

Characterization of transcriptional enhancers
in plants using *Arabidopsis thaliana* and *Zea
mays* as model species

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Johan Zicola

aus Nîmes (Frankreich)

Köln, 2017



Max Planck Institute
for Plant Breeding Research

Die vorliegende Arbeit wurde am Max-Planck-Institut für die Pflanzenzüchtungsforschung in Köln in der Abteilung für Entwicklungsbiologie der Pflanzen (Direktor Prof. Dr. G. Coupland) angefertigt.

Berichterstatter: Prof. Dr. George Coupland

Berichterstatter: Prof. Dr. Ute Höcker

Prüfungsvorsitzender: Prof. Dr. Juliette de Meaux

Tag der Disputation: 06.06.2016



MAX-PLANCK-GESELLSCHAFT

Abstract

Enhancers are non-coding regions of the genome controlling the transcription of genes located at distances ranging from kilo to mega bases. Enhancers act mainly as binding platforms for multiple transcription factors and were shown to interact directly with the promoter of their target genes. Since few enhancers are characterized in plants, we developed a pipeline to find new enhancers based on previous knowledge from plant and animal case studies. To do so, the two plant models *Arabidopsis thaliana* and *Zea mays* were used. Both functional and genome-wide approaches were used to study known enhancers in *A. thaliana* and *Z. mays* and identify putative enhancers in *Z. mays*. The characterization of novel enhancers in plants provides more insight about gene regulatory networks in plants. *In fine*, a better understanding of gene regulation could help improving important traits of cultivated plants by conventional breeding or biotechnological improvement.

In *A. thaliana*, enhancers of the *FLOWERING LOCUS T (FT)* were characterized by inducing DNA methylation using Inverted Repeats (IRs). Indeed, DNA methylation was often associated with the inactivation of *cis*-regulatory elements such as enhancers and promoters in mammals and for some cases in plants. DNA methylation deposition at the previously characterized enhancer *Block C* led to downregulation of *FT* expression and subsequent late flowering phenotype in *FT*-inducing growth conditions. A novel putative enhancer of *FT*, named *Block E*, was found downstream of the gene and shared several features with *Block C* such as accessible chromatin, conserved sequences among Brassicaceae, and putative transcription factor binding sites present at *Block C*. Transgenic lines containing an IR targeting a part of *Block E* displayed a late flowering phenotype, which was not as strong as for *Block C*, but significantly higher than other lines used as control and displaying a mild late flowering phenotype. We conclude that IR-targeted DNA methylation is a useful tool that can allow characterization of known enhancers and discovery of putative ones.

In *Z. mays*, differential chromatin accessibility, Histone 3 lysine 9 acetylation (H3K9ac) enrichment level, and gene expression profile were obtained for two different tissues (young seedling leaves and husk) in order to define enhancers' location, activity, and associated genes. Enhancers were previously shown to be in most cases located in accessible chromatin, and to contain associated histone marks such as H3 lysine 27 acetylation (H3K27ac), H3 lysine 4 monomethylation (H3K4me1), or H3K9ac. We first verified whether our data could properly define known or putative enhancers and the expression of their target genes. Indeed, we could find both high chromatin accessibility and H3K9ac enrichment level for the known enhancers of

teosinte branched1 (tb1) and *booster1 (b1)*, and the corresponding expression fold change of the target genes across the two tissues. By combining chromatin accessibility and H3K9ac enrichment level data, we could determine ≈ 2000 candidate enhancers in intergenic regions. We could finally associate for each tissue about 20 differentially expressed genes to about 20 putative tissue-specific enhancers, which will be characterized by transient expression assays in the future.

Zusammenfassung

Enhancer sind nicht-kodierende Regionen des Genoms, die die Transkriptionsrate von Genen kontrollieren, von denen sie einige Tausenden bis zu Millionen von Basenpaaren entfernt sind. Enhancer sind Bindungsplattformen für mannigfache Transkriptionsfaktoren, und es ist nachgewiesen worden, dass Enhancer physisch direkt mit Zielgenpromotoren interagieren können. Da über Enhancer in Pflanzen wenig bekannt ist, war es Ziel dieser Arbeit, ein Arbeitsprogramm zu entwickeln, das neue Enhancer aufgrund aus durch Modellstudien in Pflanzen und Tieren bekannter gemeinsamer Eigenschaften aufgespürt. Um dies zu erreichen, wurden Untersuchungen an den Pflanzen *Arabidopsis thaliana* und *Zea mays* durchgeführt. Sowohl funktionelle also auch genomweite Ansätze dienten dazu bekannte Enhancer aus beiden Modellen weiter zu untersuchen mit dem Ziel, neue Enhancer in *Z. mays* zu beschreiben. Charakterisierung neuer Enhancer hilft, die regulatorischen Gennetzwerke besser zu verstehen. Im Endeffekt kann solch besseres Verstehen der Genregulation helfen, Nutzpflanzen entweder durch konventionelle oder molekulare Züchtung in ertragsrelevanten Eigenschaften zu verbessern.

In *A. thaliana* wurden Enhancer des *FLOWERING LOCUS T (FT)* durch künstliche Methylierung untersucht, die durch sogenannte „inverted Repeats (IRs)“ gezielt an diesen induziert wurde. Der Zusammenhang zwischen DNA-Methylierung und der Inaktivierung von Enhancern und Promotoren ist in Säugetieren und auch einigen Fällen in pflanzlichen Modellen gut beschrieben. Methylierung des bekannten *FT* Enhancers *Block C* führte zur Reduktion der *FT* Expression, die mit einer Verzögerung des Blühzeitpunkts, unter Anzuchtsbedingungen in denen *FT* induziert wird, einherging. Ein neuer möglicher *FT* Enhancer, jetzt als *Block E* bezeichnet, wurde abwärts des *FT* Gens gefunden. *Block E* ähnelt *Block C* in vielen Eigenschaften wie dem Auftreten einer offenen Chromatinstruktur, Konservierung der Sequenz innerhalb der Brassicaceae und der Präsenz mehrerer gemeinsamer Transkriptionsfaktorbindestellen. Auch transgene Linien, die *Block E* überlappende IRs exprimierten, blühten signifikant später als Kontrollen, wenn auch nicht so spät, wie die entsprechenden *Block C* Linien. Zusammenfassend kann gesagt werden, dass IR-induzierte Methylierung ein Mittel ist, um Enhancer zu validieren und weiter zu untersuchen.

In *Z. mays* wurden genomweiter Profile von offenem Chromatin, Histone H3 Acetylierung an Lysin 9 (H3K9ac), sowie Expressionsprofile in zwei Geweben (junge Pflanzen und Kolbenblätter) erstellt, um Enhancerkandidaten zu lokalisieren und Zielgenen zuzuordnen. Es war beschrieben, dass Enhancer oft in offenem Chromatin lokalisieren und mit Histonmodifikationen H3 Acetylierung an Lysine 27 (H3K27ac), H3 Monomethylierung and Lysine 4 (H3K4me1) oder

H3K9ac assoziiert sind. Als erstes wurde validiert, welche dieser Modifikationen an bekannten Enhancern in *Z. mays* auftreten. Offenes Chromatin und H3K9ac Anreicherung waren an den beschriebenen Enhancern der *teosinte branched1 (tb1)* and *booster1 (b1)* Loci vorhanden, deren Expression auch in beiden untersuchten Geweben unterschiedlich war. Im Weiteren wurden durch Überlappen von offenen Chromatin- und H3K9ac-Profilen ca. 2000 Enhancerkandidaten in intergenischen Regionen aufgezeigt. Von diesen konnten ca. 20 differentiell offen und acetylierte Enhancer mit differentiell exprimierten Genen assoziiert werden. Diese Gruppe von Enhancerkandidaten soll in weiterführenden Studien experimentell validiert werden.

Table of contents

Abstract	I
Zusammenfassung	III
Table of contents	V
1 Introduction	1
1.1 One type of <i>cis</i> -regulatory elements: transcriptional enhancers	1
1.2 Importance of enhancers	2
1.3 Function of enhancers	2
1.4 Origin of enhancers	3
2 Characteristics of enhancers	4
2.1 Sequence conservation	4
2.2 Chromatin accessibility	4
2.3 Specific histone marks	5
2.4 DNA methylation	6
3 Identification of enhancers	7
3.1 DNase-seq	7
3.2 ChIP-seq	7
4 Aim of the study	8
Part I - Induced-DNA methylation represses the activity of two distinct regulatory regions of <i>FLOWERING LOCUS T</i>	9
5 Introduction	9
5.1 <i>FLOWERING LOCUS T</i>	9
5.2 Inverted repeat silencing	10
6 Materials and methods	11
6.1 Cloning	11
6.2 Transgenic plants	12
6.3 Culture conditions	12

6.4 Gene expression analysis	13
6.5 Bisulfite sequencing.....	14
6.6 smRNA sequencing.....	15
6.7 Chromatin Immunoprecipitation.....	15
6.8 Statistical analysis.....	16
7 Results.....	17
7.1 IR-induced late flowering	17
7.2 DNA methylation at <i>Block C</i>	19
7.3 H3K9me2 association with DNA methylation deposition.....	22
7.4 Newly identified regulatory element <i>Block E</i>	23
8 Discussion	25
Part II - Identification of putative enhancers in <i>Zea mays</i>	28
9 Introduction	28
9.1 <i>Zea mays</i> as genetic model.....	28
9.2 Enhancer of <i>bl</i>	28
9.3 Enhancer of <i>tbl</i>	29
9.4 Enhancer of <i>pl</i>	30
9.5 Putative enhancer of <i>ZmRap2.7</i>	30
10 Materials and methods.....	32
10.1 Culture conditions	32
10.2 Material harvest.....	32
10.3 DNase-seq.....	33
10.3.1 Nuclei preparation	33
10.3.2 DNase I digestion	33
10.3.3 Naked DNA control.....	34
10.3.4 DNA quantification	35
10.3.5 Library preparation and sequencing	36

10.4 ChIP-qPCR.....	37
10.5 ChIP-seq.....	39
10.6 RNA-seq.....	39
10.7 Bioinformatic analyses.....	41
10.7.1 Uniqueome.....	41
10.7.2 Read quality processing.....	41
10.7.3 Mapping.....	42
10.7.4 Peak calling.....	42
10.7.5 RNA-seq analysis	43
11 Results	44
11.1 Mappability of <i>Z. mays</i> genome.....	44
11.2 Chromatin accessibility profile in <i>Z. mays</i>	45
11.2.1 DNase I assay	45
11.2.2 DNase-seq analysis.....	47
11.2.3 DHSs are enriched at promoters and coding regions.....	49
11.3 Active histone mark H3K9ac profile in <i>Z. mays</i>	51
11.3.1 Validation of the histone mark H3K9ac at the <i>b1</i> locus	51
11.3.2 Genome-wide H3K9ac enrichment profile in <i>Z. mays</i>	54
11.4 Transcriptome analysis.....	56
11.5 Data validation at known enhancers	58
11.5.1 Differential gene expression	58
11.5.2 Chromatin accessibility and H3K9ac enrichment profiles	59
11.6 Genome-wide identification of enhancers.....	62
11.6.1 Data validation.....	62
11.6.2 Enhancer prediction.....	65
12 Discussion	69
13 Conclusion.....	72

14 Supplementary data	73
15 Literature	84
16 Abbreviations	98
17 Acknowledgments	102
Erklärung	103
18 Lebenslauf	104

1 Introduction

1.1 One type of *cis*-regulatory elements: transcriptional enhancers

One hallmark of multicellular organisms is the ability of different cells derived from a unicellular spore or zygote to cooperate and fulfill different tasks within the same organism (Bonner, 1998, 2000). As a result, cells specialize and acquire a specific repertoire of expressed genes that varies from one cell type to the other. This differential regulation of genes within different cells is modulated *Cis*-Regulatory Elements (CREs) which act as binding platforms for multiple Transcription Factors (TFs). Gene expression is therefore determined by the binding of TFs within CREs, encoding together a specific transcriptional output (Davidson, 2006). CREs encompass promoters, transcriptional enhancers (hereafter enhancers), silencers, and insulators (Jeziorska et al., 2009). CREs are typically located directly upstream of genes, i.e. in promoter regions, but can also be present in introns, Untranslated Regions (UTRs), downstream and upstream distal regions, and even in coding regions (Wray et al., 2003). The term enhancer was firstly used for a viral sequence of the Simian Virus 40 (SV40) able to increase the transcription, in an orientation and distance-independent manner, of the β -globin gene in transient expression assays (Banerji et al., 1981; Moreau et al., 1981). The term “*cis*-regulatory module” was also proposed as replacement for “enhancer” since it could gather CREs acting either as stimulator or repressor of gene transcription (Davidson, 2001). However, the term enhancer is now used strictly for elements promoting gene expression while silencers and insulators are representing distinct categories of CREs (Maston et al., 2006). Another semantic issue derives from the dichotomy made between promoters and enhancers. This dichotomy was questioned by recent studies that found many similarities between these two CREs: some promoters can act as enhancers for distal genes (Li et al., 2012; Leung et al., 2015), while some enhancers can initiate transcription independently of promoters (Vernimmen et al., 2007; Kowalczyk et al., 2012). A unified model was proposed in which enhancers and promoters would be considered as a single class of functional elements that recruit and initiate RNA Polymerase II (Pol II) at varying rates in a context-dependent manner (Andersson et al., 2015). Nonetheless, the term enhancers is still useful as it implies their distal nature compared to promoters that are usually adjacent to their target gene. It is of course arbitrary to set a limit distance from genes’ Transcriptional Start Sites (TSSs) to separate a promoter element from an enhancer but semantically ambiguous CREs can be considered correctly either as proximal enhancers or distal promoter elements.

