Electronic Journal of Biotechnology 62 (2023) 1–12

Contents lists available at ScienceDirect

Electronic Journal of Biotechnology

journal homepage: www.elsevier.com/locate/ejbt

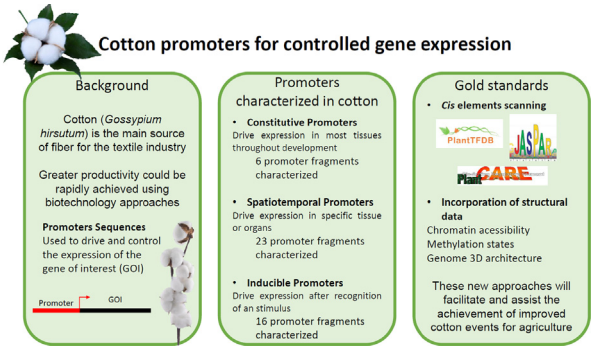
Review

Cotton promoters for controlled gene expression

Ana Luiza Atella [a,*], Maria Fatima Grossi-de-Sá [b], Marcio Alves-Ferreira [a]

[a] Departamento de Genética, Universidade Federal do Rio de Janeiro (UFRJ), Cidade Universitária, Rio de Janeiro, RJ 219410-970, Brazil
[b] Embrapa Genetic Resources and Biotechnology (EMBRAPA-CENARGEN), Brasília, DF, Brazil

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 5 July 2022
Accepted 22 December 2022
Available online 13 January 2023

Keywords:
Cis element
Constitutive promoters
Cotton
Cotton improvement
Crop
Gossypium hirsutum
Inducible promoters
Promoters
Spatiotemporal promoters

ABSTRACT

Cotton (*Gossypium hirsutum*) is one of the most important crops and is the main source of fiber for the textile industry, but its productivity is still hampered by several challenges, such as pests, diseases and abiotic stresses. This scenario has increased interest in achieving cotton events with greater productivity and sustainability through biotechnological approaches. An essential component of these strategies is the controlled expression of the gene of interest, suggesting that promoters are a key element. These promoters are generally divided into three types: constitutive, spatiotemporal, and inducible. However, to date, this diversity of promoter activity has not been as widely explored in cotton improvement. In this review, we provide an overview of cotton promoters that can be used to achieve fine-tuning of expression, facilitating decision-making and improving the ability to develop desirable traits in cotton plants. In addition, we present new approaches to identify promoters that may be useful for the development of new tools for cotton improvement.

How to cite: Atella AL, Grossi-de-Sá MF, Alves-Ferreira M. Cotton promoters for controlled gene expression. Electron J Biotechnol 2023;62. https://doi.org/10.1016/j.ejbt.2022.12.002.

Contents

Peer review under responsibility of Pontificia Universidad Católica de Valparaíso

* Corresponding author.
E-mail address: analuiza.atella@gmail.com (A.L. Atella).

https://doi.org/10.1016/j.ejbt.2022.12.002

1. Introduction

Thousands of years ago, aboriginal people discovered the outstanding properties of cotton, resulting in a unique domestication process for each of the four domesticated species of the genus *Gossypium*: two from the Americas, *G. hirsutum* and *G. barbadense*, and two from Africa-Asia, *G. arboreum* and *G. herbaceum* [1]. This history resulted in a legacy where cotton is one of the most important crops and the main source of fiber for the textile industry, as well as a source of seed oil. Domesticated cotton species are currently cultivated in more than 75 countries worldwide, and the allotetraploid species *G. hirsutum*, which is widely known as upland cotton, has dominated the field, accounting for ~90% of cotton production worldwide [2,3]. However, the broad use of cotton has resulted in new challenges, such as increasing pest pressure, new diseases, drought, salinity, and fluctuations in weather conditions due to global environmental changes.

Biotechnology has been successfully applied for cotton improvement and greater productivity. To date, more than 67 cotton GM events have been approved [4]. In these events, various candidate genes, including the widely known *Bacillus thuringiensis cry* genes, have been introduced into cotton via genetic engineering methods, enhancing the agronomic performance of cultivars [5]. In addition to the isolation and characterization of new genes, an essential step in obtaining events is the controlled transcriptional regulation of genes of interest using a specific promoter with a desirable pattern of activity. Therefore, the identification of promoters able to drive different patterns of temporal and spatial expression of transgenes is crucial for biotechnological applications.

A promoter may be divided into different regions with a modular structure. The core promoter is defined as the minimal nucleotide sequence that directs transcription initiation by RNA polymerase II, typically comprising the transcription start site (TSS) and extending up to the -35 and +35 nucleotide positions. This core promoter region contains the TATA box and other known sequences that are essential to drive basal levels of expression [6,7,8]. However, the activity of a promoter is affected by other DNA elements that may be located in proximal and distal regions. The proximal region is described as comprising the region adjacent to the TSS and extending up to -250 bases, whereas the distal region comprises sequences at greater distances [9,10]. The small modular sequences present in these regions are known as *cis* elements and regulate the spatiotemporal expression of a gene through interactions with trans-acting factor to induce expression at required sites and within a specific time frame [8,11].

These promoters are divided into different classes based on the patterns of gene expression that they drive. The most common categories are constitutive, spatiotemporal and inducible promoters [9]. In constitutive promoters, promoter activity must be detected throughout development, and most tissues are minimally affected by environmental factors. In contrast, spatiotemporal promoters show a specific pattern of activity, directing expression of the downstream gene in a particular tissue or developmental stage. Inducible promoters also function in a specific manner and are modulated by external stimuli, such as biotic and abiotic stresses and the presence of phytohormones or chemicals [12,13]. This diversity of promoters with different activity patterns allows the choice and use of promoters oriented toward target traits, which facilitates cotton improvement events. Although there are several promoters already well characterized, few of them were tested in cotton plants. Therefore, identification and the use of cotton-specific promoters is a more reliable strategy. Moreover, cotton presents unique features in crop plants, such as fiber development, highlighting the importance of the use of cotton-specific promoters to manipulate these traits. Puspito et al. [14] have reported a decrease in transgenic proteins in 74 transgenic plants, possibly due to a weak performance of promoter region. In contrast, Ribeiro et al. [15] have used a cotton endogenous promoter to drive the expression of a toxic protein, obtaining a new GM cotton with high transgene stability and resistance to the cotton boll weevil. Rathore et al. [16] have also used a cotton-derived promoter to drive the expression of an RNAi system, achieving a cotton commercial cultivar with lower gossypol content that is being commercialized in the United States since 2019 [4,16].

In this review, we summarize the knowledge of promoters already obtained from cotton (*G. hirsutum*) studies and present more robust technologies for the description of future promoters in this species. This information will contribute to better decision-making among the tools used to regulate expression with the promoters already available, consequently improving our ability to develop new cultivars with desirable traits.

2. Promoters characterized in cotton

We categorized promoters according to their activity as constitutive, spatiotemporal and inducible promoters to better summarize the content and facilitate decision-making. Information such as qPCR primers, adjacent coding sequences and the actual regulatory region isolated were analyzed using BLASTn to identify the corresponding endogenous gene in *G. hirsutum* for each promoter fragment. The regulatory sequences were retrieved from the CottonGen database and scrutinized using the PLACE database to determine plant *cis*-acting regulatory DNA elements (Table S1; Table S2) [17,18].

2.1. Constitutive promoters

Genes under the control of constitutive promoters are expected to be expressed in most tissues throughout development [9]. The cauliflower mosaic virus (CaMV) 35S promoter is the most

commonly used constitutive promoter in commercial transgenic cotton cultivars because of its high levels of transgene expression (Table S3) [19]. However, in addition to being highly used, the CaMV 35S expression profile has been the subject of debates regarding its stability in transgenic cotton cultivars [20,21,22]. When investigated in cotton plants using GFP as a reporter system, the CaMV35S promoter showed varying levels of activity among different tissues and no detectable activity in the early stages of embryogenesis [23]. Additionally, trait combinatorial approaches have increased in the crop improvement field, requiring the characterization and use of different promoters to avoid transgene silencing. Researchers have attempted to characterize new constitutive promoters with high and constant expression patterns throughout development in cotton to overcome these limitations.

To date, four native constitutive promoters have been described for *G. hirsutum* (Table 1). The first cotton constitutive promoter was described for a β-galactosidase gene (Gh_D01G1634). The promoter fragment of the gene GhGal1 (1770 bp) showed high activity in several tissues, such as flowers, trichomes, cotyledons, stems and fruits. However, leaves and roots showed less activity [24]. Viana et al. [25] identified and characterized the promoter fragment (1049 bp) of *GhGDRP85* (Gh_D11G0229), named *uceA1.7*. The *uceA1.7* activity level is similar to that of CaMV35S in leaves and stems and higher than that of CaMV35S in roots (2-fold) and floral tissues (7-fold) [25,26]. Two other promoters from genes encoding putative cinnamate 4-hydroxylases show activity similar to the 35S promoter. The cotton genes *GhC4H1* and *GhC4h2* (Gh_D10G1845 and Gh_A13G2057, respectively) are conspicuously expressed in several tissues, indicating the potential utility of their promoters [27]. Using particle bombardment, promoter fragments of 939 bp for *GhC4H1* and 1243 bp for *GhC4H2* were evaluated in cotton tissues, confirming the promoter activity in ovules and fibers that were the target tissues of bombardment. Based on the results, the authors inferred that p*GhC4H1* and p*GhC4H2* could be constitutive promoters, but *in vivo* analyses in cotton or heterologous systems are required to validate these results [27].

2.2. Spatiotemporal promoters

Spatiotemporal promoters are expected to drive expression in a specific manner, as their sites of action are either a specific organ, tissue or a stage of development. This type of promoter facilitates the expression of target genes more precisely, as expression only occurs when and where the promoter is active [9]. To date, several promoters with spatiotemporal patterns of expression have been described in cotton (Table 2).

Vegetative organs of cotton are commonly overlooked compared to reproductive structures. However, these organs are also targets for several diseases inflicted by bacteria, fungi and pests, such as lepidopterans. Therefore, the characterization of promoters with high activity in these vegetative organs is fundamental to fight against diseases [28].

