|----------------------|------------|------------------------|-------------|-------------|-----|--------|--------|
| ATGA2OX1      | NC_003070.9      | 844180    | 1540                 | 3          | 1                      | 29511561–29513100 | Reverse      | 111-1329 | 1-110 | 1330-1540 |
| ATGA2OX2      | NC_003070.9      | 839983    | 2508                 | 3          | 1                      | 10537457–10540044 | Forward      | 313-2114 | 1-312 | 2115-2588 |
| ATGA2OX3      | NC_003071.7      | 818019    | 1870                 | 3          | 2                      | 14356988–14558857 | Forward      | 115-1695 | 1-114 | 1696-1870 |
| ATGA2OX4      | NC_003070.9      | 841217    | 2444                 | 3          | 1                      | 17698598–17701041 | Forward      | 58-2237  | 1-57  | 2238-2444 |
| ATGA2OX5      | NC_003074.8      | 3768767   | 1983                 | 3          | *                      | 5870869–5872851  | Forward Pseudogene | *   | *     |        |
| ATGA2OX6      | NC_003070.9      | 839508    | 2852                 | 3          | 1                      | 486800–489651    | Forward      | 165-2592 | 1-164 | 2593-2852 |
| ATGA2OX7      | NC_003070.9      | 841518    | 2366                 | 3          | 1                      | 18894973–18891838 | Forward      | 77-2247  | 1-76  | 2248-2366 |
| ATGA2OX8      | NC_003075.7      | 827868    | 4089                 | 5          | 4                      | 11302682–11306770 | Forward      | 70-3920  | 1-69  | 3921-4089 |
| ATGA2OX1      | NC_003075.7      | 828645    | 1735                 | 2          | 4                      | 12990851–12992585 | Reverse      | 177-1604 | 1-176 | 1605-1735 |
| ATGA2OX2      | NC_003076.8      | 835256    | 1703                 | 3          | 5                      | 21053179–21056881 | Reverse      | 136-1493 | 1-135 | 1494-1703 |
| ATGA2OX3      | NC_003076.8      | 830611    | 1787                 | 3          | 3                      | 2243533–2245339  | Reverse      | 183-1505 | 1-182 | 1506-1787 |
| ATGA2OX4      | NC_003070.9      | 842389    | 1568                 | 3          | 1                      | 22452573–22454140 | Forward      | 1-1568   | *     | *      |
| ATGA2OX5      | NC_003070.9      | 841012    | 1871                 | 3          | 1                      | 16760616–16762486 | Reverse      | 1-1810   | *     | 1811-1871 |
| ATGA2OX1      | NC_003070.9      | 838125    | 1955                 | 2          | 1                      | 534388–5346342  | Reverse      | 265-1774 | 1-264 | 1775-1955 |
| ATGA2OX2      | NC_003075.7      | 844374    | 1884                 | 2          | 1                      | 30200466–30202349 | Reverse      | 187-1655 | 1-186 | 1656-1884 |
| ATGA2OX3      | NC_003075.7      | 828256    | 1832                 | 3          | 4                      | 11527229–11529060 | Forward      | 1-1832   | *     | *      |
| ATGA2OX4      | NC_003070.9      | 844373    | 1477                 | 2          | 1                      | 30198061–30199537 | Reverse      | 1-1477   | *     | *      |

*Represent the absence/unannotated exon, 5′-UTR and 3′-UTR.
Figure 1. Conserved motifs distribution of MEME output in putative promoters of ATGA2OXs, ATGA20OXs and ATGA3OXs. At the bottom of the diagram, the positions of nucleotides are indicated from TSSs (+1) to the upstream (−1000) bp. The different motifs are represented by different colours.
Identification of motifs in ATGA2OXs, ATGA20OXs and ATGA3OXs genes