1.2 Importance of enhancers

Enhancers were predicted to outnumber genes in *Drosophila* with 50,000 to 100,000 putative enhancers estimated (Kvon et al., 2014) and in human with over 43,000 identified enhancers (Andersson et al., 2014). Enhancers have an essential role in development and several mutations at enhancers were shown to be associated with a number of human diseases (Maurano et al., 2012; Khurana et al., 2013; Herz et al., 2014). In plants, although few enhancers are known compared to the animal field, several enhancers or putative enhancers were shown to be associated with important developmental traits such as apical dominance (Doebley et al., 1997) and flowering time in *Z. mays* (Salvi et al., 2007) and in *A. thaliana* (Adrian et al., 2010). These enhancers will be further introduced in the two parts of this thesis with the first part focusing on *A. thaliana* and the second part on *Z. mays*.

1.3 Function of enhancers

Enhancers are located in intergenic regions but also in introns, either of the gene they regulate or another gene. For instance, one enhancer of the *Sonic hedgehog* (*Shh*) gene in mammals is located in the intron of a gene 1 Mb upstream of the target gene (Lettice et al., 2002). Kvon et al., (2014) have shown that approximately 50% of the enhancers in *Drosophila* locate directly up or downstream of the genes and 36% locate in introns. UTRs were also shown to contain enhancers though their activity can be at a transcriptional as well as at a post-transcriptional level. For instance, transient expression experiments showed that *UbiquitinC* in human requires its 5'-UTR intron for proper expression (Bianchi et al., 2009). In mouse, the 3' UTR of *Col2a1* was shown to contain an enhancer that interacts with the *Col2a1* promoter in a permissive chromatin context (Jash et al., 2012). In Arabidopsis, introns located in UTRs are mostly present in the 5'UTR and can contain regulatory elements as it was shown for *ELONGATION FACTOR 1 α -A3* (*EF1 α -A3*) (Chung et al., 2006). The activation of the target gene by its enhancer is commonly thought to be based on the interaction between the enhancer and the target gene promoter through the binding of TFs at the two CREs, leading to the recruitment of intermediary co-activators, which *in fine* recruit Pol II to initiate transcription (Koch et al., 2011) (**Figure 1**). Although long-range interactions were already speculated at the discovery of the enhancer of SV40 (Banerji et al., 1981), the evidence of such interactions were brought only two decades later with the development of Chromosomal Conformation Capture (3C) techniques (Dekker et al., 2002). Examples of 3C studies in plants will be further developed in this thesis.

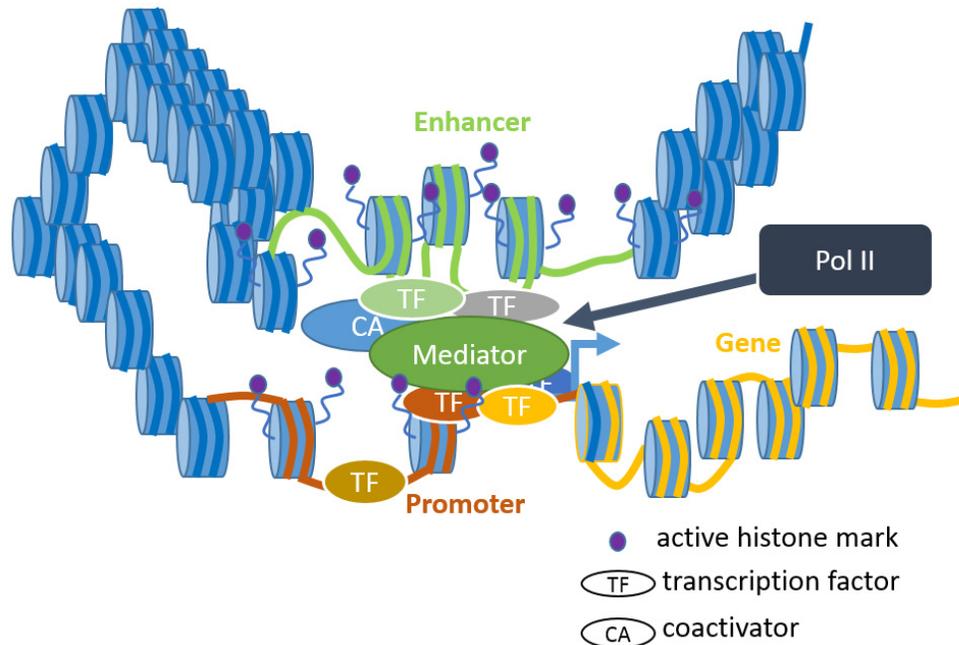


Figure 1: Simplified model of enhancer-promoter interaction. General TFs (TF, oval shapes) are recruited at the core promoter and specific TFs are recruited at both promoter and enhancer. Coactivators can mediate the interaction between TFs. The mediator complex (green oval shape) relays the signal from the different TFs and coactivators (CA, oval shape) to Pol II in order to initiate transcription. The enhancer and the promoter are located in a region poor in nucleosomes (blue cylinders), allowing TFs binding. Besides, chromatin can have a higher order of compaction. Also, nucleosomes flanking or within active enhancers and promoters can carry specific modifications at their histone tails (purple circles).

1.4 Origin of enhancers

The evolution of enhancers is not well understood but several possible mechanisms were suggested such as *de novo* evolution, transposition, co-option and promoter switching (Rebeiz et al., 2011). Chromosomal transposition can also be a source of new regulation for genes by being juxtaposed to an enhancer coming from another chromosome. For instance, translocation of oncogenes such as *c-myc* close to the enhancer of the immunoglobulin heavy-chain locus is involved in cancer development in humans (Dalla-Favera et al., 1982; Taub et al., 1982; Adams et al., 1985). Transposable Elements (TEs) can also be a source of evolution for enhancers. In 1969, Britten and Davidson suggested that TEs could be involved in gene regulation. A decade later, the term “exaptation” was proposed by Gould and Vrba (1982) to define the new regulatory functions TEs can acquire during evolution. Many lines of evidence accumulated since then to support this theory and TEs were shown to carry regulatory sequences and to be able to rewire regulatory networks by inserting nearby genes (Britten, 1997; Bejerano et al., 2006; Santangelo et al., 2007). The *Alu* elements in primates can act in some cases as enhancers (Norris et al., 1995; Vansant and Reynolds, 1995; Polak and Domany, 2006). In plants, few examples of enhancers were described but TEs are

known to be involved in the enhancer-mediated regulation of *ZmRap2.7* and *tb1* genes in *Z. mays* (Salvi et al., 2007; Studer et al., 2011). In addition to TEs, exaptation can also encompass nonfunctional regions of DNA that contain Transcription Factor Binding Sites (TFBSs) that may evolve by mutations or insertions/deletions into new CREs (Stone and Wray, 2001).

2 Characteristics of enhancers

2.1 Sequence conservation

Enhancers are DNA sequences containing multiple TFBSs. Many enhancers are under purifying selection and can be identified according to their degree of conservation in related species (Visel et al., 2007). To determine conserved regions, sequence comparisons, termed phylogenetic footprinting and shadowing, can be performed either between distant species or related species, respectively (Tagle et al., 1988; Gumucio et al., 1992; Boffelli et al., 2003). Conserved Non-coding Sequences (CNSs) were identified at the putative enhancer of *ZmRap2.7* in *Z. mays* although no TFBSs have been identified (Salvi et al., 2007). In *A. thaliana*, phylogenetic shadowing was used to determine CREs in the intron of *AGAMOUS*, although the method could not find a strong conservation at two known TFBSs (Hong et al., 2003). Also, several CNSs were found around *FT* but only one, excluding the proximal promoter, was shown so far to have a *cis*-regulatory function (Adrian et al., 2010). In *Drosophila*, enhancers were shown to diverge importantly between close species, although they retain the same activity (Ludwig and Kreitman, 1995; Ludwig et al., 1998). Individual TFBSs can exhibit little conservation, either because of the degenerate nature of TFBSs or because their small size makes compensatory mutations likely (Ludwig et al., 2000; Dermitzakis and Clark, 2002). Sequence conservation can therefore be a useful tool to determine CNSs with putative regulatory functions but does not highlight CREs that are under neutral selection.

2.2 Chromatin accessibility

Chromatin accessibility varies depending on the local nucleosome occupancy and binding of other chromatin-associated proteins. The degree of accessibility affects gene transcription with accessible regions being more transcribed than closed ones (Weintraub and Groudine, 1976). Accessible regions, also called Nucleosome-Depleted Regions (NDRs), contain CREs such as promoters and enhancers (Thurman et al., 2012). Chromatin accessibility at enhancers allows binding of TFs or other regulatory factors to DNA instead of nucleosomes (Hesselberth et al., 2009; Sherwood et al., 2014). In plants, NDRs were shown to be enriched in TFBSs in both *A. thaliana* and *Z. mays* (Hesselberth et al., 2009; Vera et al., 2014). Finally, several intergenic NDRs identified in *A. thaliana* were validated as enhancers using reporter assays (Zhang et al., 2012b; Zhu et al., 2015).

2.3 Specific histone marks

Histone marks are post-translational modifications of histones that modulate chromatin accessibility (Bannister and Kouzarides, 2011). Nucleosomes within or flanking enhancers display specific histone marks. In mammals, H3K4me1 was found at both active and inactive enhancers (Heintzman et al., 2007). H3K27ac, H3K9ac, and H3 lysine 14 acetylation (H3K14ac) were associated with active enhancers (Creyghton et al., 2010; Heintzman et al., 2009; Karmodiya et al., 2012) (**Figure 2**). In contrast, H3 lysine 27 trimethylation (H3K27me3) and H3 lysine 9 trimethylation (H3K9me3) were associated with poised or/and inactive enhancers (Rada-Iglesias et al., 2011; Zentner et al., 2011). In *Z. mays*, the active enhancer of *b1* was shown to be enriched in H3K9ac and H3K14ac (Haring et al., 2010). In *A. thaliana*, a recent study revealed a positive correlation between inactive enhancers and H3K27me3; however, the association of active enhancers with H3K27ac was less clear (Zhu et al., 2015). Finally, intergenic NDRs in rice are strongly associated with Histone 4 lysine 12 acetylation (H4K12ac) (Zhang et al., 2012a), an histone mark also shown to be associated with active enhancers in mammals (Nagarajan et al., 2015). In conclusion, plant enhancers seem to be generally associated with H3 and H4 acetylation marks.