Table 1
Summary of constitutive promoters characterized in cotton. Gh_code indicates the gene locus ID annotated in the cotton genome NBIv1.1. Category indicates the major category of promoter activity. Transgenic indicates the plant species in which the promoter activity was tested, and Method indicates reporter system used. Reference indicates the digital object identifier.

Gh_code	Category	Transgenic	Method	Gene	Fragment	Reference
Gh_A13G2057	Constitutive	*G. hirsutum*	GUS	*C4H2*	1243 bp	[27]
Gh_D01G1634	Constitutive	*N. tabacum*	GUS	*GhGal1*	1770 bp	[24]
Gh_D10G1845	Constitutive	*G. hirsutum*	GUS	*C4H1*	939 bp	[27]
Gh_D10G2314	Constitutive	*N. benthamiana*	GUS	*GhVTC1*	1600 bp	[58]
Gh_D11G0229	Constitutive	*A. thaliana*	GUS	*GhGDRP85*	1049 bp	[25]
Gh_D10G2314	Constitutive	*N. benthamiana*	GUS	*GhVTC1*	240 bp	[58]

Table 2
Summary of spatiotemporal promoters characterized in cotton. Gh_code indicates the gene locus ID annotated in the cotton genome NBIv1.1. Category indicates the major category of promoter activity, whereas promoter activity indicates the specific spatiotemporal activity of promoters. Transgenic indicates the plant species in which the promoter activity was tested, and Method indicates reporter system used. Reference indicates the digital object identifier.

Gh_code	Category	Promoter Activity	Transgenic	Method	Gene	Fragment	Ref.
Gh_A10G0327	Organ-specific	Trichomes and roots	*A. thaliana*	GUS	*GhCesA4*	2574 bp	[40]
Gh_A10G0567	Organ-specific	Fiber	*G. hirsutum*	GUS	*GhGDSL*	951 bp	[43]
Gh_A11G1801	Organ-specific	roots, stems and leaves	*A. thaliana*	GUS	*GhWRKY42*	1943 bp	[29]
Gh_A11G1801	Organ-specific	roots, stems and leaves	*A. thaliana*	GUS	*GhWRKY42*	1407 bp	[29]
Gh_A11G1801	Organ-specific	roots, stems and leaves	*A. thaliana*	GUS	*GhWRKY42*	778 bp	[29]
Gh_A13G1835	Organ-specific	Flower and fruit	*A. thaliana*	GUS	*GhAO-like1*	351 bp	[34]
Gh_A13G1850	Organ-specific	Seeds	*A. thaliana*	GUS	*GhFAD2-1*	1279 bp	[38]
Gh_D01G1634	Organ-specific	Stamens and stigma	*A. thaliana*	GUS	*BGal*	356 bp	[35]
Gh_D04G1457	Organ-specific	Anthers and stipules	*A. thaliana*	GUS	*GhRING1*	1000 bp	[37]
Gh_D08G2100	Organ-specific	Leaves and Roots	*N. benthamiana*	GUS	*GhAO1*	1600 bp	[30]
Gh_D08G2100	Organ-specific	Leaves and Roots	*N. benthamiana*	GUS	*GhAO1*	720 bp	[30]
Gh_D08G2407	Organ-specific	Leaves and Flowers	*A. thaliana*	GUS	*GhFT*	1000 bp	[35]
Gh_D03G0971	Organ-specific	Stamens and stigma	*A. thaliana*	GUS	*GhPL*	435 bp	[35]
Gh_D09G2288	Organ-specific	Stamens and stigma	*A. thaliana*	GUS	*PME*	300 bp	[35]
Gh_D13G1899	Organ-specific	Flower and fruit	*G. hirsutum*	GUS	*Arf*	2292 bp	[32]
Gh_A05G1347	Stage specific	Fiber elongation	*G. hirsutum*	GUS	*E6*	2640 bp	[42]
Gh_A05G1347	Stage specific	Fiber elongation	*G. hirsutum*	GUS	*E6*	1186 bp	[42]
Gh_A10G2323	Stage specific	Fiber elongation	*G. hirsutum*	GUS	*Expansin*	2192 bp	[42]
Gh_A08G0129	Stage specific	Fiber initiation, elongation and secondary cell wall synthesis	*G. hirsutum*	GUS	*LTP*	2983 bp	[42]
Gh_D10G0333	Stage specific	Fiber secondary cell wall synthesis	*G. hirsutum*	GUS	*CelA1*	2898 bp	[42]
Gh_A05G1365	Tissue-specific	Boll wall	*G. hirsutum*	GUS	*GhPRP3*	1587bp	[44]
Gh_A05G1647	Tissue-specific	Boll wall, fiber and petal tissues	*G. hirsutum*	GUS	*GhCHS1*	723 bp	[44]
Gh_A13G0613	Tissue-specific	Tapetum tissue of anthers	*G. hirsutum*	GUS	*AEG1*	1500 bp	[45]

The *GhWRKY42* gene (Gh_A11G1801) is expressed at high levels in vegetative organs and at low levels in reproductive tissues. Using a GUS activity assay, a *GhWRKY42* promoter fragment with a size of 1943 bp confirmed the expression profile, showing activity in roots, stems and leaves, but no activity in reproductive tissues. When scanned for *cis* elements in the *GhWRKY42* promoter fragment of 1 kb, we reported the occurrence of four different root-specific elements, including 12 copies of the root-specific ROOTMOTIFTAPOX1 element and five copies of NTBBF1ARROLB element, which is described as related to root and shoot specificity (Fig. 1). Different fragments of the same promoter were also isolated, and the 391 bp fragment was unable to activate expression, suggesting that critical *cis* elements are located between the 391 bp and 778 bp fragments that maintained the original activity [29]. Xin et al. [30] identified a promotor that also shows high expression in vegetative organs. An analysis of the *pGhAO1* (Gh_D08G2100) promoter fragment (1920 bp) revealed strong reporter gene expression in tobacco trichomes, leaves and roots. When *cis* element scanning was performed, the 1 kb promoter fragment contained 10 copies of the root-specific ROOTMOTIFTAPOX1 element and one stem-specific RGATAOS element (Fig. 1) [30].

Proper development of reproductive organs is necessary for fiber formation, highlighting the importance of achieving controlled transcriptional regulation within these organs. In addition, reproductive tissues may be the site of action of insect pests, such as *Anthonomus grandis*, which reduces fiber yield and leads to substantial losses in productivity [31]. Therefore, the identification of promoters specific to reproductive organs is crucial for cotton improvement [5]. Four different promoter fragments of the ADP-ribosylation factor1 gene named *Arf* (Gh_D13G1899) were isolated from cotton, displaying higher GUS activity in reproductive organs than 35S and decreased activity in leaves. In addition, the 819- and 390-bp promoter fragments differentiated themselves by presenting a more specific activity profile. These fragments showed higher activity in the bud husk, stigma, anther, boll husk and fiber, but relatively low expression in petal, ovary and ovule tissues [32]. The promoter of the cotton putative homolog of FLOWERING LOCUS T (Gh_D08G2407), a gene well-known to participate in the flowering pathway, was investigated by Sang et al. [33] using a deconstructive approach. Fragments of 1, 1.5 and 1.8 kb showed high reporter gene expression with activity present in leaves, flowers and siliques of *Arabidopsis thaliana* [33]. When scanned for *cis* elements, the 1 kb *pGhFT* promoter was shown to contain two leaf-specific S1FBOXSORPS1L21 elements, as well as several seed-specific elements and three pollen-specific GTGANTG10 elements (Fig. 1). A 351 bp promoter fragment of an ascorbate oxidase-like1 gene (Gh_A13G1835) was isolated and characterized by Lambret-Frotté et al. [34]. Using GUS histochemical assays, the authors showed that *pGhAO-like1* specifically drives GUS activity in stamens and carpels of transgenic *Arabidopsis* plants. When scanned for *cis* elements, a large number of different seed-specific *cis* elements, as well as a fiber-specific L1BOXATPDF1 element and two pollen-specific GTGANTG10 elements, were detected in the 1 kb *pGhAO-like1* fragment [34]. Artico et al. [35] characterized three promoters specific to reproductive organs:

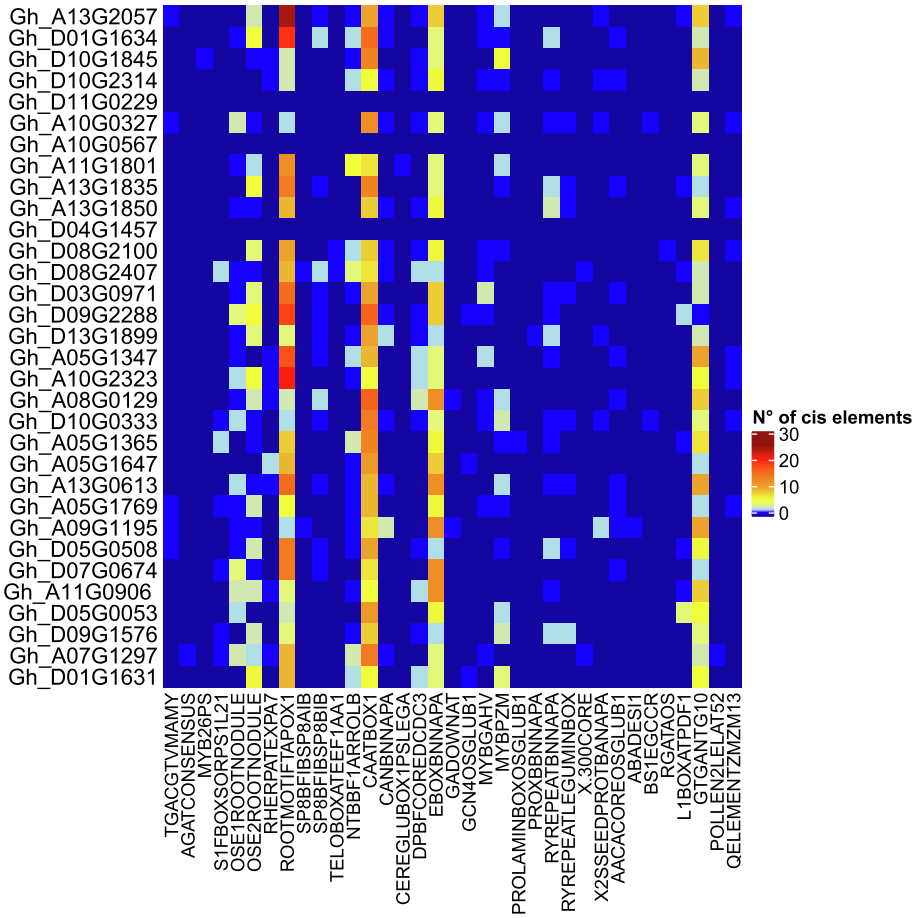


Fig. 1. Heatmap of *cis* elements related to spatiotemporal specificity present in promoters characterized in cotton. *Cis* element descriptions were retrieved from the PLACE database. Scale bar, number of *cis* elements occurring in promoters. Blue denotes a low number of *cis* element occurrences, yellow denotes a medium number, and red denotes a high number.

