Identification and validation of maize enhancers
A cartography of the maize regulatory genome
Weber, B.B.M.

Publication date
2018

Document Version
Other version

License
Other

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 1

Plant enhancers: A call for discovery

Blaise Weber*, Johan Zicola*, Rurika Oka*, Maike Stam

*: Equal contributors

This Chapter was initially published as a review in Volume 21, issue 11 of journal “Trends In Plant Science”

Abstract

Higher eukaryotes typically contain many different cell types, displaying different cellular functions that in addition are influenced by biotic and abiotic cues. The different functions are characterized by specific gene expression patterns that in part are mediated by regulatory sequences such as transcriptional enhancers. Recent genome-wide approaches have identified thousands of enhancers in animals, reviving the interest in enhancers in gene regulation. Although the regulatory roles of plant enhancers are as critical as those in animals, genome-wide approaches are only very recently applied to plants. Here we review characteristics of enhancers at the DNA and chromatin level in plants and other species, their similarities and differences, and techniques widely used for the genome-wide, high-throughput discovery of enhancers in animal systems that can be implemented in plants.
Introduction: enhancers in gene regulation

The vast majority of eukaryotes consist of numerous different cell types. In a given organism, the different cell types possess the same DNA, and it is fascinating that such diversity of cell types can arise from one and the same set of chromosomes. Cells of all organisms are in addition able to respond abiotic and biotic environmental cues, such as light, temperature, chemicals and pathogens. Crucial for the successful production of highly specialized cell types and their response to external signals is a correct temporal and spatial regulation of gene expression [1]. This is for a large part accomplished through the activation and repression of the relevant cis-regulatory elements, such as transcriptional enhancers (hereafter referred to as enhancers) and silencers, at the correct moment in time and space [2,3]. Enhancers are non-coding DNA sequences that can be bound by multiple transcription factors (TFs) to activate the expression of genes localized up to mega-bases away (Figure 1A) [4,5]. Silencers are DNA elements that repress gene expression [3]. Both enhancers and silencers can be located up- or downstream of their target genes and function in an orientation-independent manner [6]. Enhancing and silencing functions can also be combined into one and the same DNA element, such as shown for the light-inducible and tissue-specific regulatory elements of ab80 and rbcS-3A in pea (Pisum sativum) [7–10]. This review focuses on enhancers.

The general mechanisms by which enhancers are activated and trigger gene expression are well studied [11]. Enhancers are generally activated by the binding of pioneer TFs, followed by the recruitment of co-activators such as histone acetyltransferases and chromatin remodelers that altogether increase chromatin accessibility [12]. This increased accessibility promotes binding of other TFs, leading to transcriptional activation of the target genes [12]. To do so, enhancers physically interact with the promoters of their cognate genes (Figure 1B). Ultimately, transcription is initiated by RNA polymerase II at the Transcription Start Site (TSS) of the gene [13].

In the last decades, several examples of enhancers have been identified and characterized in different species, including yeast, fungus, animals and plants (e.g. [14–21]). These examples have mainly been identified using low-throughput methods such as enhancer trapping, promoter deletion analysis, recombinant analysis and quantitative trait locus mapping. The recent development of affordable next-generation sequencing technologies, in combination with the identification of general enhancer features, especially DNA and chromatin features, allowed the genome-wide identification of enhancers in a high-throughput manner. This led to the discovery of
Chapter 1

Figure 1. Schematic illustrations of chromatin features and associated proteins observed at enhancer regions and target genes in animals. (A) Enhancer located distantly from its target gene. Presence of H3K4me1 and absence of H3K4me3 distinguishes enhancers from promoters. (B) Active enhancer physically interacts with the promoter of its target gene through protein complexes. (C) Inactive enhancers are associated with H3K27me3 and H3K4me1. (D) Active enhancers are associated with nucleosome-depleted regions as well as H3K4me1, H3K9ac and H3K27ac (annotated as H3 acetylation).

over 43,000 enhancer candidates in the human genome [22] and up to 100,000 predicted enhancers in Drosophila (Drosophila melanogaster) [23]. Remarkably, genes are often shown to be regulated by more than one enhancer [23–25]. The crucial roles of enhancers in gene regulation have been emphasized in studies linking enhancers not only with proper embryonic development and the specialization of cell types, but also with a large set of diseases, including cancer [2,4,26–29]. Plant genomes are very likely to also contain numerous (distant) enhancers. So far, we do not know how many enhancers are present across plant species, and their chromatin features are poorly characterized, except for a few examples such as the enhancers of the Pea plastocyanin (PetE) gene in pea [30], the booster1 (b1) gene in maize (Zea mays ssp. mays L.) [14,31,32], and the enhancer of the FLOWERING LOCUS T (FT) gene in
Arabidopsis (Arabidopsis thaliana) [17,33] (Table 1). The first genome-wide study identifying enhancers in Arabidopsis based on chromatin features [34] reflects the renewed interest for enhancers in plants.