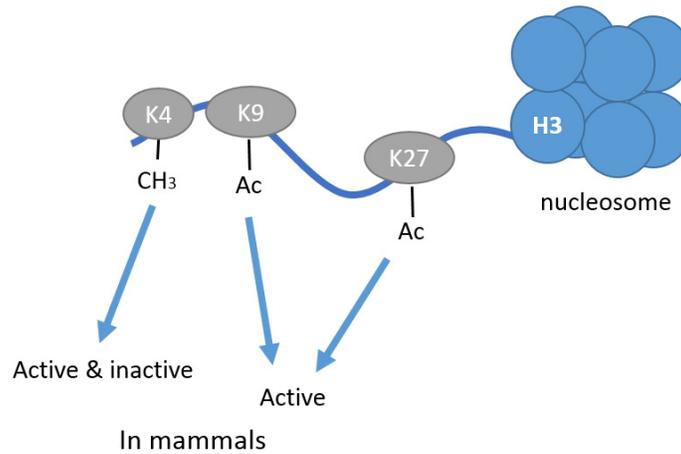


Figure 2: Acetylation of amino-terminal tail of histone H3 at lysine in position 4, 9, and 27 were shown to be enriched at enhancer locations in mammals. More specifically, H3K4me1 was associated with both active enhancers while H3K27ac and H3K9ac were associated with active enhancers.

2.4 DNA methylation

DNA methylation is associated with transcriptional silencing in plants and animals (Law and Jacobsen, 2010). In mammals, DNA methylation at enhancers was associated with downregulation of the target genes (Stadler et al., 2011; Blattler et al., 2014). Furthermore, changes in DNA methylation pattern at enhancers can be involved in oncogenesis (Aran et al., 2013). The dynamics between changes in DNA methylation at enhancers and gene expression was used to identify promoters and enhancers in different cell lines in mice and humans (Stadler et al., 2011; Hwang et al., 2015). In plants, DNA methylation at promoters and enhancers was also associated with repression of gene expression, e.g. *FLOWERING WAGENINGEN (FWA)*, *TOO MANY MOUTHS (TMM)*, and *FT* in *A. thaliana* (Kinoshita et al., 2007; Deng et al., 2014; Deng and Chua, 2015); *p1* and *b1* in *Z. mays* (Sidorenko and Peterson, 2001; Haring et al., 2010). In *Solanum lycopersicum*, fruit ripening was associated with demethylation at promoters, allowing the TF RIPENING INHIBITOR (RIN) to bind and activate genes involved in ripening (Zhong et al., 2013). However, there is not yet evidences suggesting dynamic changes in DNA methylation at enhancers associated with changes in gene expression in plants such as observed in animals. In plants, *de novo* DNA methylation is established by the RNA-dependent DNA Methylation (RdDM) pathway. Briefly, a single-stranded RNA (ssRNA) produced by the plant-specific RNA Polymerase IV (Pol IV) is processed into a double-stranded RNA (dsRNA) by RNA-dependent RNA polymerase 2 (RDR2). The dsRNA is then cut into 24-nucleotides (nt) small RNAs (smRNAs) by DICER-LIKE 3 (DCL3) that are methylated at their 3'-ends by HUA ENHANCER 1 (HEN1) to be loaded into ARGONAUTE 4 (AGO4). AGO4 is recruited through AGO4-bound smRNAs base-pairing with complementary Pol V-generated scaffold transcripts. AGO4 interacts indirectly with DOMAIN REARRANGED METHYLTRANSFERASE 2 (DMR2), which establishes DNA methylation in all cytosine contexts at the region of homology with the initial Pol IV-generated RNA (Matzke et al., 2015). It was shown that Pol II was able, as Pol V, to generate scaffold transcripts and interact with AGO4-siRNAs to induced DNA methylation (Zheng et al., 2009). Note that Pol IV and Pol V are plant-specific polymerase subunits (Haag and Pikaard, 2011).

3 Identification of enhancers

Several methods can be used to identify enhancers genome-wide (Shlyueva et al., 2014). In this thesis, two of them are exploited in *Z. mays*: DNase-seq and ChIP-seq.

3.1 DNase-seq

Active enhancer sequences are usually located in regions of low nucleosome occupancy and are therefore sensitive to nuclease activity (Gross and Garrard, 1988). DNase I Hypersensitive Sites (DHSs) can be determined by partial digestion of chromatin with the endonuclease DNase I followed by sequencing of the small fragment representing the accessible fraction of the genome (DNase-seq) (Hesselberth et al., 2009). Alternatively, the ends of large digested DNA fragments can be sequenced (Boyle et al., 2008). In addition to DHSs, TFBSs can also be identified from DNase-seq data (Neph et al., 2012). However, the intrinsic cleavage bias of DNase I was shown to affect TFBS prediction (He et al., 2014). Additionally, the length of the digested fragments sequenced can either highlight TF occupancy (<200bp) or nucleosome architecture (200-400bp) (Vierstra et al., 2014).

3.2 ChIP-seq

Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) allows to identify DNA regions based on their associated modifications or proteins (e.g. histone marks, TFs and Pol). The principle is based on the recovery of protein-bound DNA sequences using an antibody recognizing the DNA-associated protein of interest. As mentioned before, several histone marks can define enhancers such as H3K4me1 or H3K27ac (Heintzman et al., 2007; Creighton et al., 2010). In addition, TFs and co-activators such as CBP/p300 can be used to identify enhancers (Visel et al., 2009). However, the most relevant combination of histone marks or TFs for enhancer identification in plants remain to be determined.

4 Aim of the study

The present study aims to identify and characterize enhancers in plants using *A. thaliana* and *Z. mays* as models. Few enhancers are described in plants compare to other organisms such as *Drosophila* or mice. However, some of these plant enhancers are known to be involved in important agronomical traits such as flowering time or apical dominance. The regulation of the florigen *FT* in *A. thaliana* is one example of a complex regulation mediated by CREs, including an upstream enhancer linked to photoperiod-dependent induction. In the other plant model, *Z. mays*, confirmed and putative enhancers were shown to regulate the expression of genes involved in domestication such as *tb1*, in flowering such as *ZmRap2.7*, or in pigmentation such as *b1* and *pericarp color1 (p1)*. We think that many other enhancers remain to be discovered. The rise of high-throughput sequencing technologies and the recent knowledge gained in the field of chromatin and epigenetics offer new tools to identify and characterize novel CREs, including enhancers. This thesis will be divided into two independent parts:

In the first part, an enhancer silencing approach based on IR-mediated DNA methylation was established using the enhancer *Block C* at the *FT* locus. The screening of late flowering phenotypes in transgenic lines, changes in *FT* expression, and associated DNA methylation was performed across four generations. Furthermore, the approach was used to survey other potential CREs at the *FT* locus. The outcome of this research would be to provide an interesting tool to characterize enhancers in plants.

In the second part, genome-wide profiles of chromatin accessibility and enhancer-associated histone mark were generated and combined to identify candidate enhancers in maize. A subset of these candidates will be cloned to test their activity through transient expression assays. Furthermore, stable transgenic lines of *Z. mays* were generated to target IR-mediated DNA methylation at a putative enhancer and confirm its activity. The outcome of this research would be to provide new insight on maize *cis*-regulation and make available to the scientific communities a list of putative enhancers for different genes that can be involved in important developmental traits.

Part I - Induced-DNA methylation represses the activity of two distinct regulatory regions of *FLOWERING*

LOCUS T

5 Introduction

5.1 *FLOWERING LOCUS T*

The *FLOWERING LOCUS T* (*FT*) locus encodes a transcription cofactor produced in the phloem companion cells in leaves and that migrates through the phloem towards the shoot apical meristem to trigger the floral transition by activating several meristem identity genes such as *APETALAI* (*API*) and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*SOCI*) (Andrés and Coupland, 2012; Bratzel and Turck, 2015). In the facultative Long Day (LD) plant *Arabidopsis thaliana*, *FT* expression is promoted by the transcription factor CONSTANS (CO) in a photoperiod-dependent manner (Andrés and Coupland, 2012). CO-mediated *FT* expression under LD was shown to require the presence of the proximal promoter and of a second regulatory region located 5.3-5.7 kb upstream of the Transcription Start Site (TSS) of *FT* (Adrian et al., 2010; Liu et al., 2014). The distal regulatory region, named *Block C*, is conserved across several species of Brassicaceae and contains several consensus Transcription Factor Binding Sites (TFBSs) (Adrian et al., 2010). Among the conserved TFBSs found at *Block C*, a conserved CCAAT box is required for full promoter activity (Cao et al., 2014). CCAAT box motifs are binding sites for trimeric Nuclear Factor Y (NF-Y) transcription factors (Kumimoto et al., 2008). CO is able to interact with NF-YB and NF-YC subunits to form a trimeric complex thereby potentially replacing NF-YA, but the binding of this complex to the CCAAT motif has not been demonstrated (Ben-Naim et al., 2006; Wenkel et al., 2006). *In vitro*, CO can bind directly to a loosely defined TGTG(N2-3)ATG consensus motif present in several copies at *FT* proximal promoter (Tiwari et al., 2010). Mutational analysis performed for this consensus motif indicated that it is required for induction of reporter genes under the control of the full-length *FT* promoter (Adrian et al., 2010). *Block C* was shown to be depleted in the repressive histone mark Histone 3 lysine 27 trimethylation (H3K27me3) that otherwise covers the *FT* locus including its upstream and downstream intergenic regions (Adrian et al., 2010) and its location corresponds to a DNase I Hypersensitive Site (DHS), which defines accessible chromatin (Zhang et al., 2012b). Altogether, *Block C* characteristics indicate that it acts as a distal transcriptional enhancer of *FT*. Enhancers act mainly as binding platforms for multiple transcription factors and were shown to interact directly with the promoter of their target genes

where they recruit the transcription machinery (Shlyueva et al., 2014). Two recent studies applied Chromosomal Conformation Capture (3C) to measure interactions between regulatory regions at the *FT* locus (Liu et al., 2014; Cao et al., 2014). The first study found predominant interaction of both, *Block C* and the TSS of *FT* with a region located 4 kb upstream of the TSS, while direct interaction between *Block C* and the TSS was rather low (Liu et al., 2014). The second study showed different interactions, one taking place between the TSS and a region 2 kb upstream and the other taking place between the TSS and *Block C* (Cao et al., 2014). Therefore, the contribution of *Block C* and other regulatory regions to *FT* expression in the native chromatin context still requires further analysis.

5.2 Inverted repeat silencing

In plants, the possibility to induce DNA methylation through the expression of Inverted Repeats (IR) provides a tool to alter gene activity by Transcriptional Gene Silencing (TGS) (Mette et al., 2000). DNA methylation at gene promoters or enhancers can affect gene expression by directly disrupting transcription factor binding in both plants and animals (Ehrlich and Ehrlich, 1993) or by recruiting repressor complexes that alter chromatin accessibility (Lewis et al., 1992; Meehan et al., 1992). In *Z. mays*, IR-mediated methylation of the enhancer of *booster1 (b1)* gene was associated with TGS (Sloan et al., 2014). In *A. thaliana*, IR-mediated methylation of the promoter of *FLOWERING WAGENINGEN (FWA)* induced TGS (Kinoshita et al., 2007), while targeting the first intron of *FT* was recently shown to increase *FT* expression and cause early flowering (Deng and Chua, 2015). The activation of *FT* expression was attributed to a decreased access of the repressive MADS-domain transcription factor *FLOWERING LOCUS C (FLC)* to its cognate binding site in the target region (Deng and Chua, 2015). We assumed that IR-mediated DNA methylation could affect the activity of the putative enhancer *Block C* and therefore lead to TGS of *FT*.

Here, we report the downregulation of *FT* expression by inducing IR-mediated DNA methylation at its enhancer *Block C*. DNA methylation is accompanied by an increase of H3 lysine 9 dimethylation (H3K9me2) at *Block C* in IR-containing plants, indicating a potential change of the chromatin state from accessible to close. We also show that symmetric DNA methylation level is partially maintained upon the IR loss but then stably retained through generations and associated with a mild late flowering phenotype. We bring an additional evidence that *Block C* is acting as an enhancer of *FT* under inductive photoperiod conditions. At last, we define a novel candidate enhancer, located downstream of *FT* and named *Block E*, which also affects *FT* expression when targeted by an IR.