pGhPME-like1 (Gh_D09G2288) with a size of 300 bp, *pGhβGal-like1* (Gh_D01G1634) with a size of 356 bp and *pGhPL-like1* (Gh_D03G0971) with a size of 435 bp. For the *pGhPME-Like1* construct, activity was specifically observed in stamens and pollen grains. Activity was also detected in the style and stigma in later stages of development. *pGhβGal-like1* and *pGhPL-like1* activity was found in pollen grains, anthers, the upper part of carpels and the style [35]. Recently, the use of the 300 bp *pGhPME-like1* fragment (pFS1) driving the expression of *Cry10Aa* showed the higher accumulation of bt protein in the stamen and carpels than events in which the uce1a constitutive promoter drove *Cry10Aa* expression [36]. The results of the in planta feeding bioassays and field simulation confirmed that the higher accumulation of Cry10Aa was active in the stamen and carpels and, consequently, enhanced cotton resistance to cotton boll weevil herbivory [36]. These results highlight the importance of identifying spatiotemporal promoters in cotton. The *GhRING1* (Gh_D04G1457) promoter was isolated by Ho et al. [37]. GUS expression driven by the 1 kb fragment of *pGhRING1* was high in stipules and anthers in transgenic *Arabidopsis* plants, whereas no activity was detected in leaves or roots [37].

In addition to the identification of promoters with activity predominantly in reproductive organs, the identification of seed-specific promoters is relevant for crop improvement. Cotton seed oil is also a pivotal commodity, and its biochemical manipulation by genetic engineering might trigger novel traits, such as higher oil productivity and quality (fatty acid profile) [31]. Liu et al. [38] reported the activity of the *GhFAD2-1* (Gh_A13G1850) promoter fragment (1279 bp) as exclusive to seeds in transgenic *Arabidopsis* plants. Regarding *cis* elements, the 1 kb *pGhFAD2-1* fragment contains twenty occurrences of seed-specific elements, as well as five occurrences of pollen-specific elements [38].

Cotton fiber quality is mainly reflected in the end product, creating a high demand for biotechnological approaches aiming to improve fiber quality while also increasing fiber yields [39]. The cellulose synthase catalytic subunit 4 (*GhCesA4, Gh_A10G0327)* gene plays an important role in cellulose biosynthesis during cotton fiber development. *GhCesA4* promoter activity was investigated by Kim et al. [40] in *Arabidopsis* and cotton plants. The 2574 bp fragment shows strong GUS activity in trichomes and root tissues. Using the 1355 bp fragment, the authors showed that the GUS activity did not change in trichomes but started to be detected in vascular tissues of stems, leaves and roots. In addition, the 693 bp fragment drove GUS activity in root vascular tissues [40].

Developmental stage-specific promoters show activity in a specific phase of the life cycle of the whole plant or a specific tissue and are temporally restricted. In cotton, these promoters are usually characterized during fiber development and its stages. The development of fibers consists of four different stages: fiber initiation, fiber elongation, secondary cell wall deposition and maturation [41]. Since fiber development is a complex process, emerging manipulations for the control of fiber quality and yield require the characterization of promoters with stage-specific patterns of expression [39]. Chen and Burke [42] isolated several promoters in cotton that showed specific activity in different stages of fiber development. The activity of the *E6* (Gh_A05G1347) and *expansin gene* (Gh_A10G2323) promoter fragments was mainly detected in the early development stage, and the promoters were more active in the cell elongation phase. A shorter 1186 bp fragment of the *E6* promoter showed no difference in activity levels [42]. The authors also showed that the promoter fragment of *CelA1* (Gh_D10G0333) drove specific GUS activity in the secondary cell wall deposition stage, while *pLTP* (Gh_A08G0129) drove specific activity in the first three developmental stages [42]. GDSL lipases are known to be expressed in cotton fibers, but their roles are still unknown. The 951 bp *pGhGDSL* (Gh_A10G0567) promoter fragment was isolated and investigated in transgenic cotton by Yadav et al. [43]. The assays indicated weak activity in ovules and fibers at 0 DPA that increased gradually and peaked at 19 DPA; activity was still detected at 25 DPA but with lower intensity [43]. All promoters described drive stage-specific expression and therefore are of interest in the genetic manipulation of fiber quality [42].

Tissue-specific promoters drive spatiotemporal activity in a highly specific pattern and are exclusive to a single tissue. Lightfoot et al. [44] described one promoter that is a good representation of this attribute. The *GhPRP3* (Gh_A05G1365) and *GhCHS1* (Gh_A05G1647) promoters were isolated as fragments of 1587 bp and 723 bp, respectively, and then fused to GUS. Investigations of activity using bombardment in cotton tissues showed *pGhPRP3* activity in the boll wall, whereas *pGhCHS1* drove activity in boll wall and petal tissues [44]. Regarding the *cis* elements present in these promoters, pGhPRP3 contains several copies of the seed-specific CAATBOX1 and a fiber-specific element, whereas *pGhCHS1* contains eleven copies of the same seed-specific CAATBOX1, eight copies of the seed-specific EBOXBNNAPA and two pollen-specific elements. Paritosh et al. [45] isolated a *pAEG1* (Gh_A13G0613) promoter fragment with a size of 1500 bp and detected its activity in a tissue-specific manner. The fragment was investigated in cotton transgenic lines and showed activity exclusively in anther tapetum [45]. Regarding the *cis* elements in the 1 kb promoter fragment, *pAEG1* contains ten copies of the pollen-specific GTGANTG10 element (Fig. 1).

Last, Zhang et al. [46] identified a tissue-specific *pGhZU* with bidirectional activity. It was found to be the intergenic region between the TSSs of *Ghrack*1 (Gh_D09G1710) and *Ghuhrf1* (Gh_D09G1711). Bidirectional promoters are characterized by a genomic region of DNA that initiates transcription in both orientations [46]. Activity of this promoter in one direction (*GhZUf*) drives higher levels of expression in the tip of the leaf and apical meristem regions, whereas activity of the promoter in the other direction (*GhZUr*) results in higher expression in young tissues such as leaf epidermal hairs [47]. The use of this type of promoter may be advantageous for strategies such as gene stacking. In transgenic plants, the use of multiple copies of the same promoter to drive the expression of different genes may result in promoter silencing [48].

2.3. Inducible promoters

Inducible promoters show activity in a specific manner that requires the recognition of an external stimulus to drive the expression of downstream genes. The controllable expression driven by inducible promoters is very important for basic research and crop improvement, and these promoters are frequently used because the gene of interest can be switched off or on, avoiding energy waste or, more importantly, deleterious effects on the plant [49]. Inducible promoters described in cotton to date are divided into three categories: inducible by abiotic stresses, inducible by biotic stresses and inducible by phytohormones (Table 3).

Abiotic stresses are known to limit cotton productivity, causing approximately 73% yield loss worldwide [28]. Among the diverse abiotic stresses, drought and salinity are the two main factors affecting production; therefore, interest in describing promoters with inducible activity upon exposure to these abiotic stresses is increasing. For example, the *GhRDL1* (Gh_D05G0508) gene showed high sequence homology to *Arabidopsis thaliana* RD22 (At5g25610), a gene related to the response to dehydration [48]. Dass et al. [50] tested promoter inducibility in plants transformed with a construct containing the 302 bp *pGhRDL1* promoter fused to the reporter gene uidA. Inducible promoter activity was observed after 3 h of polyethylene glycol (PEG) treatment and continued to increase, reaching activity levels that were three-fold higher after 6 h than the activity observed in the absence of stress [50]. We reported two occurrences of *cis* MYBCORE elements related to drought and

Table 3
Summary of inducible promoters characterized in *G. hirsutum*. Gh_code indicates the gene locus ID annotated in the cotton genome NBIv1.1. Category indicates the major category of promoter activity, whereas promoter activity indicates the specific inducible activity of promoters. Transgenic indicates the plant species in which the promoter activity was tested, and Method indicates reporter system used. Reference indicates the digital object identifier.

Gh_code	Category	Promoter Activity	Transgenic	Method	Gene	Fragment	Reference
Gh_A05G1769	Abiotic Stress	Dehydration and salinity	*A. thaliana*	GUS	*GhDBP1*	1482 bp	[52]
Gh_A09G1195	Abiotic Stress	Dehydration, salinity and ABA	*N. tabacum*	GUS	*LEA D113*	1383 bp	[53]
Gh_D05G0508	Abiotic Stress	Dehydration	*N. benthamiana*	GUS	*GhRDL1*	302 bp	[50]
Gh_D07G0674	Abiotic Stress	Dehydration	*N. tabacum*	GUS	*GhRGP1*	624 bp	[51]
Gh_A11G0906	Biotic Stress	Wounding	*G. hirsutum*	GUS	*GhHB12*	905 bp	[59]
Gh_D05G0053	Biotic Stress	*M. incognita*	*N. benthamiana*	GUS	*GHNTR1*	1559 bp	[60]
Gh_D07G0674	Biotic Stress	Wounding	*N. tabacum*	GUS	*GhRGP1*	624 bp	[51]
Gh_D09G1576	Biotic Stress	*V. dahliae*	*A. thaliana*	GUS	*Ve*	2061 bp	[61]
Gh_D09G1576	Biotic Stress	*V. dahliae*	*A. thaliana*	GUS	*Ve*	979 bp	[61]
Gh_A07G1297	Phytohorme	Gibberellin	*A. thaliana*	GUS	*GhMPK11*	1309 bp	[57]
Gh_A09G1195	Phytohorme	Abscicic acid	*N. tabacum*	GUS	*LEA D113*	158 bp	[53]
Gh_A10G0327	Phytohorme	Auxin	*N. benthamiana*	GUS	*GhCesA4*	1482 bp	[55]
Gh_D01G1631	Phytohorme	Auxin	*G. hirsutum*	GUS	*GhMYB9*	1231 bp	[56]
Gh_D08G2100	Phytohorme	Auxin	*N. benthamiana*	GUS	*GhAO1*	1920 bp	[30]
Gh_D08G2100	Phytohorme	Auxin	*N. benthamiana*	GUS	*GhAO1*	1760 bp	[30]
Gh_D10G2314	Phytohorme	Ethylene	*N. benthamiana*	GUS	*GhVTC1*	1901 bp	[58]