What are the general properties displayed by enhancers? Large-scale animal studies comparing several molecular features showed that, depending on their activity state, enhancers are characterized by different DNA and chromatin features [11]. Inactive enhancers are typified by low chromatin accessibility and the presence of specific histone modifications, e.g. trimethylation of lysine 27 of histone H3 (H3K27me3) (Figure 1C) [35,36]. Active enhancers generally display high chromatin accessibility and histone acetylation, eRNA transcription and low DNA methylation [22,35–40] (Figure 1D). Specific histone modifications, such as H3K4me1, indicate enhancers independent of their activity level. Despite a growing interest, there is no study reporting a comparison of several chromatin features at enhancers in plants yet.

In this review, we (i) provide an overview on the current knowledge on enhancers in plants, including their molecular characteristics, (ii) discuss the potential commonalities and differences between plant and animal enhancers, and (iii) discuss and compare the different techniques available to identify and characterize enhancers in plants, focusing mainly on high-throughput methods based on next-generation sequencing approaches. Finally, we will provide directions for future research.

Enhancers in Plants

One of the first enhancers described in plants dates back to 1985 when Simpson et al. [41] reported an enhancer of the chlorophyll a/b-binding protein gene AB80 in pea. Since then, other enhancers have been identified and characterized in different plant species (examples given in Table 1). At first, plant enhancers were primarily characterized using promoter deletion assays, electrophoretic mobility shift assays (EMSA), and DNaseI footprinting [15,42–44], subsequently chromatin features were investigated as well. The enhancers of the hydroxyproline-rich glycoprotein (HRGP) gene in maize and the PetE gene in pea are among the first enhancers examined for accessible chromatin and histone acetylation, respectively [30,44]. Currently, a hepta-repeat region of the b1 gene in maize is one of the enhancers of which its chromatin features are best-characterized. When
Table 1. Examples of currently known plant enhancers and their associated characteristics