The MEME web server was used in the identification and analysis of motifs in 15 GA oxidase genes, and a total of 5 conserved motifs were discovered (Table 3). The identified motifs in the 5′-upstream regions are shown as coloured boxes (Figure 1). Two motifs (MfII and MfIII) were shared by most (80.00% and 73.33%) of GA genes, respectively. MfIV was common (100%) to all GA genes and its distribution was almost uniform through the promoter regions and not clustered around a specific site. The sequence logo of MfIV was indicated (Figure 2). MfI was not observed in ATGA2OXs (ATGA2OX2, ATGA2OX3, ATGA2OX4, ATGA2OX6, ATGA2OX7, ATGA2OX8), ATGA20OXs (ATGA20OX1, ATGA20OX3, ATGA20OX4) and ATGA3OXs (ATGA3OX1, ATGA3OX3). The result revealed that MfV was absent in ATGA2OXs (ATGA2OX1, ATGA2OX3, ATGA2OX4, ATGA2OX7), ATGA20OXs (ATGA20OX1, ATGA20OX3, ATGA20OX4, ATGA20OX5) and ATGA3OXs (ATGA3OX1, ATGA3OX3). Graphical analysis showed that more than half (53.33%) of the motifs were distributed throughout the promoter regions on both strands. When the distribution of motifs was compared in the last 100 (−900 to −100) bp, it covered only 20%, while in the first 100 (+1 to −100) bp it indicated 46.66%. Additionally, positional clustering of more than half (55.31%) of the motifs was found between +1 and −500 nucleotides, suggesting that most of them may be part of core promoters of associated genes.

The query motifs from MEME tools were directly submitted to TOMTOM. Among 1790 TFs in JASPAR-CORE databases, 15 TFs were identified by TOMTOM that matched the conserved motifs forwarded from MEME. However, only a Myb-related TF with the lowest p-value (1.86e−04) was found to match from plant (A. thaliana) species. The rest of the TFs were from animal species (mainly from mouse and human) and may be this calls for further analysis to investigate TFs homology between animal and plant species. Therefore, the plant database called Plant CARE was selected to analyse the cis-acting regulatory motifs in the promoter sequences to overcome the above-mentioned problems. The predicted promoter regions of each gene were submitted to the Plant CARE database. The result revealed that numerous known regulatory elements that match the identified motifs were accessed. The output of Plant CARE, name of motifs, consensus sequence and their function are presented in Table 4. Different regulatory elements with a different rate of occurrence were found in all GA gene families. However, the output of Plant CARE showed that CAAT-box and TATA-box were the most frequently identified motifs in the promoter region of all GA oxidase gene families (Figure 3).

Among the identified motifs such as ATC-motif, Gap-box, GA-motif, Sp1, G-box, Box 4, chs-CMA1a, TCT-motif, GT1-motif, LAMP-element, LS7, GATA-motif, ATCT-motif and Box II were identified as light-responsive cis-acting regulatory elements. Two gibberellin-responsive elements (GARE-motif, P-box) were identified in the 5′-regulatory regions of GA genes. Although the synergic effect of other plant growth regulators and GA is not clear, auxin-responsive cis-acting element (AuxRR-core) was found only in two genes (ATGA2OX1, ATGA20OX5). Salicylic acid-responsive element (TCA-element), abscisic acid-responsive element (ABRE), defense- and stress-responsive element (TC-rich repeats) were other regulatory elements identified in the promoter regions of GA oxidase gene families with various occurrence.

CpG islands analysis in ATGA2OXs, ATGA20OXs and ATGA3OXs genes

A complete absence of CGIs in the putative promoter sequences of all GA oxidase genes suggests that the expression of all these genes is regulated by other regulatory elements rather than CGIs. Excluding two genes (ATGA2OX7, ATGA20OX4), high density of CGIs was identified in the gene bodies of all GA oxidase genes (Table 5). The CLC Genomics Workbench

Figure 2. MEME tool output of sequence logos for common motif (MfIV) of ATGA2OXs, ATGA20OXs and ATGA3OXs.
confirmed that all of the CGIs in the body of the genes were highly clustered between ATG and +1000bp.