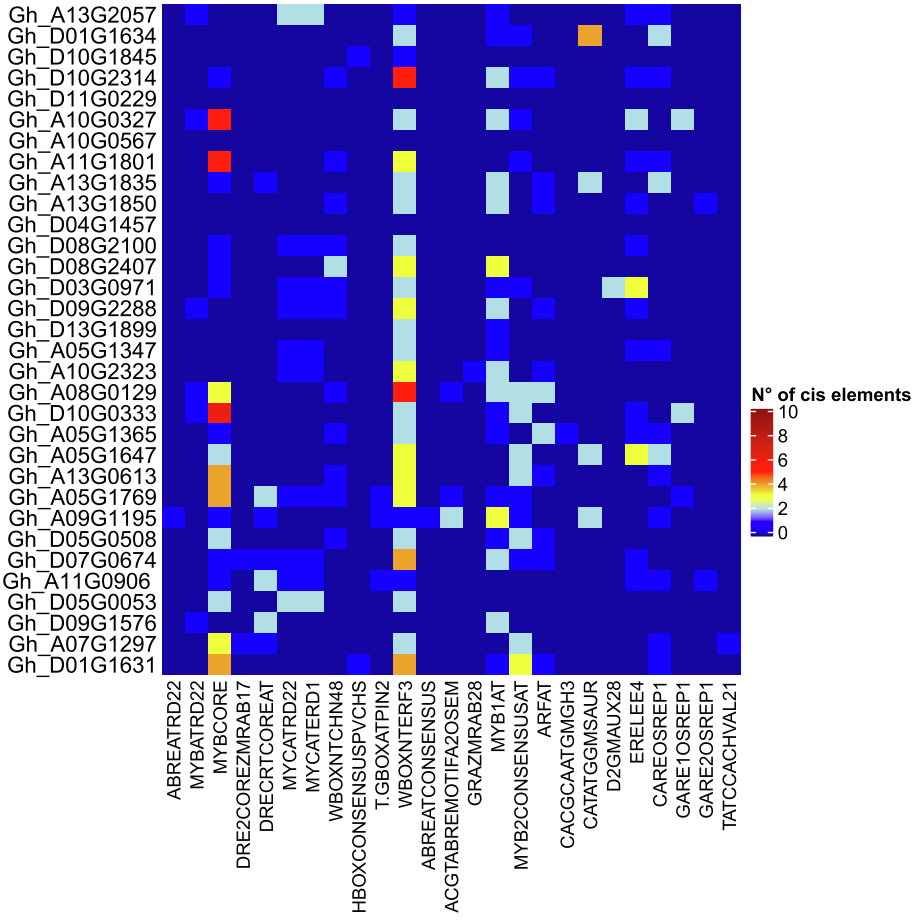


Fig. 2. Heatmap of the occurrence of *cis* elements related to inducibility in promoters characterized in cotton. *Cis* element descriptions were retrieved from the PLACE database. Scale bar, number of *cis* elements occurring in promoters. Blue denotes a low number of *cis* element occurrences, yellow denotes a medium number, and red denotes a high number.

occurrences of WBOXNTCHN48 and WBOXNTERF3 wounding response elements in the 1 kb *pGhRDL1* fragment (Fig. 2). Wu et al. [51] characterized the 624 bp promoter fragment of *GhRGP1* (Gh_D07G0674) in tobacco plants. Inducible promoter activity was observed in plants subjected to drought stress, reaching activity levels 1.76-fold higher than those in unstressed plants. In addition to drought, the *pGhRGP1* promoter also showed a 31% increase in activity in plants subjected to ABA treatment when compared to nontreated plants [51]. Regarding *cis* elements present in the 1 kb fragment of *GhRGP1*, five different *cis* elements related to the drought response and four copies of the WBOXNTERF3 *cis* element related to the wounding response were reported (Fig. 2). Dong et al. [52] identified a gene (*GhDBP1*;Gh_A05G1769) whose expression was induced in cotton by high salinity and drought. The 1482

bp *GhDBP1* promoter fragment was isolated and ligated to the uidA reporter gene and investigated in transgenic *Arabidopsis* lines using GUS assays. The authors showed a 3.9-fold increase in the activity of *GhDBP1* following treatment with NaCl. Similar results were obtained, with a 3.8-fold increase in GUS activity, 2.5-fold increase with mannitol treatment and 5.9-fold increase with drought stress [52]. We reported four occurrences of the drought-related MYB-CORE *cis* element and copies of DRECRTCOREAT, MYCATRD22 and MYCATERD1 elements, which are also related to drought stress, in the 1 kb *pGhDBP1* fragment (Fig. 2). Luo et al. [53] evaluated a late embryogenesis-abundant (LEA) gene D113 (Gh_A09G1195) promoter using a deconstructive approach. Fragments of 1383, 974, 578 and 158 bp were fused to GUS, and tobacco transgenic lines were subjected to salt and drought treatment. When subjected to drought and salt stress, plants carrying the constructs with 1383, 974 and 578 bp promoter fragments showed a high induction level (15- to 24-fold). In addition to being induced by abiotic stresses, *pGhLEA D113* is also responsive to the phytohormone ABA. A 10-fold increase was observed for the 1383 bp and 578 bp fragments [53]. Regarding *cis* element occurrence, the 1 kb promoter fragment of GhLEA D113 only contains one copy of the ABREATRD22, MYBCORE and DRECRTCOREAT drought-related *cis* elements (Fig. 2).

Phytohormones have a vital role in regulating cotton fiber development; therefore, promoters responsive to hormones may be particularly useful for the genetic manipulation of cotton [54]. In addition, promoters activated by phytohormones have a broad spectrum of potential applications to control expression, as they are active even in pathways participating in different stress responses and developmental processes. Wu and collaborators isolated a 1482 bp promoter fragment of the cellulose synthase gene *GhCesA4* (Gh_A10G0327). The constructs of the fragment promoter fused to GUS were used to transform tobacco plants. *GhCesA4* promoter activity was observed after treatment with the phytohormones NAA and 2,4-D [55]. Zhang et al. [56] isolated the promoter of *GhMYB9* (Gh_D01G1631), a gene encoding a protein from the R2R3-MYB transcription factor family. The 1487 bp *pGhMYB9* fragment was isolated, fused to GUS and used to produce transgenic cotton lines. GUS activity was increased 1.5-fold in fibers subjected to IAA treatment [56]. The 1 kb promoter fragment of *GhMYB9* contains one ARFAT auxin-responsive *cis* element. Xin and collaborators identified an ascorbate oxidase (*GhAO*1; Gh_D08G2100) gene that shows inducible transcription by auxin, promoting interest in characterizing its promoter. Different promoter fragments were fused to GUS and tested in tobacco leaves using transient expression assays. When exposed to 1 mg/L IAA, promoter fragments of 1920 and 1760 bp showed increases of 3- to 4-fold in GUS activity compared to nontreated plants [30]. Wang et al. [57] identified the *GhMPK11* (Gh_A07G1297) gene that encodes a mitogen-activated protein kinase involved in sensing and transducing internal or external signals to downstream effectors. A 1309 bp *pGhMPK11* promoter fragment was isolated, fused to GUS, and subsequently transformed into *Arabidopsis* plants. GUS activity driven by the promoter increases exclusively after exposure to GA3 [57]. The *GhVTC1* (Gh_D10G2314) promoter was investigated by Song et al. [58] using a deconstructive approach with different fragments. The 1901 and 1600 bp fragments showed increased GUS activity in tobacco leaves treated with ACC, whereas promoter fragments with sizes of 1360, 640 and 240 bp were not able to drive inducible GUS activity [58]. Regarding *cis* element occurrence, *pGhVTC1* contains *cis* elements related to several phytohormones, with three different occurrences of ABA-responsive elements and one occurrence each of auxin-, ethylene- and gibberellin-responsive elements.

Different biotic stresses, such as pests and pathogens, affect cotton production worldwide. Historically, cotton crops are known to require large amounts of and more frequent pesticide applications to maintain their productivity. Nevertheless, disease losses account for approximately 11% of US cotton production annually [28]. Therefore, insect- and disease-resistant cotton events are of high interest, requiring knowledge of promoters inducible by biotic stimuli that might help to produce resistant plants. Two promoters were characterized recently by He et al. [59] that are activated by wounding in cotton. The GhHB12 (Gh_A11G0906) gene encodes an HD-ZI class I transcription factor that has been previously described as being induced by stresses. A 905 bp *pGhHB12* promoter fragment was fused with the GUS reporter gene and transformed into cotton. In these plants, inducible GUS activity was detected upon wounding stress [59]. When scanned for *cis* element occurrence, the 1 kb fragment of *pGhHB12* was shown to contain T/GBOXATPIN2 and WBOXNTERF3 wounding-related *cis* elements. Wu et al. [51] identified a cotton promoter, *pGhRGP1* (Gh_D07G0674), that, in addition to being drought-inducible (see above), is activated by wounding [52]. To date, only two cotton promoters have been identified as pathogen-inducible promoters in cotton. *GhNTR1* (Gh_D05G0053) was investigated as a putative root-knot nematode (*Meloidogyne incognita*) resistance gene. A 1559 bp promoter fragment was isolated from the upstream region of *GhNTR1*, fused to the GUS reporter gene and transformed into transgenic *Arabidopsis*. GUS activity was detected in most tissues in later developmental stages and showed an inducible pattern when subjected to nematode infection [60]. The *GhVe* (Gh_D09G1576) gene is a *Verticillium* wilt resistance gene described previously by Liu et al. [61]. Promoter fragments with sizes of 2061 and 979 bp were fused to GUS. *Arabidopsis* transgenic lines expressing these fragments fused to GUS were infected with *Verticillium dahlia,* which triggered the activity of both promoter fragments [61].