<table>
<thead>
<tr>
<th>Enhancer</th>
<th>Target gene</th>
<th>Organism</th>
<th>Location(^{a})</th>
<th>Chromatin Accessibility(^{b})</th>
<th>Chromatin Interaction(^{c})</th>
<th>Sequence conservation among or within species</th>
<th>TF binding motif</th>
<th>Histone mark(^{d})</th>
<th>Reporter assay(^{e})</th>
<th>DNA methylation(^{f})</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hepta-repeat</td>
<td>boosterl (b1) (GRMZM2G172795)</td>
<td>Zea mays</td>
<td>100 kb upstream</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes, among different maize lines(^{g})</td>
<td>ND</td>
<td>H3ac (A)</td>
<td>Yes</td>
<td>Yes</td>
<td>[14, 3, 132, 121]</td>
</tr>
<tr>
<td>Block C</td>
<td>FLOWERIN G LOCUS T (FT) (A1t65480)</td>
<td>Arabidopsis thaliana</td>
<td>5 kb upstream</td>
<td>Yes(^{h})</td>
<td>Yes</td>
<td>Yes, among Brassicaceae</td>
<td>CCAAT-box, REalpha, I-box</td>
<td>H3K9K1 4ac (A)</td>
<td>Yes(^{i})</td>
<td>Yes</td>
<td>[17, 3, 53, 75]</td>
</tr>
<tr>
<td>Region C</td>
<td>LATERAL SUPPRESS OR (LAS) (AT1G55580)</td>
<td>Arabidopsis thaliana</td>
<td>3.2 kb downstream</td>
<td>Yes</td>
<td>ND</td>
<td>Arabis alpina, Arabidopsis lyrata and Capsella rhabella</td>
<td>ND</td>
<td>H3K27m e3 (I)</td>
<td>ND</td>
<td>ND</td>
<td>[128]</td>
</tr>
<tr>
<td>P268 PetE enhancer</td>
<td>pea plastocyanin gene (PetE)</td>
<td>Pisum sativum</td>
<td>177 bp upstream</td>
<td>Yes</td>
<td>ND</td>
<td>ND</td>
<td>HMG-I/Y binding motif</td>
<td>H3ac and H4ac (A)</td>
<td>Yes</td>
<td>ND</td>
<td>[30]</td>
</tr>
<tr>
<td>P1-rr distal enhancer</td>
<td>pericarp color 1 (p1) (GRMZM2G084799)</td>
<td>Zea mays</td>
<td>6 kb upstream</td>
<td>Yes</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>Yes</td>
<td>[124, 129]</td>
</tr>
<tr>
<td>Vegetative to generative 1 (Vgt1)</td>
<td>ZmRap2.7 (GRMZM2G700665)</td>
<td>Zea mays</td>
<td>70 kb upstream</td>
<td>ND</td>
<td>ND</td>
<td>Conserved among late flowering lines</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>[16, 130]</td>
</tr>
<tr>
<td>AB80 enhancer</td>
<td>AB80 chlorophyll a/b binding protein (CAB) gene</td>
<td>Pisum sativum</td>
<td>400 bp upstream</td>
<td>ND</td>
<td>ND</td>
<td>Yes, among the CAB genes in N. plumbaginifoli a and wheat</td>
<td>G-boxes, GATA-box</td>
<td>ND</td>
<td>Yes</td>
<td>ND</td>
<td>[41, 123, 131]</td>
</tr>
<tr>
<td>Enhancer-like element</td>
<td>Ribulose 1,5-biphosphate carboxylase small subunit (rbcS) genes SS3.6, E9, 3A, 3C</td>
<td>Pisum sativum,</td>
<td>400 bp upstream</td>
<td>ND</td>
<td>ND</td>
<td>Yes, among Pisum sativum, Solanum lycopersicum, wheat, Nicotiana plumbaginifoli a, Antirrhinum majus</td>
<td>Box II: Homolog y with SV40 core enhancer GT motif, Box III: homolog y with human β-interferon enhancer and adenoviru s 5 E1A enhancer</td>
<td>ND</td>
<td>Yes</td>
<td>ND</td>
<td>[8, 9, 15]</td>
</tr>
<tr>
<td>TACPyA T repeats</td>
<td>Chalcone synthase A (chsA)</td>
<td>Petunia hybrida</td>
<td>-67 to -53 bp upstream</td>
<td>ND</td>
<td>ND</td>
<td>Yes, with Antirrhinum majus</td>
<td>Two TACPyA T motifs</td>
<td>ND</td>
<td>Yes</td>
<td>ND</td>
<td>[132, 133]</td>
</tr>
<tr>
<td>L3 enhancer</td>
<td>Putative: AT-hook motif nuclear-localized protein 22</td>
<td>Arabidopsis thaliana</td>
<td>4 kb upstream</td>
<td>Yes</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>Yes</td>
<td>[34]</td>
</tr>
</tbody>
</table>
Distal cis-element (DICE)  |  Bx1  |  Zea mays  |  140 kb upstream  |  ND  |  ND  |  ND  |  ND  |  ND  |  ND  |  Low methylation  |  [134]

th1 enhancer  |  teosinte branched 1 (tb1)  |  Zea mays  |  60 kb upstream  |  ND  |  ND  |  ND  |  ND  |  ND  |  Yes  |  ND  |  [135, 136]

Egg apparatus -specific enhancer (EASE)  |  ND  |  Arabidopsis thaliana  |  ND  |  Yes  |  ND  |  Yes, among different Arabidopsis accessions  |  ND  |  ND  |  Yes  |  ND  |  [21]

MATURE MINOR VEIN ELEMENT 1 (MMVE1)  |  ND  |  Arabidopsis thaliana  |  ND  |  ND  |  ND  |  Yes, among Brassicaceae API and ARF binding motifs, CACGTG motif  |  ND  |  Yes  |  ND  |  [20]

HRGP enhancer  |  Hydroxyproline-rich glycoprotein (HRGP)  |  Zea mays  |  1380 to 220 bp Upstream  |  Yes  |  ND  |  ND  |  ND  |  ND  |  ND  |  [44]

ND stands for not determined

a Location of enhancer relative to the TSS of its target gene
b Correlation between enhancer activity and chromatin accessibility
c Physical interaction observed with target gene upon gene activation
d Enrichment in histone marks: (A) for active enhancer, (I) for inactive enhancer
e Activity of enhancer measured in reporter assay (e.g. transient or transgenic minimal reporter assay or enhancer trap)
f DNA methylation at enhancer is associated with silencing of the enhancer
 g Sequence is conserved, but number of repeats varies
 h Exact locations of interactions do not agree between publications
 i Zicola and Turck, unpublished data

active, the hepta-repeat displays several hallmarks of active enhancers in mammals such as accessible chromatin, H3 acetylation and low DNA methylation [14,31,32]. Most other plant enhancers are less well-characterized, and the list of their associated characteristics is incomplete. Recent studies in Arabidopsis and rice (Oryza sativa) [34,45] are the first reporting the use of chromatin features to identify cis-regulatory elements in a genome-wide, high-throughput manner.
Characteristics of enhancers

Enhancer regions display specific characteristics including the presence of TF binding motifs, chromatin accessibility, particular histone modifications, eRNA expression, low DNA methylation and physical interactions with their target genes [11]. Altogether, these signatures can help to identify enhancers in a genome. For better enhancer prediction and characterization, multiple features should be studied in parallel, also given particular features can as well be displayed by other cis-regulatory elements, the TSS or coding regions of genes. Below, we discuss enhancer characteristics identified in animals that can be used to identify and characterize plant enhancers.