6 Materials and methods

6.1 Cloning

IR constructs were produced using the Gateway system (Invitrogen). Target regions of interest were cloned with specific primers containing attB1 and attB2 sites (Invitrogen). The amplified product was sequenced and cloned into the Donor Vector pDONR207 (Invitrogen) through a BP reaction, according to Gateway's manual. The resulting plasmid was transformed into chemical competent *Escherichia coli* DH5- α strain (Invitrogen) using the heat shock method (Hanahan, 1983). The transformed bacteria were screened on Luria broth (LB) plate containing gentamicin (10 mg/l). The insert was then introduced by LR reaction into the destination vector pJawohl8-RNAi (AF408413, <http://www.ncbi.nlm.nih.gov/nuccore/AF408413>, Max Planck Institute for Plant Breeding Research). The recombined plasmid was introduced into *E. coli* DH5- α by heat shock and cultivated on LB plates containing ampicillin (10 mg/l). The plasmid of four resistant colonies were sent to sequencing with the proper primer to control insert sequence. The plasmid was then introduced into *Agrobacterium tumefaciens* strain GV3101 (pMP90RK) (Hellens et al., 2000) by electroporation and transformed bacteria were selected on LB plates containing the antibiotics rifampicin (50 mg/l), kanamycin (25 mg/l), gentamycin (10 mg/l), and carbenicillin (40 mg/l). The presence of the construct was verified by PCR and three independent validated colonies were pooled for plant transformation.

Primers with attB sites (in upper case) used to generate the different IR constructs and the genomic coordinates of the target regions (based on TAIR10 coordinates):

Block C (chr1:24325788..24326334)

F: 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTcatttgctgaacaaaatct-3'

R: 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTaaacgttggaataggaagtatg -3'

Col-0 insertion (chr1:24327957..24328261)

F: 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTttgatgacaaagggcactca-3'

R: 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTtagattggcaagtggatgagg-3'

Block B (chr1:24329393..24329770)

F: 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTaaaagacaagtggcagatacgtt-3'

R: 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTaaatattggacaggagactcag-3'

CaRG boxes III and IV (chr1:24330158..24330535) (data not shown in this thesis)

F: 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTtctacagttgtaggctatggtt-3'

R: 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTttttataacaagcgccata-3'

Block E (chr1:24334630..24335026)

F: 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTatgaaaacgctctcctccaa-3'

R: 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTggatgctggttacgagcataag-3'

6.2 Transgenic plants

Agroinfiltration was performed by floral dipping (Clough and Bent, 1998) using the *A. tumefaciens* strain GV3101 (pMP90RK) and Columbia (Col-0) as *A. thaliana* background accession. To select transgenic T1 plants, T1 seedlings at stage four true-leaves were sprayed every four days, four times in total with a solution of glufosinate (250 mg/l, BASTA, Bayer). About 30 resistant plants were kept to provide T2 seeds. T2 seeds were sterilized with ethanol before selection on Petri dishes (100 mm x 15 mm) containing Growth Medium (GM) [half-strength Murashige and Skoog (0.5x MS) medium, 0.7% agar] containing 50 µg/ml of glufosinate. Seeds were put in recycled RNA purification columns (Qiagen) and shaken with 700 µl of 75% ethanol for 7 min. The tubes were centrifuged to discard ethanol and shaken a second time 5 min with 700 µl of 100% ethanol. The sterilized seeds were sown on plates, stratified for 3-5 days at 4°C and transferred to Percival cabinets. After three weeks of growth at 24°C, 16h light/8h dark, resistance to glufosinate was scored to select lines with single locus insertions by 3:1 segregation. T3 plants homozygous for the transgene were screened on 0.5x GM plates containing 50 µg/ml of glufosinate. T3 lines that showed loss of glufosinate resistance were kept to check whether residual DNA methylation triggered by the IR in hemizygous T2 was present. To do so, T2 seeds resistant and sensitive to glufosinate were sown to obtain homozygous transgenic and non-transgenic T3. Two independent lines of T3 homozygous for the transgene and without the transgenes were used for further analysis.

6.3 Culture conditions

For Chromatin Immunoprecipitation (ChIP) and smRNA sequencing experiments, seeds were sterilized with ethanol, sown on Petri dishes (100 mm x 15 mm) containing GM, stratified for 3-5 days at 4°C, and transferred to Percival cabinets in Long Day conditions (LD; 16h light/8h dark) at 22°C. For gene expression analysis and flowering time analysis, seeds were sown on soil, stratified for 3-5 days at 4°C. For gene expression analysis, stratified seeds were transferred into Percival cabinets in LD conditions at 22°C. For flowering time analysis, stratified seeds were transferred either in Percival cabinets or greenhouses at 22°C. Either LD or Mid Day conditions (MD; 12h light/12h dark) were used.

6.4 Gene expression analysis

Aerial parts of two-week-old seedlings were collected at ZT16 into 2-ml microcentrifuge tubes containing two tungsten carbide beads and flash frozen into liquid nitrogen. Frozen samples were grinded using the TissueLyser (Qiagen) and RNA extraction was performed with TRIzol (ThermoFisher Scientific) following manufacturer's protocol. RNA was resuspended in 100 µl of RNase-free water and 5 µg was treated with DNase I with the DNA-free DNA Removal Kit (Ambion) according to manufacturer's protocol. RNA was split into two fractions of 2.5 µg to perform complementary DNA (cDNA) synthesis using SuperScript II Reverse Transcriptase (RT, ThermoFisher Scientific) and following manufacturer's protocol. RT enzyme was added to only one of the two fractions in order to verify by PCR whether genomic DNA (gDNA) contamination was present. After cDNA synthesis, samples were diluted two times and a control PCR was performed using primers amplifying a portion of the *PROTEIN PHOSPHATASE 2A SUBUNIT A3 (PP2A)* gene for the two fractions. An amplicon of 155 bp was found only for the fraction in which RT enzyme was added. Quantitative PCR (qPCR) on LightCycler480 (Roche) was performed using a 10 µl PCR mix [0.5 µl 20× EvaGreen dye (Biotium), 5 µl of ¼ cDNA dilution (≈ 40 ng), 0.5 µl forward primer and reverse primer (10mM), 0.1 µl Taq polymerase (Bio-Budget, Germany), 1 µl 10× Eva Buffer (0.7mM dNTPs, 0.14M Tris-HCl pH 8.8, 0.6M KCl, and 36mM MgCl₂), 2.4 µl dH₂O]. The amplification cycle was 95°C 10 s, 58°C 15 s, 72°C 20 s, followed by a melting curve analysis of 95°C 5 s, 65°C 1 min, increase from 65°C to 97°C with 2.5°C/s. Calibration curve was prepared for each primer using 10 µl of cDNA diluted 1/5, 1/25, 1/125, and 1/625 in order to calculate primer efficiency. *FT* expression relative to *PP2A* housekeeping gene expression was defined using the average of crossing point (CP) values calculated by the LightCycler480 for three technical replicates. Averaged CP values were used in Pfaffl's equation:

$$\text{ratio} = \frac{(E_{\text{target}})^{\Delta\text{CP}_{\text{target}}(\text{control-sample})}}{(E_{\text{ref}})^{\Delta\text{CP}_{\text{ref}}(\text{control-sample})}}$$

Defined as the ratio of *FT* target gene expressed in the line tested versus control Wild Type (WT) in comparison to the *PP2A* reference gene. E_{target} is the qPCR efficiency of the *FT* primers; E_{ref} is the qPCR efficiency of the *PP2A* primers; $\Delta\text{CP}_{\text{target}}$ is the CP deviation of control – sample of *FT* transcript; $\Delta\text{CP}_{\text{ref}}$ = CP deviation of control – sample of *PP2A* transcript (Pfaffl, 2001).

Primers *FT* cDNA (244 bp cDNA, 1896 bp genomic DNA (gDNA))

F: 5'- GGTGGAGAAGACCTCAGGAA -3'

R: 5'- ACCCTGGTGCATACACTGTT -3'

Primers *PP2A* (155 bp for both cDNA and gDNA)

F: 5'- CAGCAACGAATTGTGTTTGG -3'

R: 5'- AAATACGCCCAACGAACAAA -3'

6.5 Bisulfite sequencing

Ten-day-old seedlings were harvested at ZT16 and the aerial part sampled into 2-ml microcentrifuge tubes with two tungsten carbide beads and flash frozen into liquid nitrogen. Samples were grinded using the TissueLyser (Qiagen) and genomic DNA was extracted using DNeasy Plant Mini Kit (Qiagen). The gDNA was eluted in 100 µl elution buffer (Tris-HCl pH 8) and a 5 µl fraction was separated on agarose gel by electrophoresis for a quality check. Bisulfite conversion was performed using 20 µl of gDNA (≈200 ng) with the Epiect Bisulfite Kit (Qiagen, 59104) according to manufacturer's instructions with small modifications. The conversion cycle program was adjusted as described in Foerster and Mittelsten Scheid (Kovalchuk and Zemp, 2010) with a 5 min longer denaturation step at 95°C and two additional hours conversion at 60°C. The conversion efficiency was measured by amplifying the bisulfite-treated DNA and the non-treated gDNA with control primers described in Foerster and Mittelsten Scheid (Kovalchuk and Zemp, 2010). Regions of interest were amplified using degenerated primers designed with the Kismeth webtool (Gruntman et al., 2008), cloned into TOPO vectors using the TA cloning system (Invitrogen), and transformed into *E. coli* DH5-α. At least eight colonies were selected for amplification using forward and reverse M13 universal primers. The PCR products were purified by Polyethylene Glycol (PEG) precipitation. An equal volume of PEG solution (20% PEG 8000, 2.5M NaCl) and PCR reaction were mixed, incubated 15 min at 37°C, centrifuged at 4000 x g for 40 min at 25°C. The supernatant was removed and the DNA pellet was washed once with 70% ethanol, dried and resuspended in 20 µl of 10mM Tris-HCl pH 8.5 buffer. DNA was sent for sequencing using universal M13 reverse primer. The sequences were aligned using MEGA5.2 (Tamura et al., 2011) and the methylation analysis was performed using the webtool CyMATE (Hetzl et al., 2007). Primers used for the amplification of bisulfite converted DNA:

Conversion control region AT5G66750 (F1-R amplify non-converted DNA, F2-R amplify converted DNA) (≈561 bp)

F1: 5'-CGTCTGGTGATTCACCCACTTCTGTTCTCAACG-3'

F2: 5'-TGTTTGGTGATTTATTTATTTTTGTTTTTAATG-3'

R: 5'-CTCTCACTTTCTATCCCATCTA-3'

Block C 5' region (584 bp)

F: 5'-GGAATYAGTTYGAYTGAAATTATGT-3'

R: 5'-GTTGATGATAGTGAAGTGAGA-3'

Block C 3' region (342 bp)

F: 5'-TTTTATYTGATTTGGGGTTYAAAAA-3'

R: 5'-AACTTCAATTCATCATCTCTCTTT-3'

Col-0 insertion (349bp)

F: 5'-ATGGGAAGATAATTGAAGGATTAT-3'

R: 5'-ATCCRRCAAACCTACRTTTACAA-3'

Block B (359 bp)

F: 5'-ATTATGYGAYATATGGTGGTTAGAA-3'

R: 5'-ATAAAATTTAACRTATCTRCCACTT-3'

M13 universal primers

F: 5'-GTTTCCAGTCACGAC-3'

R: 5'-CAGGAAACAGCTATGAC-3'

6.6 smRNA sequencing

Total RNA was extracted from whole two-week-old seedlings grown in GM plates using TRIzol reagent (ThermoFisher Scientific) according to manufacturer's protocol. A quality check of the RNA was performed on the Agilent 2100 Bioanalyzer before and after the smRNA purification using PureLink miRNA Isolation Kit (Ambion) according to manufacturer's protocol. Sequencing was performed on Illumina Hi-seq 2500 to generate single-end 100 bp reads. Ten million reads were generated per library. The adapters were trimmed, and the reads were mapped on the Arabidopsis TAIR10 reference genome using BWA (Li and Durbin, 2009).

6.7 Chromatin Immunoprecipitation

ChIP was performed as previously described (Reimer and Turck, 2010) using whole two-week-old seedlings harvested at ZT16. A volume of 1 μ l of α -SWN rabbit serum (Eurogentec, DE10103, rabbit SA66993, PPI2485) as no-antibody control and 4 μ l of H3K9me2 antibodies (Diagenode pAb-060-050, lot N°A90-0042) was used for immunoprecipitation. Eluted DNA was amplified by qPCR on the Roche LightCycler480. A calibration curve was performed with one input sample with the dilutions 1/5, 1/25, 1/125, 1/625. Input samples were diluted at 1/20 while immunoprecipitated DNA samples were diluted at 1/5 to be used in a 10 μ l qPCR [0.5 μ l 20 \times EvaGreen dye (Biotium), 5 μ l of diluted DNA, 0.5 μ l forward primer and reverse primer (10mM), 0.1 μ l Taq polymerase (Bio-Budget, Germany), 1 μ l 10 \times Eva Buffer, 2.4 μ l dH₂O]. The amplification cycle was 95°C 10 s, 58°C 15 s, 72°C 20 s, followed by a melting curve analysis with 95°C 5 s, 65°C 1 min, increase from 65°C to 97°C with 2.5°C/s. Average Crossing Points (CP) values of three technical replicates were used to calculate the relative quantity of DNA for each sample and finally the enrichment by normalization to input (%input).