3. Strategies used to characterize cotton promoters

Despite its importance, the identification and characterization of promoters have not received much attention in cotton improvement. Therefore, most genetically modified cotton events that have been commercialized carry promoters from other organisms [4] (Table S3). This fact brings attention to the necessity of additional studies focusing primarily on cotton promoter identification and characterization in cotton.

Several authors have opted to use previous studies on global expression analysis as a source of information to identify promoters useful for cotton biotechnology improvement. Artico et al. [35] used *G. raimondii* cDNA libraries comprising bud and boll, the single transcriptome data that were available online at that moment. These data were then mined to select flower-specific genes. After this step, the putative homolog of *G. hirsutum* was identified by sequence comparison, and its expression was evaluated using qPCR analysis. Three new promoters specific to reproductive organs were identified in a single effort; one of them, which is mentioned above (*pGhPME-like1;* named PFS1), was successfully used to drive the expression of Cry10Aa and confer cotton resistance to the boll weevil. Yadav et al. [43] used microarray data from different developmental stages of cotton to filter genes expressed preferentially at the secondary cell wall deposition stage in fibers, also obtaining the desired highly specific pattern of promoter activity. Similarly, Paritosh et al. [45] used information obtained from microarray experiments to select candidates with specific expression profiles, resulting in the characterization of a promoter with a highly specific pattern of activity in anther tapetum. These examples highlight the strength of analyzing previous expression data as a tool to infer possible promoter activity patterns, select better candidates for experimental investigation and promote better targeting of efforts.

Strategies based on mining global expression data have also been used for several plant species, including crops such as maize, rice, poplar and peanut, achieving specific patterns of promoter activity [62,63,64]. *Gossypium spp.* genomes are now available in the CottonFGD database and cover a variety of biological processes, such as stress responses and developmental series, as well as multiple tissues [65]. We identified each gene mentioned in the review in the CottonFGD database and evaluated their expression profiles to validate the utility of the data (Fig. 3, Table S4). For most of the genes, the RNA-seq expression profile was consistent with the data from promoter characterization assays *in vivo*.

For instance, the late embryogenesis-abundant (LEA) gene (Gh_A09G1195) showed the highest expression levels in stress condition libraries, which included mostly drought- and salinity-treated samples. Its promoter activity (D113) was originally described as inducible by drought and salinity [53]. Therefore, the expression data may be a useful tool to infer patterns of gene expression and, consequently, novel promoters useful for cotton biotechnology.

However, some cases of incongruences between the description of promoters and the available expression data for their respective downstream genes in *G. hirsutum* have been noted. For example, the *GhMYB9* (Gh_D01G1631) promoter directed high GUS expression to fibers and flowers of transgenic cotton [56]. The *GhMYB9* expression profile obtained using RNAseq showed low levels of constitutive expression in all libraries available (Fig. 3, Table S4). These discrepancies may be attributed to multiple factors, but the fact that the isolated promoter fragment evaluated may not contain all *cis* elements that are important for gene expression may be the most relevant factor. *Cis* elements are present 1 kb upstream of the start codon and at larger distances, still contributing to gene expression. Thus, since transcriptional regulation depends on the type, number, position, and combination of *cis* elements, fragmentation of the whole sequence context in which this promoter is inserted may result in experimental promoter activity that differs from that observed in planta [11,49].

As next-generation sequencing technologies have reduced the cost of sequencing while increasing throughput and speed, the possibility of answering complex questions in plant biotechnology has increased [66]. However, the design and understanding of RNA-seq experiments demand attention, as artifacts have been reported [67,68]. Therefore, variability may occur at both the biological and technical levels. Biological variability may arise from neglect of standardization when collecting plant material. In contrast, technical variability arises even at the same biological unit due to experimental processes, such as variations in lane, flow cell or library preparation. Therefore, randomization and replication avoid confounding of biological results with systematic experimental effects and are crucial aspects to ensure reproducibility and accuracy of the interpretation of results from an RNA-seq experiment [69,70,71].

These limitations have not prevented the successful use of expression data, as mentioned above [62,63,64]. In addition, the development of new techniques has improved the power of RNA-seq in determining expression profiles in plants and, consequently, the prediction of promoter activity.

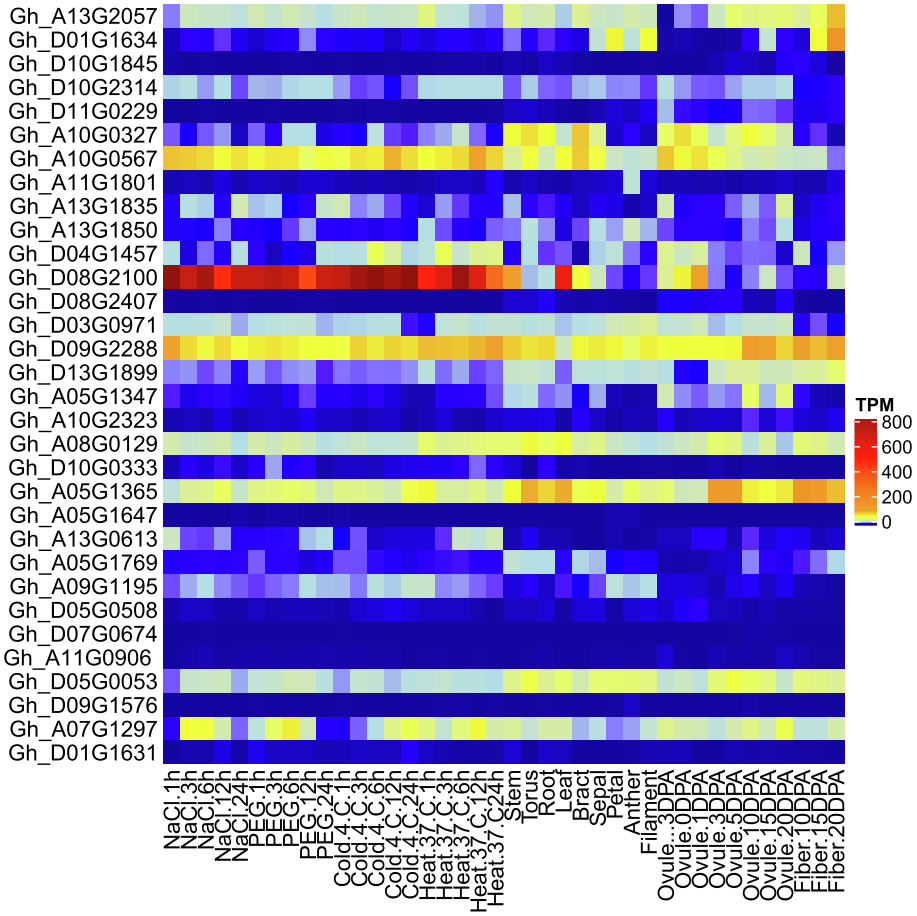


Fig. 3. Heatmap of the expression profiles of 33 genes with characterized promoters in cotton. The RNA-seq data were downloaded from CottonFGD and comprised 40 libraries that included different tissues and organs, fiber development, ovule development, drought stress, cold stress, heat stress and salt stress. Scale bar, expression levels reported as TPM values. Blue denotes a low level of expression, yellow denotes medium levels of expression, and red denotes high levels.

For instance, single-cell genomics has emerged as a powerful source of information. The emerging technology of single-cell RNA sequencing enables higher resolution measurements, revealing cell-to-cell variation that is masked in conventional bulk sequencing and enabling researchers to obtain a detailed understanding of complex biological processes such as transcriptional regulation [72]. This technique has already been extensively applied in different plant species, showing its power in elucidating new transcriptional landscapes [73,74]. For example, cell-specific profiling in *Arabidopsis* and rice has shown that both abiotic and biotic stresses lead to dramatically different responses in various cell types. However, common bulk sequencing analysis diluted plant response signals and overlooked this cell-specific variation [75,76,77]. Cotton fibers have also served as an early single-cell model because of their easy isolation before modern single-cell strategies were developed. Using a single-cell approach to investigate gene expression profiles in cotton fiber and various *Arabidopsis* tissues, the xylem of *Arabidopsis* was identified as the best model for secondary wall cellulose synthesis in cotton fibers in addition to being morphologically recognized as leaf trichomes, reinforcing the potential of this technique to identify novel expression profiles even in cotton [78]. Nonetheless, the evolution of this technology is expected to play an important role in providing a better understanding of transcriptional regulatory mechanisms and therefore facilitate promoter characterization [79].

4. Gold standard strategies to characterize cotton promoters

An important step in overcoming the challenge of understanding the mechanisms of gene regulation is the ability to identify *cis* regulatory elements. In recent years, due to its high importance in transcriptional regulation, several tools have been developed to explore the *cis* element landscape. Several databases are available for *cis* element deposition along with predictive models to scan occurrences of described *cis* elements in new promoters [17,80,81,82]. These tools have often been used to predict possible patterns of activity and play a crucial role in promoter mining, definition of which promoter fragment will be used in *in vivo* analysis and further characterization of these fragments [83,84,85].

Among the available studies describing cotton promoters, descriptions of the occurrences of *cis* elements are a crucial part of the knowledge obtained. However, *cis* element scanning of these promoters is often performed without regard to their parameters and how these parameters may affect the interpretation of results. Most papers reviewed have utilized sequence-based models where a *cis* element is represented by an invariable nucleotide sequence that must match exactly within promoter sequences to be considered an occurrence. However, transcription factors have differences in binding affinity for all of their potential binding sites, and this information is necessary for a better understanding of the regulatory network. *Cis* element scanning models that consider differential binding affinity were developed through the popularization of Chip-seq experiments that provided information on the entire set of potential binding sites and allowed the complete specificity to be estimated [86].

The position weight matrix (PWM) is a model that utilizes differential binding affinity to represent *cis* elements. In this model, a matrix represents the frequency of each nucleotide at each position of the motif, which also includes differential binding affinity patterns of transcription factors. Based on the PWM, a score can be calculated for any sequence corresponding to the sum of all values at each position [86]. However, standard PWM models usually assume that each position within the matrix contributes independently to the affinity score and are unable to represent cases of dependence between positions that have been previously observed. After considering this property, models have increased in complexity with respect to dependencies and generally provide better results but require a higher expertise for use [84].