**Discussion**

A total of 17 GA oxidase gene sequences of *A. thaliana* were retrieved using NCBI and TAIR databases. Prediction of TSSs location for all these genes is the first and crucial step in the discovery of promoter regions as well as for further analysis of gene regulation. The existence and location of regulatory motifs in promoter regions can be analysed based on the distance they have from TSS. Thus, prediction and selection of TSS with the highest prediction score most probably tell us a region of the promoter that regulates the transcriptional activities in a given gene. Currently, several computational methods including NNPP have been used for in silico TSS prediction even though it was proposed earlier [23]. Many TSSs were located in the 5'-UTRs of the genes. The TSSs located in the 5'-UTRs demonstrated the relative position of motifs that are highly preferred as binding sites by some classes of TFs. The distribution of motifs in the 5'-UTRs around TSS preferred as binding sites by certain groups of TFs has been investigated in plants [24,25]. The experimental analysis performed on *EF1a*-A3 gene of *A. thaliana* that contains introns in the 5'-UTR showed increased gene expression levels up to 10-fold in transgenic plants [26]. There is a poor investigation of motif consensus sequences, their regulatory effect and intron regions in 5'-UTRs. Within 1 kb of the sequences, more than one TSS was identified in most genes, which is in agreement with the works that identified the location of TSSs in different organisms including *A. thaliana* [27], *S. scrofa* [28], *H. seropedicae* [29]. The MEME web server is among well-known algorithms that allow users to discover conserved short motifs in DNA. It is widely used in the discovery of novel conserved motifs in various organisms [28–30].

**Table 4.** Description of identified cis-acting regulatory elements (motifs) 5' upstream of TSSs (+1) of *A. thaliana* ATGA2OXs, ATGA2OXs and ATGA3OXs gene families. The 1 kb of cis-acting regulatory elements was analysed using the Plant CARE database.