Primers used for qPCR amplification of ChIP samples:

Block C (212b bp)

F: 5'-AAAGGATTGGATGAGTGCAAA-3'

R: 5'-TCTTGACATGGAGCGAAAGA-3'

cl2-ta22 (TE AT4G03790) (208 bp)

F: 5'-AATCTGGGAGGAGAGGAGGA-3'

R: 5'-CATAGACGAACCGCCTTGTT-3'

6.8 Statistical analysis

Statistical analyses were performed using the R Statistical Programming Language (R Core Team, 2014). For flowering time analysis, the number of true rosette leaves per plant, with about 8 to 15 plants per group tested, was used as response variable. For ChIP-seq analysis, the values of %input of two biological replicates were used as response variable. For gene expression analysis, the ratios of two biological replicates were used as response variable. For the three analyses, an analysis of variances (ANOVA) followed by a Dunnett's *post hoc* test were performed (risk alpha=5%) after verifying the normality and homoscedasticity of the data. For gene expression analysis, T3 and T5 generations were tested separately.

7 Results

7.1 IR-induced late flowering

Transgenic lines expressing an IR of the conserved element *Block C* were generated in the accession Col-0 background, called hereafter WT, with the aim to target DNA methylation at *Block C* and assess the effect on *FT* expression (**Figure 3 a**). Two independent lines (#15 and #27) containing an insertion of the transgene at a single locus (segregating 3:1 for herbicide resistance) were selected for further analysis. For each line, one individual T3 sibling having lost the transgene (#15-3 and #27-3) and one having retained the transgene at the homozygous state (#15-2 and #27-4) were propagated till generation T6 for further analysis (**Figure 3 b**). Flowering time was assessed for each generation in strongly *FT*-inductive LD photoperiod and moderately inductive MD conditions. In LD conditions, flowering time was significantly delayed in both transgenic lines compared to WT for the three generations tested (T3 to T5). The non-transgenic lines #15-3 and #27-3 showed a trend towards mild late flowering across generations, although the difference with WT was not always significant (**Supplementary Figure 29**).

To better compare flowering time across generations and reduce the variation due to different growth conditions across the experiments, four generations (T3 to T6) for each line were grown simultaneously under MD conditions in one growth cabinet. As expected, flowering was clearly delayed compared to WT for the transgenic lines with little variation across generations (**Figure 3 c**). Transgenic lines flowered at a similar number of leaves as the *ft-10* mutant that does not produce FT protein due to a Transferred DNA (T-DNA) insertion in the first intron (Yoo et al., 2005) (**Figure 3 c**). The trend towards delayed flowering was again observed in the non-transgenic lines, but the significance could not be reliably confirmed (**Figure 3 c**). *FT* expression was assessed for the four lines at generations T3 and T6. Two-week-old seedlings were harvested at the high *FT* expression ZT16 (16h after the beginning of the light period). The transgenic lines #15-2 and #27-4 showed a significant reduction of *FT* expression for both generations while non-transgenic lines #15-3 and #27-3 showed a significant reduction only at generation T5 (**Figure 3 e**). Lower *FT* expression in the transgenic lines correlated with delayed flowering while the non-transgenic lines showed only a partial reduction of *FT* expression corresponding to their mild late flowering phenotype. In conclusion, presence of the IR targeting *Block C* clearly correlated with delayed flowering in inductive photoperiod and reduced expression of *FT* while a memory effect after removal of the IR was marginal.

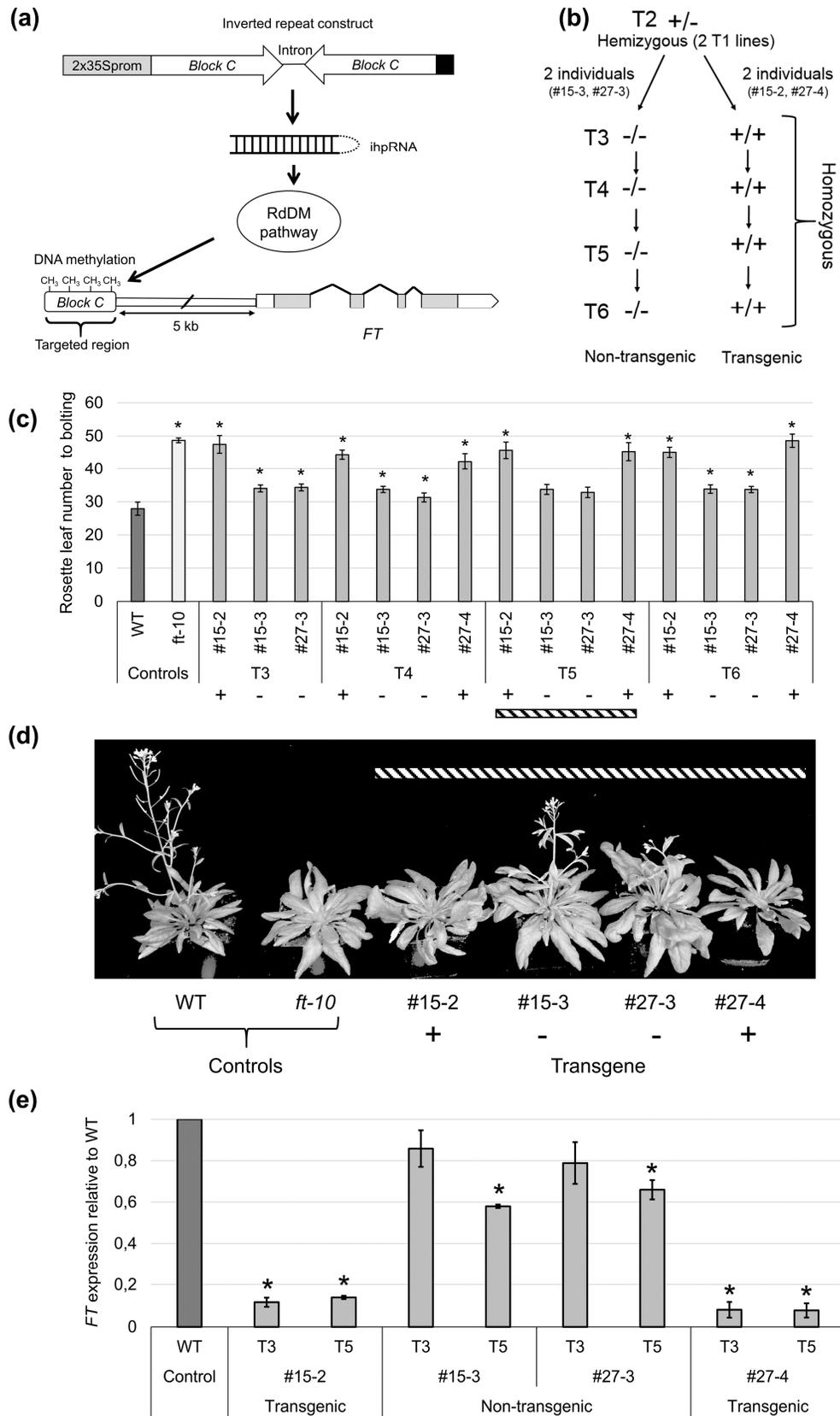


Figure 3: (a) Cartoon of IR-mediated DNA methylation at *Block C*. An IR containing a region of 550 bp which includes *Block C* (380 bp) is driven by the strong cauliflower mosaic virus (CaMV) 35S promoter with duplicated enhancer, separated by the intron 1 of *WRKY33*, and terminated by the 35S terminator of CaMV (black box). During transcription, an intron-containing hairpin RNA (ihpRNA) is formed and can be further processed into the RdDM pathway to trigger *de novo* DNA methylation at the endogenous *Block C*. (b) Scheme of the two independent lines propagated from T2 till T6 generation. Two independent T2 transgenic lines hemizygous for the transgene were propagated and both homozygous transgenic and non-transgenic siblings were selected. These four derived lines were propagated from T3 to T6 generation. (c) Flowering time in MD conditions for the transgenic and non-transgenic lines across four generations. The error bars indicate Standard Errors of the Mean (SEM) (10 plants by modality) and the horizontal black line indicates the average value of flowering time for WT. The T-DNA insertion line *ft-10* is the *FT* knockout allele and was used as late-flowering control. The transgenic line #27-4 at T3 is not displayed in **Figure 2 a** due to bad germination. (d) Phenotype of one representative plant from WT and *ft-10* and the four lines at generation T5 from the same experiment (indicated by the hatched box in **Figure 3 c**). (e) *FT* expression measured at ZT16 in two-week-old seedlings grown in LD conditions. *FT* expression was normalized to the expression of the housekeeping gene *PP2A* and the WT normalized value was set to one. The error bars indicate SEM (two biological replicates). For statistical analyses in figures c and e, ANOVA followed by *post hoc* Dunnett's tests were performed for each generation (one-sided, WT as control group). Asterisks indicate a significant difference with WT (α -risk = 5%).

7.2 DNA methylation at *Block C*

Previous studies have shown that 24-nucleotides (nt) small RNAs (smRNAs) are required for effective TGS through DNA methylation (Hamilton et al., 2002; Melnyk et al., 2011). In the model of the canonical RNA-dependent DNA methylation (RdDM) pathway, 24-nt smRNAs are generated at loci transcribed by RNA polymerase IV (Pol IV) through the action of RNA-dependent RNA polymerase 2 (RDR2) and DICER-LIKE 3 (DCL3) (Matzke and Mosher, 2014). In the case of an IR driven by a strong promoter, double-stranded RNA (dsRNA) is formed directly after transcription by RNA Polymerase II (Pol II) due to the self-complementary structure of the IR. The dsRNA can theoretically be directly processed by DCL3 to integrate the RdDM pathway (Matzke and Mosher, 2014). However, once the ARGONAUTE 4 (AGO4)-siRNA complex is formed, *de novo* DNA methylation relies on the production of scaffold transcripts by RNA Polymerase V (Pol V) or Pol II at the endogenous target site (Matzke and Mosher, 2014). We sequenced the smRNAs in transgenic and non-transgenic lines at generation T6 to establish whether smRNAs were generated by the IR and assessed whether they were restricted to the IR target region. We found smRNAs mapping specifically to the target region in the transgenic lines but not in the non-transgenic siblings. This indicates that the presence of the transgene was required for smRNAs generation and that production of smRNAs was limited to the IR target site (**Figure 4 a**). The predominant species of smRNAs mapping to the IR target region was 21-nt in both transgenic lines (34-38%), followed by 22-nt and 24-nt species representing about 20% and 10% of smRNAs, respectively (**Figure 4 b**). The IR contained as spacer the first intron of the *A. thaliana WRKY33*

gene. Although the intronic sequence is not self-complementary and should therefore not form dsRNAs, it generated a similar spectrum of smRNAs mapping to this genomic location, but with very few reads (15 for line #15-2 and 62 for line #27-4) compared to *Block C* (**Supplementary Figure 30**). This indicates that single-stranded RNAs (ssRNAs) derived from the construct could, at a small level, generate smRNAs. Transgenic line #27-4 generated about seven times more smRNAs than line #15-2 though reduction of *FT* expression in these two lines was comparable indicating that the effect was saturated (**Figure 4 a, b**).