Recent studies have also incorporated structural data that, in addition to *cis* element identification, is able to improve the predictions of functionality of these *cis* elements. These structural analyses rely on information regarding chromatin states, methylation states, genome 3D structures, and interactions between transcription factors. Regarding the dynamic chromatin environment, studies of *cis* elements have shown that transcription factors preferentially bind to elements in accessible chromatin regions containing a reduced number of nucleosomes, and only a small portion of pioneer factors are able to bind even when nucleosomes are present [87,88]. In addition, several types of histone modifications are possible within promoter regions, resulting in active or repressive marks corresponding to transcriptionally active and inactive chromatin, respectively [89]. In cotton, events generated by targeted sequence insertion presented transcriptional gene silencing through differential DNA methylation of the transgene promoter as a source of variation in expression patterns for the gene of interest, highlighting the effect of epigenetic modifications [90]. Additionally, the 3D architecture of chromosomes in the nucleus enables long-range interactions within topologically associating domains, facilitating interactions between distant *cis* elements that are established by architectural proteins [91,92]. Investigations of the 3D genome architecture in cotton highlighted promoter-centered interactions, suggesting that a large number of genes are regulated by different regulatory elements over long distances and describing new regulatory levels of gene expression in cotton species [93]. Thus, these contexts must be considered to achieve a more precise prediction of functional *cis* elements in promoters and encourage a better understanding of the landscape of transcriptional regulatory mechanisms [84]. The high throughput of information for transcriptional patterns and occurrences of *cis* elements in promoter architecture enable the more rapid description of promoters. However, functional characterization of the isolated promoter fragment is still an intrinsic process of promoter description. Different gene expression validation systems are available, and the correct interpretation of their results depends on the expression system used and its particular features.

Transient expression systems are commonly the most commonly used strategies to assay promoter activity because of their speed and suitability. Protoplasts provide a cell system that is easily manipulated by a broad range of treatments, such as light, temperature, hormones, metabolites and pathogen-derived elicitors. However, spatiotemporal analysis of promoter activity patterns cannot be assessed by transient expression in protoplasts [94,95]. The biolistic bombardment of plants with the construct of interest is also widely used [35,44]. In contrast to protoplasts, biolistics may target different tissues, facilitating the investigation of promoter activity in a spatiotemporal manner [96]. However, transient expression often leads to high initial levels of gene expression compared to stable expression. This effect may occur due to the introduction of a larger amount of DNA and transient extrachromosomal expression, in which the subject is not submitted to a chromatin structure context and therefore does not reflect its full functionality [97]. In contrast to transient expression, stable transformation is time-consuming but provides the most robust information on promoter function. Last, heterologous systems for validating promoter activity may also lead to artifacts. Model plants such as *Arabidopsis* and tobacco are often used for promoter characterization; however, the phylogenetic distance to *G. hirsutum* and the absence of specific anatomical structures such as fibers in these plants may affect results for activity patterns of spatiotemporal promoters. On the other hand, the data from heterologous species might reinforce the relevance of a promoter as a

biotechnology tool if they coincide with observed cotton promoter activity. Therefore, the use of cotton as a host and heterologous species when the promoters are intended for biotechnological applications is strongly recommended.

5. Conclusions and future perspectives

Several cotton promoters with potential for biotechnological application have been characterized in the last decade. As an example of these promoters' potential, the *UceA1.7* promoter fragment derived from the GhGDRP85 gene improves cotton boll weevil resistance in cotton plants, and the fiber-specific promoter from the *E6* gene was applied as a reporter gene to elucidate the effect of abiotic stresses on fiber development [36,42]. In addition, the wide use of high-throughput RNA-seq combined with new technologies, such as single-cell genomics and analyses of chromatin states, methylation states, nucleosome positioning, genome 3D structures, and interactions between transcription factors, will substantially increase the availability of information in this field and quickly improve the knowledge of the mechanisms of gene regulation by promoters. The high-resolution characterization of cotton promoters will enable the identification of shorter, high-activity promoters for distinct uses in biotechnology. However, naturally occurring promoters may still contain undesirable properties. The logical method will be synthetic biology using *cis* elements as building blocks to design a new synthetic promoter [98]. Synthetic switches have also been proposed as a fast approach for obtaining controllable gene expression. These genetic switches are generally composed of a sensor domain that detects signals within a cell and an actuator domain that regulates gene expression and may be constructed with several different components in various ways, providing higher personalization of expression [99]. Similar to synthetic approaches, new strategies for *in vivo* promoter engineering using sequence-specific endonucleases have been proposed as a fast method for crop improvement. This strategy shows the rapid success of manipulating expression through *cis* element-targeted editing and does not involve the addition of a transgene, promoting public acceptance, and it is not subjected to OGM regulation in several countries [100]. These new approaches will facilitate and assist the achievement of improved cotton events for agriculture.

Author contributions

- Data collection and draft manuscript preparation: AL Atella; M Alves-Ferreir.
- Revision of the results and approval of the final version of the manuscript: MF Grossi-de-Sá; M Alves-Ferreira.

Financial support

This research was supported by the Foundation for Research Assistance of the Rio de Janeiro State (FAPERJ) and the National Council for Scientific and Technological Development (CNPq).

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary material

https://doi.org/10.1016/j.ejbt.2022.12.002.

References

[1] Lee JA, Fang DD. Cotton as a World Crop: Origin, History, and Current Status. In: Fang DD, Percy RG, editors. Cotton 2015;57:1–23. https://doi.org/10.2134/agronmonogr57.2013.0019
[2] Wendel JF, Cronn RC. Polyploidy and the evolutionary history of cotton. Advances in Agronomy 2003;78:139–86. https://doi.org/10.1016/S0065-2113(02)78004-8.
[3] Wendel JF, Grover CE. Taxonomy and Evolution of the Cotton Genus, *Gossypium*. In: Fang DD, Percy RG, editors. Cotton 2015;57:25–44. https://doi.org/10.2134/agronmonogr57.2013.0020.
[4] International Service for the Acquisition of Agribiotech Applications (ISAAA). GM Approval Database - Cotton 2021. [cited February 16, 2022]. https://www.isaaa.org/gmapprovaldatabase/crop/default.asp?CropID=7&Crop=Cotton.
[5] Chakravarthy VSK, Reddy TP, Reddy VD, et al. Current status of genetic engineering in cotton (*Gossypium hirsutum* L): an assessment. Crit Rev Biotechnol 2012;34(2):144–60. https://doi.org/10.3109/07388551.2012.743502. PMid: 23190258.
[6] Butler JEF, Kadonaga JT. The RNA polymerase II core promoter : a key component in the regulation of gene expression. Genes Dev 2002;16:2583–92. https://doi.org/10.1101/gad.1026202. PMid: 12381658.
[7] Juven-Gershon T, Kadonaga JT. Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol 2009;339(2):225–9. https://doi.org/10.1016/j.ydbio.2009.08.009. PMid: 19682982.
[8] Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. Nat Rev Mol Cell Biol 2018;19:621–37. https://doi.org/10.1038/s41580-018-0028-8. PMid: 29946135.
[9] Porto MS, Pinheiro MPN, Batista VGL, et al. Plant promoters: An approach of structure and function. Mol Biotechnol 2014;56:38–49. https://doi.org/10.1007/s12033-013-9713-1. PMid: 24122284.
[10] Biłas R, Szafran K, Hnatuszko-Konka K, et al. *Cis*-regulatory elements used to control gene expression in plants. Plant Cell Tissue Organ Cult 2016;127:269–87. https://doi.org/10.1007/s11240-016-1057-7.
[11] Vedel V, Scotti I. Promoting the promoter. Plant Sci 2011;180(2):182–9. https://doi.org/10.1016/j.plantsci.2010.09.009. PMid: 21421359.
[12] Mehrotra R, Gupta G, Sethi R, et al. Designer promoter: An artwork of *cis* engineering. Plant Mol Biol 2011;75:527–36. https://doi.org/10.1007/s11103-011-9755-3. PMid: 21327513.
[13] Kummari D, Palakolanu SR, Kishor PBK, et al. An update and perspectives on the use of promoters in plant genetic engineering. J Biosci 2020;45:119. https://doi.org/10.1007/s12038-020-00087-6. PMid: 33097676.
[14] Puspito AN, Rao AQ, Hafeez MN, et al. Transformation and evaluation of Cry1Ac+Cry2A and *GTGene* in *Gossypium hirsutum* L. Front Plant Sci 2015;6:943. https://doi.org/10.3389/fpls.2015.00943. PMid: 26617613.
[15] Ribeiro TP, Arraes FBM, Lourenço-Tessutti IT, et al. Transgenic cotton expressing Cry10Aa toxin confers high resistance to the cotton boll weevil. Plant Biotechnol J 2017;15(8):997–1009. https://doi.org/10.1111/pbi.12694. PMid: 28081289.
[16] Rathore KS, Sundaram S, Sunilkumar G, et al. Ultra-low gossypol cottonseed: Generational stability of the seed-specific, RNAi-mediated phenotype and resumption of terpenoid profile following seed germination. Plant Biotechnol J 2012;10(2):174–83. https://doi.org/10.1111/j.1467-7652.2011.00652.x. PMid: 21902797.
[17] Higo K, Ugawa Y, Iwamoto M, et al. PLACE : a database of plant *cis*-acting regulatory DNA elements. Nucleic Acids Res 1998;26(1):358–9. https://doi.org/10.1093/nar/26.1.358. PMid: 9399873.
[18] Yu J, Jung S, Cheng CH, et al. CottonGen: A genomics, genetics and breeding database for cotton research. Nucleic Acids Res 2014;42(D1):1229–36. https://doi.org/10.1093/nar/gkt1064. PMid: 24203703.
[19] Odell JT, Nagy F, Chua NH. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature 1985;313:810–2. https://doi.org/10.1038/313810a0. PMid: 3974711.
[20] Benfey PN, Ren L, Chua NH. Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. EMBO J 1990;9:1677–84. https://doi.org/10.1002/j.1460-2075.1990.tb08291.x. PMid: 2347301.
[21] Terada R, Shimamoto K. Expression of CaMV35S-GUS gene in transgenic rice plants. MGG Mol Gen Genet 1990;220:389–92. https://doi.org/10.1007/BF00391743.
[22] Yang N-S, Christou P. Cell type specific expression of a CaMV 35S-GUS gene in transgenic soybean plants. Dev Genet 1990;11(4):289–93. https://doi.org/10.1002/dvg.1020110407.
[23] Sunilkumar G, Mohr LA, Lopata-Finch E, et al. Developmental and tissue-specific expression of CaMV 35S promoter in cotton as revealed by GFP. Plant Mol Biol 2002;50:463–79. https://doi.org/10.1023/A:1019832123444. PMid: 12369622.
[24] Wu A, Liu J. Isolation of the promoter of a cotton β-galactosidase gene (*GhGal1*) and its expression in transgenic tobacco plants. Sci China, Ser C Life Sci 2006;49(2):105–14. https://doi.org/10.1007/s11427-006-0105-7. PMid: 16704113.
[25] Viana AAB, Fragoso RR, Guimarães LM, et al. Isolation and functional characterization of a cotton ubiquitination-related promoter and 5'UTR that drives high levels of expression in root and flower tissues. BMC Biotechnol 2011;11:115. https://doi.org/10.1186/1472-6750-11-115. PMid: 22115195.