| Motif        | Sequence       | Function                                                                 |
|--------------|----------------|--------------------------------------------------------------------------|
| ARE          | AAACCA         | cis-acting regulatory element essential for the anaerobic induction        |
| ATC-motif    | AGTAATCT       | Part of a conserved DNA module involved in light responsiveness            |
| ATCT-motif   | AATCTAATCC     | Part of a conserved DNA module involved in light responsiveness            |
| AuxRR-core   | GGTCCAT        | cis-acting regulatory element involved in auxin responsiveness            |
| Box II       | TGGTATAA       | Part of a light-responsive element                                        |
| CARR-box     | CAACCACTCC     | Common cis-acting element in promoter and enhancer regions                |
| CGTCA-motif  | CGTCA          | cis-acting regulatory element involved in the MeJA-responsiveness          |
| ERE          | ATTTTAAA       | Ethylene responsive element                                                |
| GATA-motif   | GATAGGA        | Part of a light-responsive element                                        |
| GCN4_motif   | TGAGTCA        | cis-regulatory element involved in endosperm expression                    |
| GT1-motif    | GGTAAA         | Light responsive element                                                  |
| LAMP-element | CTTATATCA      | Part of a light-responsive element                                        |
| LS7          | CAGATTATTTT    | Part of a light-responsive element                                        |
| LTR          | CCGAAA         | cis-acting element involved in low-temperature responsiveness             |
| MYB          | TAAACA         | cis-acting regulatory element involved in the regulation of drought inducible gene expression |
| MYB-like sequence | TAAACA     | Abiotic stress response                                                   |
| Myb-binding site | CAACAG     | Biotic and abiotic stresses response                                       |
| O2-site      | TGGACG(T/G)/A/G| cis-acting regulatory element involved in zein metabolism regulation     |
| TATA-box     | TATA           | Core promoter element around –30 of transcription start                   |
| TC-rich repeats | GTTTTCTTAC  | cis-acting element involved in defense and stress responsiveness          |
| TC-motif     | TCTTAC         | Part of a light-responsive element                                        |
| TGA-element  | AACGAC         | Auxin-responsive element                                                  |
| TGACG-motif  | TGACG          | cis-acting regulatory element involved in the MeJA-responsiveness          |
| chs-CMA1a    | TTACTTAA       | Part of a light-responsive element                                        |
| ABRE         | AGCTGT         | cis-acting element involved in the abscisic acid responsiveness           |
| Box 4        | ATTAAT         | Part of a conserved DNA module involved in light responsiveness           |
| G-box        | TACGTG         | cis-acting regulatory element involved in light responsiveness            |
| MYC          | CAAATG         | cis-acting element involved in response to drought and abscisic acid induction |
| Sp1          | GGGCGG         | Light responsive element                                                  |
| At-rich element | ATGAAATCAA  | Binding site of AT-rich DNA binding protein (ATBP-1)                      |
| P-box        | CTTTTG         | Gibberellin-responsive element                                            |
| TCA-element  | CCACTTTTTT     | cis-acting element involved in salicylic acid responsiveness             |
| GA-motif     | ATAGATAA       | Part of a light-responsive element                                        |
| W box        | TTGACC         | cis-acting regulatory element involved in direct fungal elicitor stimulated transcription of defense genes and activation of genes involved in response to wounding |
| Gap-box      | CAATGAA(A/G)   | Part of a light-responsive element                                        |
| HD-Zip 1     | CAATTA/TGATGG  | Element involved in differentiation of the palisade mesophyll cells       |
| GARE-motif   | TCGTGG         | Gibberellin-responsive element                                            |
| RY-element   | CATGCAATG      | cis-acting regulatory element involved in seed-specific regulation        |
In the present study, more than half of the motifs were distributed throughout the promoter regions on both strands. Differential distributions among motifs in GA oxidase genes have demonstrated the functional divergence in the course of evolution. This is in agreement with a previous study that reported 61 motifs in GA oxidase genes, and concluded that the distribution of motifs was different among genes [30]. However, the positional clustering of most of the motifs in GA oxidases was very close to TSS. This suggests that most of them may be part of core promoters of associated genes. Core promoters are the 5′-upstream regions immediate to TSS and contain regulatory motifs which serve as binding sites for different TFs.

The query motifs from MEME tools were directly submitted to TOMTOM, an algorithm that searches a database for motif-motif similarity at statistically significant $P$-values [22]. Statistically, the smaller $P$-value indicates that the probability of query motifs matching with known motifs in the database is not due to chance alone. TOMTOM was researched in JASPAR-CORE database and identified many TFs from unrelated species (mouse and human) with only a Myb-related TF with the lowest $p$-value from plant (A. thaliana).
species. This indicates that the collection of JASPAR-CORE database is poor in its content of plant regulatory elements, but is highly relevant for vertebrates. As an alternative option to TOMTOM, the Plant CARE (plant database) was used for motif–motif matching between the query motifs and known motifs of plant promoters in the database. It resulted in large sizes of matching index and suggested that the conserved motifs among plant species serve as binding sites for specific or common TFs. TFs that bind to the same conserved motif or to a specific motif serve as transcriptional regulators during various developmental stages in different cells under diverse environmental situations.