To verify whether DNA methylation was specifically induced at *Block C*, bisulfite conversion experiments were performed for the four lines to analyze the DNA methylation status for the IR target region and its flanking sequences. We confirmed previous genome-wide studies of DNA-methylation in seedlings showing that *Block C* is not methylated in Col-0 (Lister et al., 2008) (**Figure 4 c**). *Block C* is defined as a 380 bp long region containing 48 cytosines of which eight, two, and 38 cytosines are in CG, CHG, and CHH context (H being A, T or G), respectively. We found that DNA methylation was induced at the IR target region including *Block C* with levels of methylation varying from 10 to 100% for each cytosine position in the transgenic lines #15-2 (**Figure 4 c**) and #27-4 (**Figure 4 d** and **Supplementary Figure 31**). DNA methylation was also detected at sites flanking around 100 bp upstream and downstream of the IR target region although no smRNAs were found to map in these regions (**Figure 4 c**). To assess whether DNA methylation at *Block C* could spread beyond the flanking regions, a 300 bp region located 2.7 kb downstream of *Block C* and 2 kb upstream of *FT* TSS was selected. No DNA methylation was observed in the control region, indicating that DNA methylation did not spread from *Block C* towards the TSS of *FT* (**Supplementary Figure 32**). We compared DNA methylation level at generation T3 and T5 for two independent lines with transgenic (#15-2 and #27-4) and non-transgenic (#15-3 and #27-3) siblings. We found that CG methylation was maintained in both transgenic lines and to a much lower extent in non-transgenic lines with about 10% for #15-3 and a few percent for #27-3 (**Figure 4 d**). Interestingly, the remaining CG methylation seemed to be maintained from T3 to T5 generation (**Figure 4 d**). CHH methylation decreased in the transgenic lines with progressing generations while it was near zero in the non-transgenic lines #15-3 and #27-3 (**Figure 4 d**). Altogether, these results indicate that, upon IR loss, only CG methylation is partially maintained and the level of maintenance varies from a line to the other.

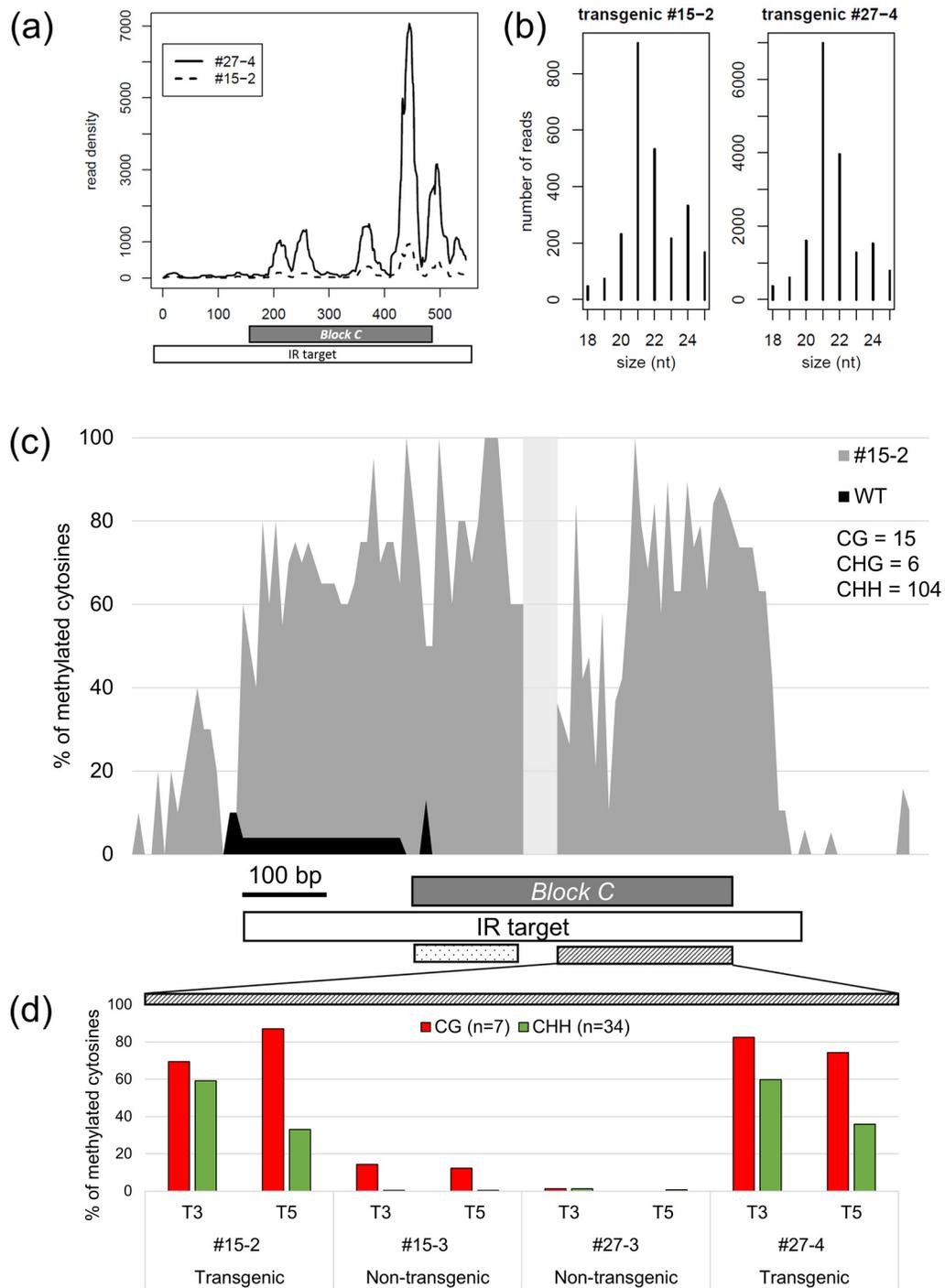


Figure 4: (a) Density of reads mapping to the IR target region for both transgenic lines #15-2 and #27-4 (generation T6). Non-transgenic lines are not shown since no smRNAs could be found. The boxes below the chart indicate the location of *Block C* and the IR target region. (b) Distribution of read sizes for both transgenic lines #15-2 and #27-4 (generation T6). The smRNAs species of 21-nt is dominant in both lines while 22-nt and 24-nt smRNAs are the second most represented species with variation between the two lines. (c) Average level of DNA methylation at each cytosine position for WT and the transgenic line #15-2 (generation T3) at the IR target region. DNA methylation is induced mainly at the IR target

region (white box) including *Block C* (grey box) in the transgenic line #15-2, while almost no DNA methylation is found in WT. The vertical light gray rectangle represents a region containing four cytosines with unknown methylation status. At least ten clones were examined. The dotted box indicates a region examined for DNA methylation in WT and transgenic lines #15-2 and #27-4 in **Supplementary Figure 31**. **(d)** Average level of DNA methylation at the *Block C* region indicated with a diagonal hatched box for two transgenic lines (#15-2 and #27-4) and their non-transgenic siblings (#15-3 and #27-3) for two different generations (T3 and T5) in CG and CHH contexts. At least eight clones were examined. The number of cytosines for each context is indicated.

7.3 H3K9me2 association with DNA methylation deposition

Methylated DNA can recruit the histone methyltransferase KRYPTONITE that dimethylates lysine 9 of histone H3 (H3K9me2), which is thought to promote chromatin compaction (Du et al., 2014). H3K9me2 in turn recruits CHROMOMETHYLASE 2 (CMT2) and CHROMOMETHYLASE 3 (CMT3) that methylate cytosines in CHH and CHG contexts, respectively (Du et al., 2015). We assessed whether induction of DNA methylation by the IR resulted in deposition of H3K9me2 at *Block C* by carrying out chromatin immunoprecipitation (ChIP) using material from the transgenic and non-transgenic lines at generation T5. The 213 bp-long amplified region overlaps with the IR target region and the flanking endogenous sequence allowing to assess H3K9me2 level only at *Block C* and not at the transgene (**Figure 5 a**). H3K9me2 levels were significantly increased in the transgenic lines compared to WT while no significant increase was observed for the non-transgenic sibling #15-3 (line #27-3 was not tested) (**Figure 5 b**).

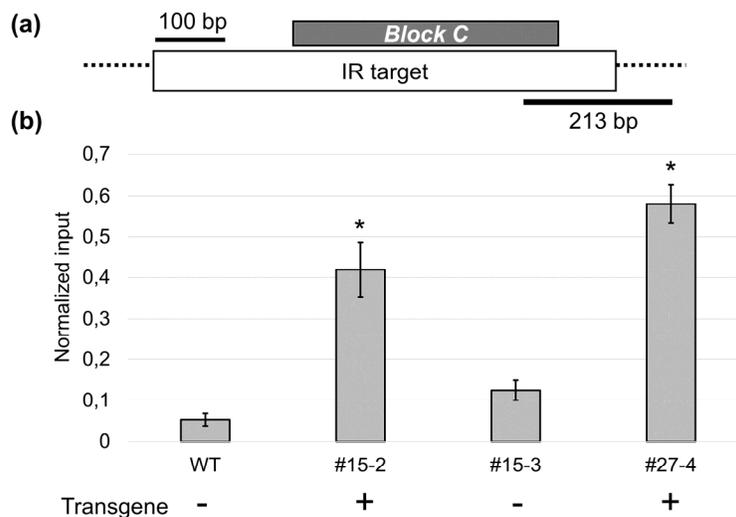


Figure 5: **(a)** An overlapping region of 213 bp between the IR target region (white box) and the flanking endogenous region (dashed line) was used as probe (thick black line) to assess the level of H3K9me2 at *Block C*. **(b)** Level of H3K9me2 present at *Block C* in WT and three lines: two transgenic lines #15-2 and #27-4, and the non-transgenic line #15-3. The % of input at *Block C* was normalized to the % of input at a TE region (gypsy-like retrotransposon AT4G03790) used as H3K9me2-rich control region. Averages of two biological replicates with SEM are shown. ANOVA followed by post-hoc Dunnett's tests were performed (one-sided, WT as control group). Asterisks indicate a significant difference with WT (α -risk = 5%).

7.4 Newly identified regulatory element *Block E*

Since DNA methylation targeted at the *cis*-regulatory region *Block C* caused downregulation of *FT* expression, we set out to interrogate whether other regions potentially involved in *FT* regulation could be surveyed with this method. We extended the previous phylogenetic analysis that had identified *Block C* (Adrian et al., 2010) by adding more orthologous sequences from the Brassicaceae family that have become available on the Phytozome database (Goodstein et al., 2012) and by including downstream sequences (**Figure 6 a**). Previously identified regions *Block C*, *B*, and *A* were all conserved across several species of Brassicaceae (**Figure 6 a**). A region downstream of the *FT* gene, which we now name *Block E*, was also highly conserved among all available sequences (**Figure 6 a**). Besides phylogenetic conservation, the overlap with high chromatin accessibility and low H3K27me3 level was a distinctive feature that *Block E* shared with *Block C* (**Figure 6 a**). Further analysis revealed that *Block C* and *E* share several super-conserved shadows that overlap with potential TFBSs, such as an I-box, a RE-box, and a CCAAT-consensus site (**Supplementary Figure 33**). This made *Block E* a good *cis*-regulatory region candidate. We also included IRs targeting the previously identified *Block B* and a region we previously described as a large insertion (Col-0 insertion) present in approximately 25% of *A. thaliana* accessions sequenced (Liu et al., 2014), including the reference Col-0 (**Figure 6 a**). DNA methylation was not found at *Block B* and *E* in Col-0 in published data, while the Col-0 insertion was shown to carry methylation (Lister et al., 2008). Our data did not reveal DNA methylation in the control region just adjacent to the 3' end of *Block B* (**Supplementary Figure 32**) and confirmed the presence of DNA methylation at Col-0 insertion (**Supplementary Figure 34**). Flowering time was performed in the T3 generation for independent lines corresponding to single locus insertions based on the segregation of the herbicide resistance marker. Two independent lines expressing an IR targeting *Block E* showed significant delayed flowering in LD conditions compared to WT, although the effect was less strong as it had been detected for *Block C* (**Figure 6 b, c**). A slight but significant late flowering was detected in lines expressing IRs directed against *Block B* and the Col-specific insertion (**Figure 6 b, c**).