[26] Basso MF, Lourenço-Tessutti IT, Busanello C, et al. Insights obtained using different modules of the cotton *uceA1.7* promoter. Planta 2020;251:56. https://doi.org/10.1007/s00425-020-03348-8. PMid: 32006110.
[27] Ni Z, Li B, Neumann MP, et al. Isolation and expression analysis of two genes encoding cinnamate 4-hydroxylase from cotton (*Gossypium hirsutum*). J Integr Agric 2014;13(10):2102–12. https://doi.org/10.1016/S2095-3119(13)60643-7.
[28] Paterson AH. Genetics and Genomics of Cotton. 2009. Springer New York, NY. 518 p. ISBN 978-0-387-70809-6. https://doi.org/10.1007/978-0-387-70810-2.
[29] Gu L, Wei H, Wang H, et al. Characterization and functional analysis of *GhWRKY42*, a group IId WRKY gene, in upland cotton (*Gossypium hirsutum* L.). BMC Genet 2018;19:48. https://doi.org/10.1186/s12863-018-0653-4. PMid: 30060731.
[30] Xin S, Tao C, Li H. Cloning and functional analysis of the promoter of an *Ascorbate oxidase* gene from *Gossypium hirsutum*. PLoS One 2016;11(9): e0161695. https://doi.org/10.1371/journal.pone.0161695. PMid: 27597995.
[31] Ahmad S, Hasanuzzaman M. Cotton Production and uses. Agronomy, Crop Protection, and Postharvest Technologies. Springer Singapore; 2020, 641 p. ISBN 978-981-15-1471-5. https://doi.org/10.1007/978-981-15-1472-2.
[32] Ren M, Chen Q, Li L, et al. Functional analysis of a reproductive organ predominant expressing promoter in cotton plants. Sci China, Ser C Life Sci 2005;48:452–9. https://doi.org/10.1360/062004-71. PMid: 16315596.
[33] Sang N, Cai D, Li C, et al. Characterization and activity analyses of the *FLOWERING LOCUS T* promoter in *Gossypium hirsutum*. Int J Mol Sci 2019;20:1–14. https://doi.org/10.3390/ijms20194769. PMid: 31561427.
[34] Lambret-Frotté J, Artico S, Muniz Nardeli S, et al. Promoter isolation and characterization of *GhAO-like1*, a *Gossypium hirsutum* gene similar to multicopper oxidases that is highly expressed in reproductive organs. Genome 2016;59(1):23–36. https://doi.org/10.1139/gen-2015-0055. PMid: 26692462.
[35] Artico S, Lambret-Frotté J, Nardeli SM, et al. Isolation and characterization of three new promoters from *Gossypium hirsutum* that show high activity in reproductive tissues. Plant Mol Biol Report 2013;32:630–43. https://doi.org/10.1007/s11105-013-0674-0.
[36] Ribeiro TP, Basso MF, de Carvalho MH, et al. Stability and tissue-specific Cry10Aa overexpression improves cotton resistance to the cotton boll weevil. Biotechnol Res Innov 2019;3(1):27–41. https://doi.org/10.1016/j.biori.2019.12.003.
[37] Ho MH, Saha S, Jenkins JN, et al. Characterization and promoter analysis of a cotton RING-type ubiquitin ligase (E3) gene. Mol Biotechnol 2010;46:140–8. https://doi.org/10.1007/s12033-010-9280-7. PMid: 20376576.
[38] Liu F, Zhao YP, Wang X, et al. Isolation and functional characterization of seed-specific *FAD2-1* promoter from cotton (*Gossypium hirsutum* L). J Plant Biochem Biotechnol 2015;24:369–75. https://doi.org/10.1007/s13562-014-0284-4.
[39] Mansoor S, Paterson AH. Genomes for jeans: Cotton genomics for engineering superior fiber. Trends Biotechnol 2012;30(10):521–7. https://doi.org/10.1016/j.tibtech.2012.06.003. PMid: 22831638.
[40] Kim HJ, Murai N, Fang DD, et al. Functional analysis of *Gossypium hirsutum* cellulose synthase catalytic subunit 4 promoter in transgenic *Arabidopsis* and cotton tissues. Plant Sci 2011;180(2):323–32. https://doi.org/10.1016/j.plantsci.2010.10.003. PMid: 21421377.
[41] Lee JJ, Woodward AW, Chen ZJ. Gene expression changes and early events in cotton fibre development. Ann Bot 2007;100(7):1391–401. https://doi.org/10.1093/aob/mcm232. PMid: 17905721.
[42] Chen J, Burke JJ. Developing fiber specific promoter-reporter transgenic lines to study the effect of abiotic stresses on fiber development in cotton. PLoS One 2015;10(6):e01298707. https://doi.org/10.1371/journal.pone.0129870. PMid: 26030401.
[43] Yadav VK, Yadav VK, Pant P, et al. GhMYB1 regulates SCW stage-specific expression of the GhGDSL promoter in the fibers of *Gossypium hirsutum* L. Plant Biotechnol J 2017;15(9):1163–74. https://doi.org/10.1111/pbi.12706. PMid: 28182326.
[44] Lightfoot DJ, Orford SJ, Timmis JN. Identification and characterisation of cotton boll wall-specific gene promoters for future transgenic cotton varieties. Plant Mol Biol Report 2013;31:174–84. https://doi.org/10.1007/s11105-012-0484-9.
[45] Paritosh K, Singh AK, Mehrotra AK, et al. Identification and characterization of the promoter of a gene expressing mainly in the tapetum tissue of cotton (*Gossypium hirsutum* L.). Plant Biotechnol Rep 2018;12:377–88. https://doi.org/10.1007/s11816-018-0501-z.
[46] Zhang M, Zheng X, Song S, et al. Spatiotemporal manipulation of auxin biosynthesis in cotton ovule epidermal cells enhances fiber yield and quality. Nat Biotechnol 2011;29:453–8. https://doi.org/10.1038/nbt.1843. PMid: 21478877.
[47] Yang J, Wang X, Hasi A, et al. Structural and functional analysis of a bidirectional promoter from *Gossypium hirsutum* in *Arabidopsis*. Int J Mol Sci 2018;19(11):3291. https://doi.org/10.3390/ijms19113291. PMid: 30360512.
[48] Wakano C, Byun JS, Di LJ, et al. The dual lives of bidirectional promoters. Biochim Biophys Acta 2012;1819(7):688–93. https://doi.org/10.1016/j.bbagrm.2012.02.006. PMid: 22366276.
[49] Hernandez-Garcia CM, Finer JJ. Identification and validation of promoters and *cis*-acting regulatory elements. Plant Sci 2014;217–218:109–19. https://doi.org/10.1016/j.plantsci.2013.12.007. PMid: 24467902.
[50] Dass A, Abdin MZ, Reddy VS, et al. Isolation and characterization of the dehydration stress-inducible GhRDL1 promoter from the cultivated upland cotton (*Gossypium hirsutum*). J Plant Biochem Biotechnol 2017;26:113–9. https://doi.org/10.1007/s13562-016-0369-3.
[51] Wu AM, Ling C, Liu JY. Isolation of a cotton reversibly glycosylated polypeptide (*GhRGP1*) promoter and its expression activity in transgenic tobacco. J Plant Physiol 2006;163(4):426–35. https://doi.org/10.1016/j.jplph.2005.06.014. PMid: 16455356.
[52] Dong CJ, Huang B, Liu JY. The cotton dehydration-responsive element binding protein GhDBP1 contains an EAR-motif and is involved in the defense response of *Arabidopsis* to salinity stress. Funct Plant Biol 2010;37(1):64–73. https://doi.org/10.1071/FP09100.
[53] Luo K, Zhang G, Deng W, et al. Functional characterization of a cotton late embryogenesis-abundant *D113* gene promoter in transgenic tobacco. Plant Cell Rep 2008;27:707–17. https://doi.org/10.1007/s00299-007-0482-9. PMid: 18066554.
[54] Xiao G, Zhao P, Zhang Y. A pivotal role of hormones in regulating cotton fiber development. Front Plant Sci 2019;10:87. https://doi.org/10.3389/fpls.2019.00087. PMid: 30838005.
[55] Wu AM, Hu JS, Liu JY. Functional analysis of a cotton cellulose synthase A4 gene promoter in transgenic tobacco plants. Plant Cell Rep 2009;28:1539–48. https://doi.org/10.1007/s00299-009-0753-8. PMid: 19655147.
[56] Zhang ZY, Zhao J, Hu Y, et al. Isolation of *GhMYB9* gene promoter and characterization of its activity in transgenic cotton. Biol Plant 2015;59:629–36. https://doi.org/10.1007/s10535-015-0545-7.
[57] Wang F, Wang C, Yan Y, et al. Overexpression of cotton *GhMPK11* decreases disease resistance through the gibberellin signaling pathway in transgenic *Nicotiana benthamiana*. Front Plant Sci 2016;7:689. https://doi.org/10.3389/fpls.2016.00689.
[58] Song W, Wang F, Chen L, et al. *GhVTC1*, the key gene for ascorbate biosynthesis in *Gossypium hirsutum*, involves in cell elongation under control of ethylene. Cells 2019;8(9):1039. https://doi.org/10.3390/cells8091039. PMid: 31492030.
[59] He X, Wang T, Zhu W, et al. GHHB12, a HD-ZIP I transcription factor, negatively regulates the cotton resistance to *Verticillium dahliae*. Int J Mol Sci 2018;19(12):3997. https://doi.org/10.3390/ijms19123997. PMid: 30545043.
[60] Zhang B, Yang Y, Wang J, et al. A CC-NBS-LRR type gene *GHNTR1* confers resistance to southern root-knot nematode in *Nicotiana benthamiana* and *Nicotiana tabacum*. Eur J Plant Pathol 2015;142:715–29. https://doi.org/10.1007/s10658-015-0646-3.
[61] Liu L, Wang W, Yang J, et al. Molecular cloning of *Ve* promoters from *Gossypium barbadense* and *G. hirsutum* and functional analysis in Verticillium wilt resistance. Plant Cell Tissue Organ Cult 2018;135:535–44. https://doi.org/10.1007/s11240-018-1485-7.
[62] Filichkin S.A., Breton G., Priest H.D., et al. Global profiling of rice and poplar transcriptomes highlights key conserved Circadian-controlled pathways and *cis*-regulatory modules. PLoS One 2011;6(6):e16907. https://doi.org/10.1371/journal.pone.0016907. PMid: 21694767
[63] Li Y, Liu X, Chen R, et al. Genome-scale mining of root-preferential genes from maize and characterization of their promoter activity. BMC Plant Biol 2019;19:584. https://doi.org/10.1186/s12870-019-2198-8. PMid: 31878892.
[64] Yuan C., Sun Q., Kong Y. Genome-wide mining seed-specific candidate genes from peanut for promoter cloning. PLoS One 2019;14(3):e0214025. https://doi.org/10.1371/journal.pone.0214025. PMid: 30921362
[65] Zhu T, Liang C, Meng Z, et al. CottonFGD: An integrated functional genomics database for cotton. BMC Plant Biol 2017;17:101. https://doi.org/10.1186/s12870-017-1039-x. PMid: 28595571.
[66] Unamba CIN, Nag A, Sharma RK. Next Generation Sequencing Technologies : The Doorway to the Unexplored Genomics of Non-Model Plants. Front Plant Sci 2015;6:1074. https://doi.org/10.3389/fpls.2015.01074. PMid: 26734016.
[67] Frith MC, Wan R, Horton P. Incorporating sequence quality data into alignment improves DNA read mapping. Nucleic Acids Res 2010;38(7): e100. https://doi.org/10.1093/nar/gkq010. PMid: 20110255.
[68] Philippe N, Boureux A, Bréhélin L, et al. Using reads to annotate the genome : influence of length, background distribution, and sequence errors on prediction capacity. Nucleic Acids Res 2009;37(15):e104. https://doi.org/10.1093/nar/gkp492. PMid: 19531739.
[69] Auer PL, Doerge RW. Statistical Design and Analysis of RNA Sequencing Data. Genetics 2010;185(2):405–16. https://doi.org/10.1534/genetics.110.114983. PMid: 20439781.
[70] Conesa A, Madrigal P, Tarazona S, et al. A survey of best practices for RNA-seq data analysis. Genome Biol 2016;17:13. https://doi.org/10.1186/s13059-016-0881-8. PMid: 26813401.
[71] McIntyre LM, Lopiano KK, Morse AM, et al. RNA-seq: Technical variability and sampling. BMC Genomics 2011;12:293. https://doi.org/10.1186/1471-2164-12-293. PMid: 21645359.
[72] Packer J, Trapnell C. Single cell multi-omics: an engine for new quantitative models of gene regulation. Trends Genet 2018;34(9):653–65. https://doi.org/10.1016/j.tig.2018.06.001. PMid: 30007833.
[73] Efroni I, Birnbaum KD. The potential of single-cell profiling in plants. Genome Biol 2016;17:65. https://doi.org/10.1186/s13059-016-0931-2. PMid: 27048384.
[74] Yuan Y, Lee H, Hu H, et al. Single-cell genomic analysis in plants. Genes (Basel) 2018;9(1):50. https://doi.org/10.3390/genes9010050. PMid: 29361790.