Transcription factor binding motifs

Enhancers are activated by the binding of TFs (Figure 1). This binding of TFs to DNA is specified by specific consensus sequences, called TF binding motifs, and/or particular chromatin features, such as histone modifications [46]. Enhancers are enriched with multiple TF binding motifs. Nearly 600 different experimentally validated TF binding motifs were reported in human (Transfac in 2003 [47]). In Arabidopsis, approximately 530 TF binding motifs are experimentally determined, e.g. by mobility shift assays or DAP-seq [48,49]. In other plant species, e.g. maize, many TF binding motifs were predicted [50]; however, few studies validated TF binding motifs experimentally [50,51].

Chromatin accessibility

The degree of chromatin accessibility impacts the binding of TFs to regulatory sequences [25,52] (Figure 1). Chromatin accessibility depends on the local nucleosome occupancy and binding of chromatin-associated proteins (Figure 1A and D). Active cis-regulatory elements such as promoters and enhancers are localized in accessible genomic regions, also called Nucleosome-Depleted Regions (NDRs) [25]. NDRs have been mapped genome-wide in Arabidopsis, maize and rice [45,53,54]. Consistent with having a regulatory role, NDRs are enriched at TF binding sites and conserved non-coding sequences in Arabidopsis and maize [53,54]. Moreover, several intergenic NDRs identified in Arabidopsis were validated as enhancers in transgenic experiments [34].
Histone modifications

Histone marks are post-translational modifications of histones that have different roles in gene regulation, including modulation of chromatin accessibility [55] (Figure 1). Nucleosomes at enhancer regions are shown to carry specific histone marks. In animals, monomethylation at lysine 4 of histone 3 (H3K4me1) is found at both active and inactive enhancers [35]. Acetylation at lysine 9, 12, 14, and 27 of H3 (H3K9ac, H4K12ac, H3K14ac, and H3K27ac) characterizes active enhancers [37,38,56], while H3K27me3 marks inactive enhancers [35]. All of these marks are, however, present at TSSs and/or coding regions as well, hampering the unequivocal identification of enhancer sequences by one histone mark. Data from an additional mark, e.g. H3K4me3, which is preferentially enriched at TSSs, can be used to distinguish TSSs from enhancers [57–59].

Which histone marks best indicate plant enhancers and their activity states is not yet entirely clear. Knowledge on such marks is slowly emerging. For instance, the active pea PetE and maize b1 enhancers were reported to be enriched in H3/H4ac and H3K9/K14ac, respectively [30,31]. Furthermore, intergenic NDRs in rice were strongly associated with H4K12ac, but also H3K27me3 [45]. Moreover, a recent study in Arabidopsis revealed a positive correlation between inactive enhancers and H3K27me3, and between active enhancers and H3K27ac, with the former correlation being more clear than the latter [34]. Altogether, the current results indicate that active plant enhancers are generally associated with H3 and H4 acetylation, while inactive enhancers appear associated with H3K27me3. At the same time, we want to emphasize that more research is required to identify the histone modifications that best detect active and inactive enhancers in plants.

Enhancer RNAs (eRNAs)

In animals, the presence of enhancer transcripts (enhancer RNAs or eRNAs) are shown to provide a good indication of active enhancers [22,40]. eRNAs are non-coding, relatively short (<2kb), capped, mostly non-polyadenylated, unspliced, and rapidly degraded by exosomes. Animal enhancers are often transcribed bidirectionally [22,40], and although the absolute eRNA transcript levels are much lower than those of protein coding genes, they correlate with those of their target genes [22]. Insight into potential roles for eRNAs is emerging. Some eRNAs are, for example, necessary to recruit TFs to enhancers [60], or to mediate enhancer-promoter interactions [61]. It can however not be excluded that part of the eRNAs have no role in gene regulation and may be products of leaky RNA pol II expression [62–64]. Recent findings suggested a significant association of non-coding RNAs with NDRs in Arabidopsis.
Chapter 1

[34]. The presence, characteristics and roles of eRNAs in gene regulation in plants remains to be further investigated.