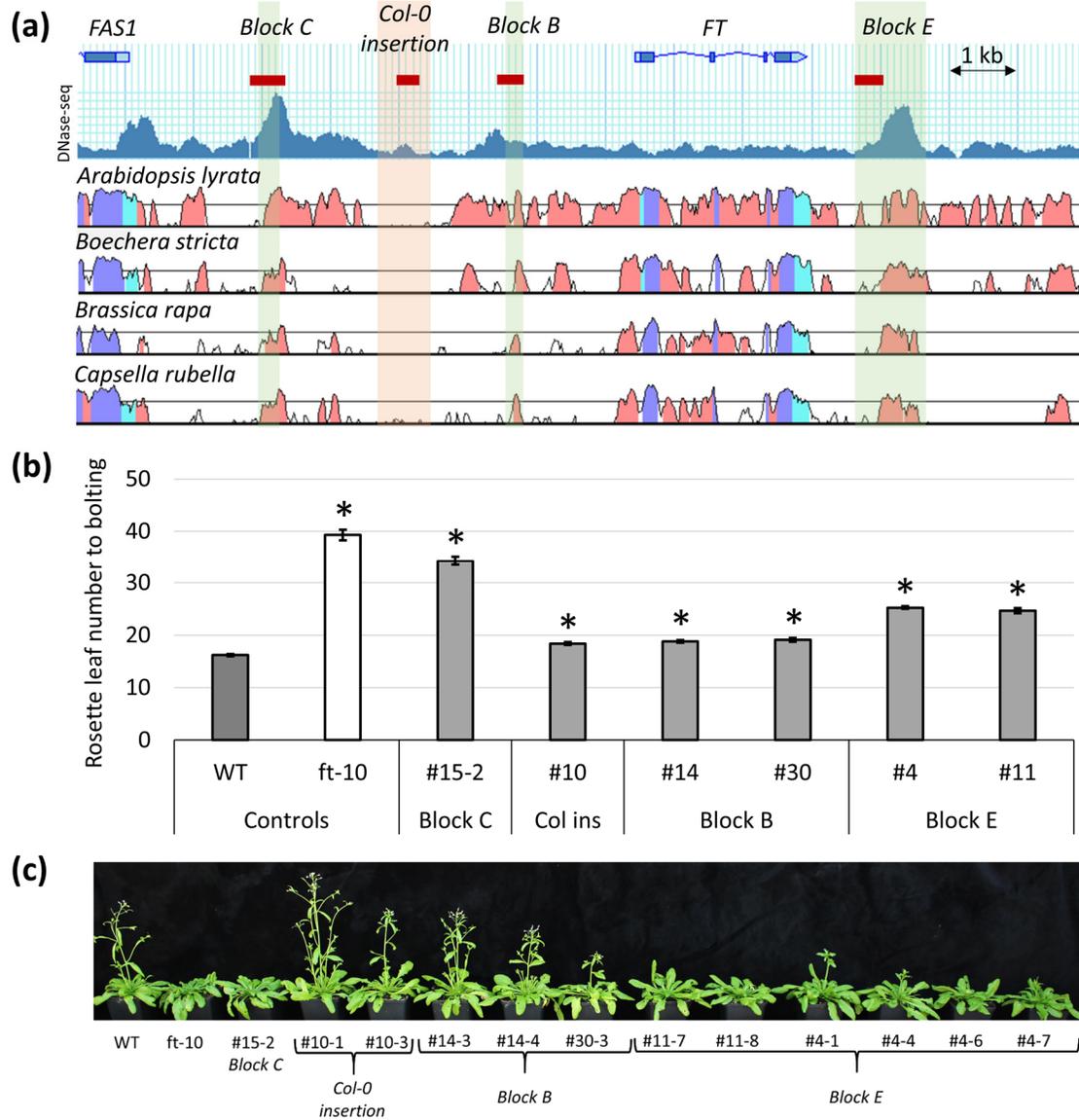


Figure 6: (a) Chromatin accessibility, H3K27me3 enrichment, and conservation level at the *FT* locus. Chromatin accessibility from published DNase-seq data (labelled DNase I) (Zhang et al., 2012b). Phylogenetic shadowing showing the conservation level between *A. thaliana* and four other species from the Brassicaceae is displayed in red, cyan, and purple for the intergenic regions and introns, Untranslated Regions (UTRs), and exons, respectively. Phylogenetic shadowing generated with Vista Point (Frazer et al., 2004). Green shadows indicate the previously defined *Block C* and *Block B*, and the newly defined *Block E*. The orange shadow indicates the location of the *Col-0* insertion. Red filled rectangles indicate regions targeted by IRs. (b) Flowering time in LD conditions for transgenic plants containing IR targeting *Block C*, *Col-0* insertion, *Block B*, and the adjacent 5' region of *Block E*. *Col-0* and *ft-10* were used as control for WT and null-allele flowering time phenotype, respectively. The error bars represent SEM (at least 10 individuals per group). ANOVA followed by post-hoc Dunnett's tests were performed (one-sided, WT as control group). Asterisks indicate a significant difference with WT (α -risk = 5%). (c) Picture of the representative plants of the lines scored for flowering in b with the first and second number indicating the line and the sibling from T2 generation, respectively.

8 Discussion

We showed that IR-mediated DNA methylation of *Block C* could lead to the downregulation of *FT* expression and was associated with late flowering and H3K9me2 deposition. We used the same rationale to characterize other putative *cis*-regulatory elements of *FT*. We found that IR targeting the previously defined *Block B* (Adrian et al., 2010) had a mild but significant effect on flowering. Furthermore, we identified a new putative *cis*-regulatory element located downstream of *FT* and named *Block E*. This region, similarly to *Block C*, was conserved across Brassicaceae, contained similar putative TFBSs, and was located in accessible chromatin. Transgenic plants containing an IR targeting a part of *Block E* showed a clear late flowering phenotype.

As negative control, we targeted a part of the Col-0 insertion, located between *Block C* and *FT*. We did not expect an effect on *FT* expression since the region is already heavily methylated in WT, mostly in symmetric contexts. However, the transgenic line obtained for this construct was mildly late flowering, with a comparable phenotype than the transgenic lines containing an IR targeting the previously defined *Block B*, located about 900 bp downstream of the Col-0 insertion. This suggests that additional symmetric and newly set asymmetric DNA methylation at the Col-0 insertion could affect the general chromatin organization and alter the function of *Block C*. Considering this result, we suggest that for the two independent IR transgenic lines for *Block B*, the mild late flowering is rather linked to a general chromatin change at the locus than to silencing of a putative *cis*-regulatory element within the region. However, additional independent IR transgenic lines for Col-0 insertion should be obtained to confirm the mild late flowering observed in the only line examined. In comparison, the stronger late flowering phenotype of several independent *Block C* and *Block E* IR lines suggests a clearer implication of these two regions in *FT* regulation. The effect of the IR on *Block C* was stronger than on *Block E* but the fact that we targeted only a fraction of the *Block E* region may result in an underestimated repressive effect. Interestingly, ChIP-seq data revealed the binding of the *FT* activator PHYTOCHROME-INTERACTING FACTOR 4 (PIF4) within *Block E* (Pedmale et al., 2016). Since our IR is adjacent to the 5' end of the predicted PIF4 binding site, new IR constructs targeting this region should be generated to fully assess the effect of *Block E* methylation on *FT* expression.

In *A. thaliana*, IRs could induce TGS when targeting either a viral enhancer driving the reporter gene *GREEN FLUORESCENT PROTEIN (GFP)* (Kanno et al., 2008; Daxinger et al., 2009) or the endogenous promoters of *TOO MANY MOUTHS (TMM)* and *FT* (Deng et al., 2014; Deng and Chua, 2015). However, we are the first to describe downregulation of gene expression by IR-mediated DNA methylation of an endogenous enhancer in *A. thaliana*. DNA methylation deposition at *Block C* in transgenic lines was associated with H3K9me2 deposition. This observation is consistent with the fact that DNA methylation and H3K9me2 are acting in a self-reinforcing loop

(Du *et al.*, 2015). H3K9me2 was associated with enhancer silencing in mammals (Zylicz *et al.*, 2015) and is required for TE silencing in plants (Bernatavichute *et al.*, 2008). Therefore, H3K9me2 at methylated *Block C* may participate to the switch of the chromatin from an active open state to an inactive heterochromatin-like state. This hypothesis is supported by a study in rice showing effective IR-mediated TGS only for the target promoters displaying both H3K9me2 and DNA methylation deposition (Okano *et al.*, 2008). In *A. thaliana*, the deposition of H3K9me2 after IR-mediated DNA methylation at the target region in *FT* intron 1 was also reported (Deng and Chua, 2015).

We also assessed the presence of IR-induced smRNAs. As predicted, the IR generated smRNAs mapping at *Block C*. The amount of smRNAs generated in the two analyzed transgenic lines differed with a fold change of about seven. Nonetheless, the downregulation of *FT* and the corresponding late flowering phenotype was similar in both lines, indicating a saturated effect. Although smRNAs were confined to sequences contained in the IR, DNA methylation spread out bidirectionally from the target region up to about 100 bp. Previous studies showed that unidirectional spreading of DNA methylation out of the IR target was linked to the generation of secondary siRNAs (Kanno *et al.*, 2008; Daxinger *et al.*, 2009). However, the viral enhancer used as target in these studies showed an intrinsic promoter activity that may explain the difference in DNA methylation spreading observed with *Block C* (Kanno *et al.*, 2008). We conclude that secondary siRNAs are not needed for spreading of IR-mediated DNA methylation in the case of *Block C*. It would be interesting to know whether Pol II and/or Pol V are transcribing nascent scaffold RNAs at *Block C* to recruit AGO4-siRNA and direct *de novo* methylation (Zheng *et al.*, 2009). We also observed that 24-nt smRNAs were not the most abundant species with only 13% of the total smRNAs while 21-22-nt smRNAs represented 60-70% of all smRNAs mapping at *Block C*. Consistently with our observations, IR targeting transgenic *GFP* and *CHALCONE SYNTHASE (CHS)* fragments were reported to induce mainly the production of 21-nt, followed by 22-nt and 24-nt smRNAs, indicating that DLC3 was less effective at dicing dsRNA than DLC4 and DCL2, which generate 21 and 22-nt smRNAs, respectively (Llave *et al.*, 2002; Wroblewski *et al.*, 2014). Interestingly, two other studies targeting the *TMM* promoter and the first intron of *FT* with IR found 24-nt smRNAs to be the main species generated (Deng *et al.*, 2014; Deng and Chua, 2015), indicating that the insertion location or the sequence of the IR may influence the processing of dsRNAs by the different DCL proteins. Although 24-nt smRNAs are responsible for *de novo* methylation in the canonical RdDM, 21-nt smRNAs generated by DCL1 were able to integrate into AGO4/6 to target DNA methylation via Pol V-mediated RdDM (Wu *et al.*, 2012). Therefore, 21-nt smRNAs may also participate to DNA methylation deposition at *Block C*.

The non-transgenic siblings that segregated out the IR targeting *Block C* contained a low but noticeable residual DNA methylation at symmetric contexts. These plants also displayed a mild late flowering phenotype which was not always significant across experiments but in most cases apparent. The lack of smRNAs targeting *Block C* in these plants indicates that residual DNA methylation was maintained across generations by CMT3 and METHYLTRANSFERASE 1 (MET1) for CHG and CG contexts, respectively. The effect of this residual DNA methylation may explain the mild late flowering phenotype observed but it may also be related to other changes in chromatin induced in the transgenic parents and transmitted to the non-transgenic offspring. Other features such as chromatin accessibility could bring more insight about the chromatin changes occurring in plants that contain IR and plants that lost it.

Here, we showed that IR-mediated DNA methylation is an interesting tool to characterize enhancers and test other putative *cis*-regulatory regions displaying specific features such as conserved sequences or increased chromatin accessibility. We also found that, compared to most transgenic approaches, IR-mediated DNA methylation was especially efficient in generating transgenic lines with homogenous phenotypes across lines and generations. In conclusion, the *FT* locus was used as model to bring additional evidence of the connection between DNA methylation, H3K9me2, and enhancer activity in plants. In the future, new *cis*-regulatory elements such as *Block E* will need further investigation to unravel the complex and fascinating gene regulation taking place at the *FT* locus. This knowledge can be applied to other species and improve our understanding of the floral transition mediated by *FT* homologs.