[75] Dinneny JR, Long TA, Wang JY, et al. Cell identity mediates the response of *Arabidopsis* roots to abiotic stress. Science 2008;320(5878):942–5. https://doi.org/10.1126/science.1153795. PMid: 18436742.
[76] Kiegle E, Moore CA, Haseloff J, et al. Cell-type-specific calcium responses to drought, salt and cold in the *Arabidopsis* root. Plant J 2000;23(2):267–78. https://doi.org/10.1046/j.1365-313x.2000.00786.x. PMid: 10929120.
[77] Marcel S, Sawers R, Oakeley E, et al. Tissue-adapted invasion strategies of the rice blast fungus *Magnaporthe oryzae*. Plant Cell 2010;22(9):3177–87. https://doi.org/10.1105/tpc.110.078048. PMid: 2085884.
[78] Betancur L, Singh B, Rapp RA, et al. Phylogenetically distinct cellulose synthase genes support secondary wall thickening in arabidopsis shoot trichomes and cotton fiber. J Integr Plant Biol 2010;52(2):205–20. https://doi.org/10.1111/j.1744-7909.2010.00934.x. PMid: 20377682.
[79] Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet 2015;16:133–45. https://doi.org/10.1038/nrg3833. PMid: 25628217.
[80] Jin J, Tian F, Yang D, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 2016;45(D1):D1040-5. https://doi.org/10.1093/nar/gkw982. PMid: 27924042.
[81] Khan A, Fornes O, Stigliani A, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res 2018;46(D1):D260–6. https://doi.org/10.1093/nar/gkx1126. PMid: 29140473.
[82] Rombauts S, Déhais P, Van Montagu M, et al. PlantCARE, a plant *cis*-acting regulatory element database. Nucleic Acids Res 1999;27(1):295–6. https://doi.org/10.1093/nar/27.1.295. PMid: 9847207.
[83] Hombach D, Schwarz JM, Robinson PN, et al. A systematic, large-scale comparison of transcription factor binding site models. BMC Genomics 2016;17:388. https://doi.org/10.1186/s12864-016-2729-8. PMid: 27209209.
[84] Lai X, Stigliani A, Vachon G, et al. Building transcription factor binding site models to understand gene regulation in plants. Mol Plant 2019;12(6):743–63. https://doi.org/10.1016/j.molp.2018.10.010. PMid: 30447332.
[85] Tompa M, Li N, Bailey TL, et al. Assessing computational tools for the discovery of transcription factor binding sites. Nat Biotechnol 2005;23:137–44. https://doi.org/10.1038/nbt1053. PMid: 15637633.
[86] Stormo GD, Zhao Y. Determining the specificity of protein–DNA interactions. Nat Rev Genet 2010;11:751–60. https://doi.org/10.1038/nrg2845. PMid: 20877328.
[87] Weber B, Zicola J, Oka R, et al. Plant Enhancers: A Call for Discovery. Trends Plant Sci 2016;21(11):974–87. https://doi.org/10.1016/j.tplants.2016.07.013. PMid: 27593567.
[88] Zaret KS, Mango SE. Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr Opin Genet Dev 2016;37:76–81. https://doi.org/10.1016/j.gde.2015.12.003. PMid: 26826681.
[89] Zhang H, Lang Z, Zhu J. Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol 2018;19:489–506. https://doi.org/10.1038/s41580-018-0016-z. PMid: 29784956.
[90] Verkest A, Bourout S, Debaveye J, et al. Impact of differential DNA methylation on transgene expression in cotton (*Gossypium hirsutum* L.) events generated by targeted sequence insertion. Plant Biotechnol J 2019;17(7):1236–47. https://doi.org/10.1111/pbi.13049. PMid: 30549163.
[91] Harmston N, Lenhard B. Chromatin and epigenetic features of long-range gene regulation. Nucleic Acids Res 2013;41(15):7185–99. https://doi.org/10.1093/nar/gkt499. PMid: 23766291.
[92] Sanyal A, Lajoie BR, Jain G, et al. The long-range interaction landscape of gene promoters. Nature 2012;489:109–13. https://doi.org/10.1038/nature11279. PMid: 22955621.
[93] Wang M, Wang P, Lin M, et al. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. Nat Plants 2018;4:90–7. https://doi.org/10.1038/s41477-017-0096-3. PMid: 29379149.
[94] Davey MR, Anthony P, Power JB, et al. Plant protoplasts: status and biotechnological perspectives. Protoplasma 2005;23(2):131–71. https://doi.org/10.1016/j.biotechadv.2004.09.008. PMid: 15694124.
[95] Jiang F, Zhu J, Liu H. Protoplasts: a useful research system for plant cell biology, especially dedifferentiation. Protoplasma 2013;250:1231–8. https://doi.org/10.1007/s00709-013-0513-z. PMid: 23719716.
[96] Taylor NJ, Fauquet CM. Microparticle bombardment as a tool in plant science and agricultural biotechnology. DNA Cell Biol 2002;21(12):963–77. https://doi.org/10.1089/104454902762053891. PMid: 12573053.
[97] Potrykus I. Gene transfer to plants: assessment and perspectives. Physiol Plant 1990;79(1):125–34. https://doi.org/10.1111/j.1399-3054.1990.tb05874.x.
[98] Ali S, Kim WC. A Fruitful decade using synthetic promoters in the improvement of transgenic plants. Front Plant Sci 2019;10:1433. https://doi.org/10.3389/fpls.2019.01433. PMid: 31737027.
[99] Andres J, Blomeier T, Zurbriggen MD. Synthetic switches and regulatory circuits in plants. Plant Physiol 2019;179(3):862–84. https://doi.org/10.1104/pp.18.01362. PMid: 30692218.
[100] Pandiarajan R, Grover A. *In vivo* promoter engineering in plants: Are we ready? Plant Sci 2018;277:132–8. https://doi.org/10.1016/j.plantsci.2018.10.011. PMid: 30466578.