**DNA methylation**

DNA methylation is associated with transcriptional silencing in both animals and plants [65], and when present at enhancers, it is shown to downregulate the expression of target genes [39, 66]. In plants, this is, for example, observed for DNA methylation at regulatory sequences of *FLOWERING WAGENINGEN (FWA)*, *TOO MANY MOUTHS (TMM)*, and *FT* in *Arabidopsis* [67–69], and *pericarp color1 (p1)* and *b1* in maize [31, 70]. In human and mouse, the DNA methylation level at numerous enhancers is dynamically regulated, negatively correlating with the activity of enhancers, allowing the identification of tissue-specific enhancers [39]. Except for a study in tomato [66], there is not much evidence that DNA methylation at *cis*-regulatory elements is regulated in a dynamic manner in plants.

**Chromatin interactions**

Enhancers and target genes must be in close proximity to allow enhancers to activate transcription (Figure 1B). Chromosomal conformation studies indeed provided ample evidence that enhancers and their target genes physically interact with each other [32, 71]. In mammals, CTCF, cohesin, a mediator complex and sometimes also eRNAs have been shown to mediate enhancer-promoter interactions [61, 72, 73]. Besides general protein factors and complexes, sequence-specific TFs are also required for enhancer-promoter interactions. The Erythroid Krüppel-like transcription Factor (EKLF) is for instance involved in establishing chromatin interactions at the active β-globin locus [74]. Chromatin interactions between distant enhancers and their target genes are also reported in plants. In maize, the 100 kb upstream hepta-repeat enhancer interacts with the TSS region of the *b1* gene when *b1* is expressed [32]. Similarly in *Arabidopsis*, two independent studies revealed interactions between the TSS and upstream regulatory regions of *FT* [33, 75] The results, however, do not agree on the exact genomic identity of the interacting regions.

**Techniques for enhancer identification**

In the following section, we will review techniques that are currently used for identifying enhancers in a genome-wide, high-throughput manner, mainly using next-generation sequencing technologies. Techniques for validating enhancer candidates are discussed as well. As each technique has inherent methodological biases and limitations, to increase the predictive
value, combining different approaches is preferred over the use of a single method. For a summary, see Table 2.

Table 2. Summary of currently used or promising techniques to identify enhancers

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Studied Aspect</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF Binding Motif</td>
<td>TF binding motifs</td>
<td>Identifies TF binding sites</td>
<td>High false discovery rate; prior knowledge on TF binding motif required</td>
<td>[81]</td>
</tr>
<tr>
<td>Scan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNase-seq</td>
<td>Open chromatin</td>
<td>TF binding motifs can be detected</td>
<td>DNase I can introduce cleavage bias, affecting TF footprint detection</td>
<td>[45,53,80]</td>
</tr>
<tr>
<td>ChIP-seq</td>
<td>Histone modifications, TFs, chromatin-</td>
<td>A wide range of targets can be studied</td>
<td>Relies on the availability of high-quality antibodies or tagged proteins</td>
<td>[89–91]</td>
</tr>
<tr>
<td></td>
<td>associated proteins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA-seq</td>
<td>Transcript levels</td>
<td>eRNA levels implicate enhancer activity, detects directionality of</td>
<td>eRNA expression is low, high sequencing depth required</td>
<td>[40]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>transcription</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAGE</td>
<td>Transcript levels</td>
<td>eRNA levels implicate enhancer activity, detects directionality of</td>
<td>Only detects capped eRNAs</td>
<td>[22]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>transcription</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRO-seq</td>
<td>Nascent transcript levels</td>
<td>eRNA transcription implicates enhancer activity</td>
<td>Challenging technique</td>
<td>[93,94]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STARR-seq</td>
<td>Enhancer mapping (and activity)</td>
<td>High-throughput identification and validation of enhancers in parallel</td>
<td>Minimal promoter used influences the set of identified enhancers</td>
<td>[96,97]</td>
</tr>
<tr>
<td>BS-seq</td>
<td>DNA methylation</td>
<td>Single bp resolution</td>
<td>High sequence depth needed; incomplete BS conversion affects data interpretation</td>
<td>[66,98,103]</td>
</tr>
<tr>
<td>Enhancer trapping</td>
<td>Enhancer activity</td>
<td>Visualizes tissue-specific pattern mediated by endogenous cts-regulatory</td>
<td>Difficult to locate trapped enhancers</td>
<td>[20,21,106,109]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sequences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3C technology</td>
<td>Chromatin interactions</td>
<td>Identifies promoter-enhancer interactions</td>
<td>Challenging technique; trade-off between number of observed interactions and resolution</td>
<td>[71,111,112,117]</td>
</tr>
<tr>
<td>Reporter assay</td>
<td>Transcriptional activity</td>
<td>Confirms activity and tissue specificity of enhancer candidates</td>
<td>Potential expression bias arising from test conditions and the minimal promoter used</td>
<td>[15,119,120,124,125]</td>
</tr>
</tbody>
</table>