Even though different regulatory elements with different rates of occurrence were found, CAAT-box and TATA-box were the most frequently identified motifs in the promoter region of all GA oxidase gene families. This suggests that the three GA gene families have common TFs that bind to these motifs and are involved in transcriptional regulation. TATA-box is a conserved motif among eukaryotes and the most frequently recognized motif of the core promoter, located approximately 30 bp upstream of TSS. In Arabidopsis, TATA-box is located around −32 bp with respect to TSS [31]. In the present study, various light-responsive cis-acting regulatory elements were identified. The existence of these light-responsive regulatory elements in the promoter of GA oxidase genes shows that the biosynthesis of GA may be regulated by light. Many of these regulatory elements have been previously identified in the promoter regions of light-regulated genes [32,33]. Some of these regulatory elements have also been observed in the promoters of potato, tomato, rice and Arabidopsis sucrose transporter gene families and suggested as they are regulated by light [34, 35]. All these and the other GA-mediated growth responses can be regulated based on cellular GA concentrations as well as regulation of GA gene families that are directly involved in the biosynthesis and catabolism of active GAs. The two identified motifs (GARE-motif, P-box) in the present work could be regulated by GA. This suggests that GAs regulate all GA genes that contain GARE and P-box in the 5′-regulatory regions. The presence of GARE-motifs in rice and Arabidopsis sucrose transporter gene families suggests that GAs may activate sucrose transport to cells during plant cellular development [35]. Mutations in the sequence of the GARE-motif caused a massive reduction in GA-induced gene expression [36].

Although the synergic effect of other plant growth regulators and GA is not clear, an auxin-responsive

| Name of putative promoter | Number of cut sites | Cut position(s) | Fragment size |
|---------------------------|---------------------|-----------------|---------------|
| ATGA20X1                  | 2                   | 454, 480        |               |
| ATGA20X2                  | No cut              |                 |               |
| ATGA20X3                  | No cut              |                 |               |
| ATGA20X4                  | 1                   | 80              | 80            |
| ATGA20X6                  | No cut              |                 |               |
| ATGA20X7                  | No cut              |                 |               |
| ATGA20X8                  | No cut              |                 |               |
| ATGA20X1                  | No cut              |                 |               |
| ATGA20X2                  | No cut              |                 |               |
| ATGA20X3                  | No cut              |                 |               |
| ATGA20X4                  | No cut              |                 |               |
| ATGA20X5                  | No cut              |                 |               |
| ATGA20X6                  | No cut              |                 |               |
| ATGA20X7                  | No cut              |                 |               |
| ATGA20X8                  | No cut              |                 |               |
| ATGA20X9                  | No cut              |                 |               |
| ATGA30X1                  | No cut              |                 |               |
| ATGA30X2                  | 1                   | 351             |               |
| ATGA30X3                  | No cut              |                 |               |

Table 5. Analysis of MspI cutting sites, fragment sizes (between 40 and 220 bp) in promoter and gene body regions of A. thaliana GA oxidase gene families.
cis-acting element (AuxRR-core) was found only in two genes. Auxins are essential plant growth regulators involved in cell division, elongation, root formation and apical dominance, and are widely used in plant tissue culture nowadays. The auxin-responsive element may have contribution in GA genes regulatory cascades. At low rates, an auxin-responsive element has been identified in OsSUT3, OsSUT4, AtSUC1, AtSUC5, AtSUC6 and AtSUC9 [35]. Salicylic acid-responsive element (TCA-element) and defense- and stress-responsive element (TC-rich repeats) were other regulatory elements identified in the promoter regions of GA oxidase gene families with various occurrence and could be involved in the regulation of GA biosynthesis or catabolism. Previously some of these regulatory elements have been investigated in various plant species [35,37,38].

The CGIs are the landing places for many TFs and most of the time are located around TSSs of genes; particularly housekeeping gene regulation is carried out through CGIs [39]. Depending on the type of genes and different organisms (for instance, vertebrates, plants), the frequency of CGIs distribution also varies, some of them cover the whole gene bodies or are located downstream of TSSs. DNA sequencing of plant genes and digestion of plant genomic DNA with methylation-sensitive restriction enzymes are among the reported techniques to check the presence of CGIs in plants [40,41]. However, there is a limited knowledge about the rate of CGI distribution in plant genomic DNA and the percentage of association with genes. A complete absence of CGIs in the putative promoter sequences of all GA oxidase genes suggests that the expression of all these genes is regulated by regulatory elements other than CGIs. This finding is in contrast with the work of Ashikawa [39], who reported that the highest (76%) CGIs were found near the 5′ ends of rice genes. Except in two genes, there was a high density of CGIs in the gene bodies of GA oxidase genes and this is in agreement with CGIs reported in rice, A. thaliana, sorghum, maize and barley [39].