**Based on DNA sequence: TF binding motif scan**

Enhancers are bound by TFs, therefore scanning genomes for TF binding motifs can contribute to the identification of enhancers [76]. However, the presence of a TF binding motif does not guarantee the functional binding of TFs in vivo, as TF binding motifs are typically less than 10 nucleotides and therefore can appear in a genome by a random chance [77,78]. In addition, TF binding is not always highly sequence-specific [46] and may therefore be hard to predict by
motif scanning. Since TFs often function in complexes, detecting clusters of TF binding motifs reduces the number of false positives [79]. In plants other than Arabidopsis, motif scanning is limited by the relatively low number of known TF binding motifs. In such case, putative TF binding motifs can be determined by DNA footprinting using DNase-seq, by ChIP-seq or by analyzing promoter sequences of co-expressed genes using multiple tissue or multiple time point data [50,51,80,81].

**Assaying chromatin accessibility**

**DNase-seq**
Active enhancer sequences are usually located in NDRs and are therefore sensitive to nuclease activity [82]. Therefore, DNase-seq is a very valuable tool to identify cis-regulatory sequence. With this method, DNase I Hypersensitive Sites (DHSs) can be identified by partial digestion of chromatin with the endonuclease DNase I, followed by sequencing of the small fragments representing the accessible fraction of the genome (DNase-seq) [80]. Alternatively, the ends of large DNA fragments, representing the less accessible fraction of the genome, can be sequenced followed by identification of DHSs [83]. DNase-seq robustly identifies DHSs, but is less sensitive in predicting TF binding motifs, due to its intrinsic cleavage bias [84].

**ATAC-seq: Assay for Transposase-Accessible Chromatin**
ATAC-seq, a technique in which the engineered Transposase Tn5 ligates accessible DNA to sequencing adapters, was shown to provide a good alternative to DNase-seq in human [85]. Importantly, with ATAC-seq, highly comparable results to DNase-seq could be obtained with 200 times less cells. Like DNase-seq, ATAC-seq can also be used for DNA footprinting [85]. ATAC-seq data sets have not yet been reported for plant tissue, but the method seems attractive, especially for analyzing tissues that are difficult to collect in large quantities.

**FAIRE: Formaldehyde-Assisted Isolation of Regulatory Elements**
Another method to identify accessible chromatin is FAIRE-seq [86]. FAIRE identifies protein-free DNA regions (*i.e.* free from nucleosomes) by cross-linking tissue or cells with formaldehyde, followed by sonication of chromatin and phenol-chloroform extraction of the nucleosome-free DNA fragments. FAIRE offers a lower resolution than DNase I-based assays as sonication provides higher background noise than DNase I digestion [87]. FAIRE-qPCR applied at the maize *b1* locus revealed FAIRE enrichment at the active hepta-repeat enhancer, demonstrating the potential of FAIRE to identify plant enhancers [32]. A FAIRE-seq protocol
Chapter 1

was developed for Arabidopsis, but no genome-wide FAIRE study has yet been reported in plants [88].

**ChIP-seq: Chromatin ImmunoPrecipitation sequencing**

Chromatin ImmunoPrecipitation followed by sequencing (ChIP-seq) can identify DNA regions based on their associated modifications or proteins (e.g. histone marks, TFs and polymerases) [11,89,90]. For instance in animals, antibodies recognizing H3K27ac and the histone acetyltransferase p300, can be used to detect active enhancers [38,91]. The modifications or proteins targeted by ChIP influence the number and types of enhancers identified. To identify TF binding sites with an almost base pair resolution, ChIP can be coupled to an exonuclease treatment (ChIP-exo; exonucleases hereby remove DNA not bound by TFs) [92]. ChIP has been adapted to plants [89,90]. However, the most relevant combination of histone marks or TFs for enhancer identification remain to be determined [31,34,45].

**Assaying transcriptional activity**

**RNA-seq based methods**

In animals, the production of eRNAs provides a good indication of enhancer activity [22]. eRNAs can be identified by different techniques. When sequencing RNAs (RNA-seq) a high sequencing depth is required to detect the low abundant eRNAs [40]. By using Cap Analysis of Gene Expression (CAGE), with which only the 5’ ends of RNAs are sequenced, eRNAs can be detected at lower sequencing depth [22]. Besides being low abundant, eRNAs are sensitive to degradation, therefore genome-wide nuclear Run-Ons (GRO-seq), which measures nascent transcript production, may provide a higher sensitivity to detect eRNAs than CAGE [93,94]. The recently developed GRO-Cap technique, which allows detection of nascent capped transcripts, maps TSSs with higher accuracy than GRO-seq and could be an interesting technique to apply in plants for eRNA detection [95].

**STARR-seq: Self-Transcribing Active Regulatory Region Sequencing**

STARR-seq is a technique developed in Drosophila to capture sequences with enhancer activity [96]. With this technique, random fragments from sheared genomic DNA are cloned between a minimal promoter incapable of driving high expression and a polyadenylation sequence. The resulting plasmids are transfected into cells, after which fragments with enhancer activity can enhance their own transcription. Hence, sequencing of polyadenylated transcripts isolated from
the transfected cells reveals the sequence and transcriptional strength of cloned DNA fragments. Not all minimal promoter-enhancer combinations lead to transcriptional activation [97]. The minimal promoter used, therefore, determines the enhancers identified. STARR-seq still has to be implemented in plants.

**BS-seq: Bisulfite sequencing**

As mentioned before, low DNA methylation levels can indicate enhancers [39]. Genome-wide DNA methylation levels can be measured using Bisulfite (BS) conversion, which converts unmethylated cytosines to thymines, followed by sequencing (BS-seq) [98]. BS-seq was first implemented in Arabidopsis, and then used in several other plant species [99–101]. BS-seq offers a single-base resolution of genome-wide DNA methylation profiles [102], allowing the precise delimitation of low methylated regions using computational tools [103]. Note that incomplete BS conversion and sequence polymorphisms can affect data interpretation.

**Enhancer trapping**

With enhancer trapping, enhancers are detected by random genomic insertion of a reporter gene that is driven by a minimal promoter that is not sufficient to drive expression [104,105]. Expression of the reporter gene can be observed when inserted adjacent to an enhancer activating the gene [106]. For plants, typically a reporter gene (e.g. GPF) driven by the minimal CaMV-35S promoter is used. Numerous enhancer trapping lines were isolated in Arabidopsis [107,108] and rice [109,110]. Most lines showed tissue-specific expression, hence this method allows identification of regulatory sequences mediating expression patterns of interest. In reality, however, only few enhancers have been identified using this method, for example the minor vein phloem specific enhancer MATURE MINOR VEIN ELEMENT1 (MMVE1) [20] and Egg Apparatus-Specific Enhancer (EASE) [21]. This suggests that it is difficult to identify the trapped enhancers; indeed, enhancers can be located distally to enhancer traps. When studying a large, complex genome, enhancer identification will be even more cumbersome.

**Characterisation of enhancers and their target genes**

**3C-based techniques: Chromosome Conformation Capture**

Chromatin Conformation Capture (3C) and its derivatives (e.g. 4C, 5C and Hi-C) measure relative interaction frequencies between different genomic regions [71,111]. In short, interacting chromosomal regions are cross-linked, followed by restriction digestion and
intramolecular ligation of the interacting fragments. Finally, interaction frequencies are quantified using qPCR (3C) or sequencing. The main strength of these methods is their ability to identify target genes of enhancers (and vice versa). 3C and 4C are the method of choice when focusing on specific enhancers or genes [112,113], while a Hi-C protocol with a resolution of 1kb allows genome-wide studies of enhancer-promoter interactions [114]. To avoid loss of sequencing capacity on interactions other than between enhancers and promoters, methods such as Capture C and Hi-Cap have been developed [115,116]. In plants, 3C was first implemented at the bl locus in maize, identifying interactions between the hepta-repeat enhancer and the bl gene [32,117]. The outcome of 3C and 4C studies on enhancer-promoter or other functional interactions in Arabidopsis is more cumbersome [33,75,118]. This can be explained by the compact genome size of Arabidopsis, hampering both the identification of the exact interacting sequences and the detection of relevant interactions above the background level of random ligation events.

**Reporter assay**

The gold standard for testing enhancer sequences is a reporter assay [11]. Usually, with this method, a candidate enhancer and a control fragment are cloned upstream of a minimal promoter driving a reporter gene, followed by introduction into the tissue or cell type of interest and measurement of reporter gene activity. In plants, one can either use transient reporter assays or generate stable transgenic lines [15,41,119–121]. For transient assays, often *Agrobacterium tumefaciens*-mediated transient assays (ATTA) are performed, usually in tobacco leaves [119]. One, however, preferably performs reporter assays in the plant species and tissues enhancer candidates are derived from, as TFs and TF binding motifs may not be conserved between species [121–123]. Methods such as particle bombardment and protoplast transformation allow testing of putative enhancer candidates in their own genetic background [124,125].