Conclusions

Among plant growth regulators, GA is considered a key regulator of plant architecture and affects different aspects of plant developmental processes. Controlling the cellular concentration of GAs in cereal crops or trees is used in the development of dwarf or semi-dwarf varieties, which provide a promising activity for increasing yield and resistance to lodging. Promoter region determinations, identification of TSS, discovery of motifs and CpGs islands are the most important steps in controlling gene expression and understanding its regulatory mechanisms. The present work focused on the prediction of putative promoter regions, TSS, CpGs islands and discovery of motifs in 15 GA gene families in A. thaliana. Among the identified motifs, many were recognized as cis-acting regulatory elements involved in GA responsiveness, light responsiveness, abscisic acid responsiveness, auxin responsiveness, defense and stress responsiveness.

Disclosure statement

The authors declare that they have no competing interests

Authors’ contributions

S.T. and M.K. equally participated in designing the study, analysed the data, prepared the manuscript and approved the final manuscript.

Availability of data and materials

The datasets in this study are included within the article. The datasets used during the current study are available from the corresponding author upon reasonable request.

Funding

This work was supported by the Department of Applied Biology, School of Applied Natural Science, Adama Science and Technology University, Ethiopia.

References

[1] Hedden P, Kamiya Y. Gibberellin biosynthesis: enzymes, genes and their regulation. Annu Rev Plant Physiol Plant Mol Biol. 1997;48:431–460.
[2] Hedden P, Sponsel V. A century of gibberellin research. J Plant Growth Regul. 2015;34(4):740–760.
[3] Dijkstra C, Adams E, Bhattacharya A, et al. Over-expression of a gibberellin 2-oxidase gene from Phaseolus coccineus L. enhances gibberellin inactivation and induces dwarfism in Solanum species. Plant Cell Rep. 2008;27(3):463–470.
[4] Lo SF, Yang SY, Chen KT, et al. A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. Plant Cell. 2008;20(10):2603–2618.
[5] Busov VB, Meilan R, Pearce DW, et al. Activation tagging of a dominant gibberellin catabolism gene (GA 2-oxidase) from poplar that regulates tree stature. Plant Physiol. 2003;132(3):1283–1291.
[6] Thomas SG, Phillips AL, Hedden P. Molecular cloning and functional expression of gibberellin 2-oxidases, multifunctional enzymes involved in gibberellin deactivation. Proc Natl Acad Sci U S A. 1999;96(8):4698–4703.
[7] Hedden P, Phillips AL. Gibberellin metabolism: new insights revealed by the genes. Trends Plant Sci. 2000;5(12):523–530.
[8] Schomburg FM, Bizzell CM, Lee DJ, et al. Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants. Plant Cell. 2003;15(1):151–163.
[9] Rieu I, Ruiz-Rivero O, Fernandez-Garcia N, et al. The gibberellin biosynthetic genes AtGA20ox1 and AtGA20ox2 act, partially redundantly, to promote growth and development throughout the Arabidopsis life cycle. Plant J. 2008;53(3):488–504.
[10] Mitchum MG, Yamaguchi S, Hanada A, et al. Distinct and overlapping roles of two gibberellin 3-oxidases in Arabidopsis development. Plant J. 2006;45(5):804–818.
[11] Rani V. Computational methods to dissect cis-regulatory transcriptional networks. J Biosci. 2007;32(7):1325–1330.