**Concluding remarks and future perspectives**

Similar to the situation in other organisms, enhancers play a crucial role in gene regulation in plants. Unlike plant enhancers, animal enhancers are very well characterized for their general properties. Based on features of the few well-characterized enhancers in plants, it appears that plant and animal enhancers share several characteristics, such as high chromatin accessibility, enrichment in histone acetylation and low DNA methylation levels. These shared features can be used for the discovery of new enhancers and subsequent in-depth characterization of the properties of plant enhancers, in active and inactive states. It will be interesting to find out
whether eRNAs also have a role in enhancer function in plants. We want to stress that more genome-wide characterization of enhancers is required to determine which combination of marks is best to identify enhancers in different activity states and discriminate enhancers from other cis-regulatory elements, such as silencers and insulators. For the time being, validation of candidate enhancer sequences is necessary.

In this review we discussed the advantages and drawbacks of different techniques used to identify and characterize enhancers. Importantly, a combination of techniques offers a higher accuracy for genome-wide enhancer detection than a single method. As an example of an innovative genome-wide technique, we highlight STARR-seq, which uses the power of high-throughput sequencing to quantitatively assess enhancer activity [96]. Other revolutionary techniques such as CRISPR-Cas9 [126] allow to test the functionality of putative enhancer sequences in vivo, circumventing issues associated with transgenic reporter assays (e.g. transgene silencing and position effects [127]). Application of these novel techniques can greatly contribute to the identification and characterization of plant enhancers. Meanwhile, differences between plant species and their genomes and how these may impact the performance of a specific technique need to be considered. In conclusion, we expect the knowledge concerning transcriptional regulation by enhancers, and thereby the knowledge on the regulatory potential in plants, to increase significantly in the very near future, and believe that many plant scientists will strongly benefit of such insights.

Acknowledgements

We gratefully acknowledge the financial support of the European Commission Seventh Framework-People-2012-ITN Project EpíTRAITS (Epigenetic regulation of economically important plant traits, no-316965) for B. Weber, J. Zicola, R. Oka and M. Stam.
Chapter 1

References

20. McGarry, R.C. and Ayre, B.G. (2008) A DNA element between At4g28630 and At4g28640 confers companion-cell specific expression following the sink-to-source transition in mature minor vein phloem. Planta 228, 839–849
Chapter 1

27 Aran, D. et al. (2013) DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol.* 14, R21
33 Liu, L. et al. (2014) Induced and natural variation of promoter length modulates the photoperiodic response of FLOWERING LOCUS T. *Nat. Commun.* 5, 4558
37 Karmodiya, K. et al. (2012) H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics* 13, 424
Chapter 1

46 Kumar, S. and Bucher, P. (2016) Predicting transcription factor site occupancy using DNA sequence intrinsic and cell-type specific chromatin features. *BMC Bioinformatics* 17, 41–50
56 Nagarajan, S. et al. (2015) H4K12ac is regulated by estrogen receptor-alpha and is associated with BRD4 function and inducible transcription. *Oncotarget* 6, 7305–7317
58 Hoffmann, B.G. et al. (2010) Locus co-occupancy, nucleosome positioning, and H3K4me1 regulate the functionality of FOXA2-, HNF4A-, and PDX1-bound loci in islets and liver. *Genome Res.* 20, 1037–1051
Chapter 1

72 Splinter, E. et al. (2006) CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. Genes Dev. 20, 2349–2354
74 Drissen, R. et al. (2004) The active spatial organization of the beta-globin locus requires the transcription factor EKLF. Genes Dev. 18, 2485–2490
75 Cao, S. et al. (2014) A Distal CCAAT/NUCLEAR FACTOR Y Complex Promotes Chromatin Looping at the FLOWERING LOCUS T Promoter and Regulates the Timing of Flowering in Arabidopsis. Plant Cell 26, 1009–1017

Song, L. et al. (2011) Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* 21, 1757–1767


Kaufmann, K. et al. (2010) Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-ChIP). *Nat. Protoc.* 5, 457–472


Melgar, M.F. et al. (2011) Discovery of active enhancers through bidirectional expression of short transcripts. *Genome Biol.* 12, R113


Chapter 1

112 van de Werken, H.J.G. et al. (2012) Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat. Methods 9, 969–972
121 Belele, C.L. et al. (2013) Specific Tandem Repeats Are Sufficient for Paramutation-Induced Trans-Generational Silencing. PLoS Genet. 9, e1003773
122 Copley, R.R. et al. (2007) Functional conservation of Rel binding sites in drosophilid genomes. Genome Res. 17, 1327–1335
128 Raatz, B. et al. (2011) Specific expression of LATERAL SUPPRESSOR is controlled by an evolutionarily conserved 3′ enhancer. Plant J. 68, 400–412