[12] Chow CN, Chiang-Hsieh YF, Chien CH, et al. Delineation of condition specific cis- and trans-acting elements in plant promoters under various endo- and exogenous stimuli. BMC Genomics. 2018;19(S2):109–180.
[13] Hedden P. The genes of the green revolution. Trends Genet. 2003;19(1):5–9.
[14] Sasaki A, Ashikari M, Ueguchi-Tanaka M, et al. Green revolution: a mutant gibberellin-synthesis gene in rice. Nature. 2002;416(6882):701–702.
[15] Lam E, Chua NH. ASF-2: a factor that binds to the cauliflower mosaic virus 35s promoter and a conserved GATA motif in cab promoters. Plant Cell. 1989;1:1147–1156.
[16] Gilmartin PM, Sarokin L, Memelink J, et al. Molecular light switches for plant genes. Plant Cell. 1992;2(5):369–378.
[17] Zhang Mq. Identification of human gene core promoters: an approach of structure and function. Mol Biotechnol. 2014;56(1):38–49.
[18] Porto MS, Nobrega Pinheiro MP, Batista VGL, et al. Plant promoters: an approach of structure and function. Mol Biotechnol. 2014;56(1):38–49.
[19] Chung BYW, Simons C, Firth AE, et al. Effect of S’UTR introns on gene expression in Arabidopsis thaliana. BMC Genomics. 2006;7:120.
[20] Morton T, Petricka J, Corcoran DL, et al. Paired-end analysis of transcription start sites in Arabidopsis reveals plant-specific promoter signatures. Plant Cell. 2014;26(7):2746–2760.
[21] Dinka H, Le MT. Analysis of pig vomeronasal receptor type 1 (V1R) promoter region reveals a common promoter motif but poor CpG islands. Anim Biotechnol. 2018;29(4):293–300.
[22] Yirgu M, Kebede M. Analysis of the promoter region, motif and CpG islands in Arac family transcriptional regulator ACP92 genes of Herbaspirillum seropedicae. Adv Biosci Biotechnol. 2019;10(6):150–164.
[23] Han F, Zhu B. Evolutionary analysis of three gibberellin oxidase genes in rice, Arabidopsis, and soybean. Gene. 2011;473(1):23–35.
[24] Molina C, Grotewold E. Genome wide analysis of Arabidopsis core promoters. BMC Genomics. 2005;6(1):25.
[25] Lam E, Chua NH. ASF-2: a factor that binds to the cauliflower mosaic virus 35s promoter and a conserved GATA motif in cab promoters. Plant Cell. 1989;1:1147–1156.
[26] Kuhn C, Franceschi VR, Schulz A, et al. Macromolecular trafficking indicated by localization and turnover of sucrose transporters in enucleate sieve elements. Science. 1997;275(5304):1298–1300.
[27] Ibrahim O, Botha CEJ, Bradley G. In silico analysis of cis-acting regulatory elements in S’ regulatory regions of sucrose transporter gene families in rice (Oryza sativa ‘Japonica) and Arabidopsis thaliana. Comput Biol Chem. 2010;34(5–6):268–283.
[28] Gubler F, Jacobsen JV. Gibberellin-responsive elements in the promoter of a barley high-pl α-amylase gene. Plant Cell. 1992;4(11):1435–1441.
[29] Timmerhaus G, Hanke ST, Buchta K, et al. Prediction and validation of promoters involved in the abscisic acid response in Physcomitrella patens. Mol Plant. 2011;4(4):713–729.
[30] Riechmann JL, Heard J, Martin G, et al. Arabidopsis transcription factors: genome-wide comparative analysis of cis-regulatory elements in 5′cis-acting regulatory regions in angiosperm genes. J Mol Evol. 1992;34(3):231–245.
[31] Ashikawa I. Gene-associated CpG islands in plants as motifs aiding in prediction of eukaryotic promoters. Genomics. 1993;15(3):421–432.
[32] Gardiner-Garden M, Frommer M. Signifcant CpG rich regions in angiosperm genes. J Mol Evol. 1992;34(3):231–245.