Part II - Identification of putative enhancers in *Zea mays*

9 Introduction

9.1 *Zea mays* as genetic model

Zea mays (*Zea mays* subsp. *mays* L.) is a cereal originated from Mesoamerica representing today the first cereal crop worldwide in terms of production (source: <http://faostat.fao.org/>). In addition to its agronomical importance, *Z. mays* is a valuable model for genetic studies due, in part, to its easily controlled pollination, easy phenotyping and incredible diversity (Coe, 2001). *Z. mays* also possesses a wide collection of genetic mutants, which facilitates gene characterization, and more recently, a first genome assembly for the inbred line B73 (Schnable et al., 2009). Compared to *Arabidopsis thaliana* (≈ 135 Mb, $n=5$), *Z. mays* is also diploid but its genome is around 20 times bigger (≈ 3.2 Gb, $n=10$), mainly due to its high content in repetitive Transposable Elements (TEs) representing $\approx 80\%$ of the genome (Schnable et al., 2009; Baucom et al., 2009). Though the number of predicted coding genes¹ in *Z. mays* (39,469) is only 1.5-fold higher than in *A. thaliana* (27,416), the genome of *A. thaliana* is far more compact with shorter intergenic regions. Therefore, we consider *Z. mays* as a good model to find distal intergenic enhancers clearly separated from promoter regions. In addition to its large intergenic regions compared to *A. thaliana* genome, several enhancers, confirmed or putative, were described in *Z. mays*: enhancers of *pericarp color1* (*p1*), *booster1* (*b1*), *teosinte branched1* (*tb1*), and *ZmRap2.7* genes (Sidorenko et al., 1999; Stam et al., 2002; Clark et al., 2006; Salvi et al., 2007). In the following, we will describe the characteristics of the confirmed enhancers of *b1*, *tb1*, *p1*, and the putative enhancer of *ZmRap2.7*.

9.2 Enhancer of *b1*

The gene *b1* encodes a Transcription Factor (TF) from the Myc family that activates the genes in the anthocyanin biosynthetic pathway (Goff et al., 1990). Different *b1* alleles generate various patterns and intensities of pigmentation (Styles et al., 1973; Radicella et al., 1992). For instance, the *B-I* allele is characterized by the presence of seven tandem repeats, named hepta-repeat, located ≈ 100 kb upstream of *b1* (Stam et al., 2002). The hepta-repeat displays a higher level for histone 3 lysine 9 acetylation (H3K9ac) and H3 lysine 14 acetylation (H3K14ac) enrichment and an increase in chromatin accessibility in tissues where *b1* is highly expressed (Louwers et al., 2009; Haring et al., 2010). Furthermore, multiple long-range interactions involving the hepta-repeat were shown

¹ The predicted number of genes for *Z. mays* as for *A. thaliana* are from the Gramene database and are prompt to frequent updates (<http://ensembl.gramene.org/> consulted on February 2016).

using Chromosomal Conformation Capture (3C) (**Figure 7**) (Louwers et al., 2009). In addition to its enhancer activity, the hepta-repeat was shown to be involved in paramutation (Stam et al., 2002). Paramutation is the communication in *trans* between two alleles to establish meiotically heritable expression states (Chandler and Alleman, 2008). In the case of *b1*, the *B'* allele contains the exact same hepta-repeat than *B-I* allele but does not enhance *b1* expression. When *B'* is crossed with *B-I*, *B'* silences *B-I* allele, becoming in its turn paramutagenic, that is to say able to silence another naïve *B-I* allele, and this in a frequency of 100% (Coe, 1959). The *B-I* hepta-repeat is only partially conserved in the sequenced inbred line B73 with only one of the seven repeats conserved at 92% and present ≈ 60 kb upstream of *b1*.

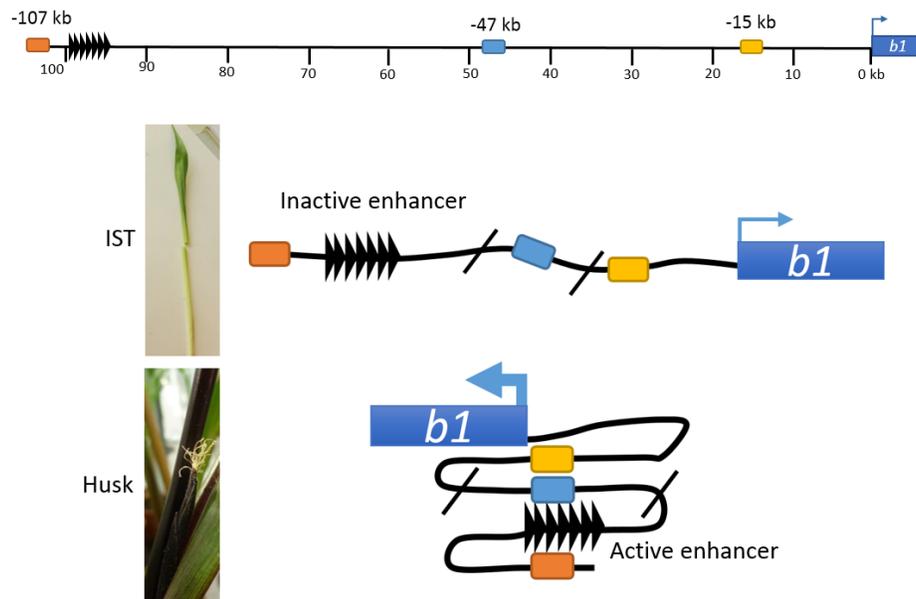


Figure 7: Simplified model of the multi-loop-mediated regulation of *b1* for the *B-I* allele. *b1* is lowly expressed in the Inner Stem Tissue of seedling (IST) with an inactive hepta-repeat (represented by seven black arrows) located ≈ 100 kb upstream of *b1* Transcriptional Start Site (TSS). In certain mature tissue such as husk, the promoter region of *b1* is interacting with the hepta-repeat and three other regions indicated by colored boxes. This multi-loop complex is mediated by unknown tissue-specific TFs (not shown). Figure adapted from Louwers et al., (2009).

9.3 Enhancer of *tb1*

The gene *tb1* encodes a TF of the TCP family and acts as a repressor of axillary bud outgrowth (Hubbard et al., 2002). In *Z. mays* wild progenitor teosinte, *tb1* is lowly expressed, resulting in high branching architecture. In *Z. mays*, *tb1* is expressed at twice the level of the teosinte allele in the immature ears and developing branch, resulting in a strong apical dominance (Doebley et al., 1997). Therefore, *tb1* was identified as a major domestication Quantitative Trait Locus (QTL) (Doebley and Stec, 1991, 1993). Later, fine mapping analyses allowed to refine the position of the QTL to a region located between 58 and 69 kb upstream of *tb1* TSS (Clark et al., 2006). This region was shown to act as an enhancer of *tb1*. Indeed, transient expression assays indicated that a *Hopscotch*

TE insertion within this region could increase the transcription of a reporter gene and is therefore likely to be responsible for the enhancer activity of this region (Studer et al., 2011).

9.4 Enhancer of *p1*

The gene *p1* encodes a *Myb*-like TF that regulates the expression of at least three genes involved in the accumulation of red flavonoid pigments in specific floral organ tissues (Styles and Ceska, 1981; Grotewold et al., 1991). An enhancer located 5 kb upstream of *p1* TSS could increase the expression of the β -glucuronidase (*GUS*) reporter gene in transient expression experiments when fused to the basal promoter of *p1* (Sidorenko et al., 1999). In addition, an *Activator* (*Ac*) insertion in the enhancer could disrupt the expression of the gene (Moreno et al., 1992). The enhancer of *p1* is contained within a 5.2 kb direct repeat also present downstream of *p1* in the characterized active *P-rr* allele (Athma and Peterson, 1991). These two repeats do not seem to derive from TEs but originate from a gene duplication event which took place about 2.75 million years ago (Zhang et al., 2000). However, *Ac* insertion in the downstream direct repeat did not affect *p1* expression (Athma et al., 1992; Moreno et al., 1992). Although the downstream direct repeat does not regulate *p1* expression, both upstream and downstream direct repeats show a decrease in DNA methylation and an increase chromatin accessibility in tissues where *p1* is highly expressed (Lund et al., 1995). Finally, a transgenic construct containing *p1* enhancer could silence the endogenous *P-rr* allele and this silencing could be transgenerationally maintained after loss of the transgene, indicating a paramutation-like phenomenon (Sidorenko and Peterson, 2001).

9.5 Putative enhancer of *ZmRap2.7*

Vegetative to generative transition1 (*Vgt1*) was identified as one major QTL for flowering time in *Z. mays* (Vlăduțu et al., 1999). Flowering time in *Z. mays* is mainly defined by the timing of the transition from vegetative to reproductive stage made by the shoot apical meristem (Irish and Nelson, 1991). Most *Z. mays* lines are quantitative short day plants; however, in the normal field conditions present in Europe or North America (long days), the meristem initiates a determined number of vegetative nodes and then converts to tassel development (Russell and Stuber, 1983). *Z. mays* in most cultivated areas is therefore relying rather on internal clues than on photoperiod to accomplish its floral transition. *Vgt1* was defined as a region of ≈ 2 kb located 70 kb upstream of *ZmRap2.7*, an *APETALA2*-like TF acting as floral repressor (Salvi et al., 2007). *Vgt1* contains Conserved Non-coding Sequences (CNSs), when compared to *Z. mays* relative species rice and sorghum, and is suggested to regulate the expression of *ZmRap2.7* (Salvi et al., 2007). The presence of a Miniature inverted-repeat TE (Mite) located within *Vgt1* was shown to correlate with the downregulation of the floral repressor *ZmRap2.7*, giving an early flowering phenotype (Salvi et al.,

2007). In addition of the Mite, a 2-bp insertion/deletion within *Vgt1* was associated with early flowering (Ducrocq et al., 2008). Although changes in *Vgt1* were correlated with changes in the expression of *ZmRap2.7*, no reporter assay experiments have been performed to conclude on the putative enhancer activity of *Vgt1*.

These four examples are the only known enhancers reported in *Z. mays*, what is relatively few compared to the abundance of enhancers described and characterized in mice or *Drosophila* (Levine, 2010; Rubinstein and Souza, 2013). Furthermore, the function of two of them is directly linked to the presence of TEs, supporting the old idea that TEs are drivers of evolution (Britten and Davidson, 1969).

In order to identify novel enhancers in *Z. mays*, including stage-specific and tissue-specific enhancers, three high-throughput sequencing-based methods were combined: DNase-seq, Chromatin Immunoprecipitation (ChIP)-seq, and RNA-seq. Most active enhancers are located in DNase I hypersensitive sites (DHSs) which correspond to accessible chromatin in a context of low-nucleosome abundance (Thurman et al., 2012). Also, histone marks H3K9ac and H3K14ac were associated with the active enhancer of *b1* in *Z. mays* (Haring et al., 2010), and more generally with active enhancers in mammals (Karmodiya et al., 2012). We performed DNase-seq to map chromatin accessibility profile and ChIP-seq to map H3K9ac enrichment profile in *Z. mays* using two tissues: immature leaves from seedlings and husk. We found that previously characterized enhancers as well as many TSSs correlate with DHSs and H3K9ac mark. In addition, we performed global gene expression analysis using RNA-seq for both tissues to identify differentially expressed genes and link them to flanking putative enhancers showing dynamic changes in chromatin accessibility and H3K9ac. We generated a list of candidates corresponding to tissue-specific and constitutive enhancers. The ChIP-seq for H3K9ac was performed by Blaise Weber and the bioinformatic analyses were performed in collaboration with Rurika Oka (Dr. Maïke Stam's group, University of Amsterdam).

10 Materials and methods

10.1 Culture conditions

The seeds used for all experiments were derived from the siblings of the sequenced B73 inbred line (obtained from Jack Gardiner). Plants were sown on soil in 9x9 cm square pots and placed in greenhouse with a constant temperature of minimum 22°C, a humidity of ≈60%, and a photoperiod of 15 to 16 h (luminosity of ≈ 15,000 lux). Two-week-old seedlings were transferred into 8 l pots if adult plants were needed.

10.2 Material harvest

Two types of tissues were used for the different experiments: (1) Inner Stem Tissue (IST) consisting of immature leaves located in the central part of young seedlings at stage V2 (two leaf collars visible) (Abendroth et al., 2011), (2) inner husk layers also called hypsophylls (hereafter referred as husk) removed from developing ears (≈2-5 cm emerging silks) (**Figure 8**). The harvested tissues were either directly flash-frozen into liquid nitrogen for subsequent DNase-seq and RNA-seq experiments, or crosslinked then flash-frozen for ChIP experiments. After being flash-frozen, samples were grinded into fine powder using a mortar and a pestle permanently cooled down with liquid nitrogen (**Figure 8**). The fine powder was then stored at -80°C into 50 ml conical tubes until nuclei extraction step.

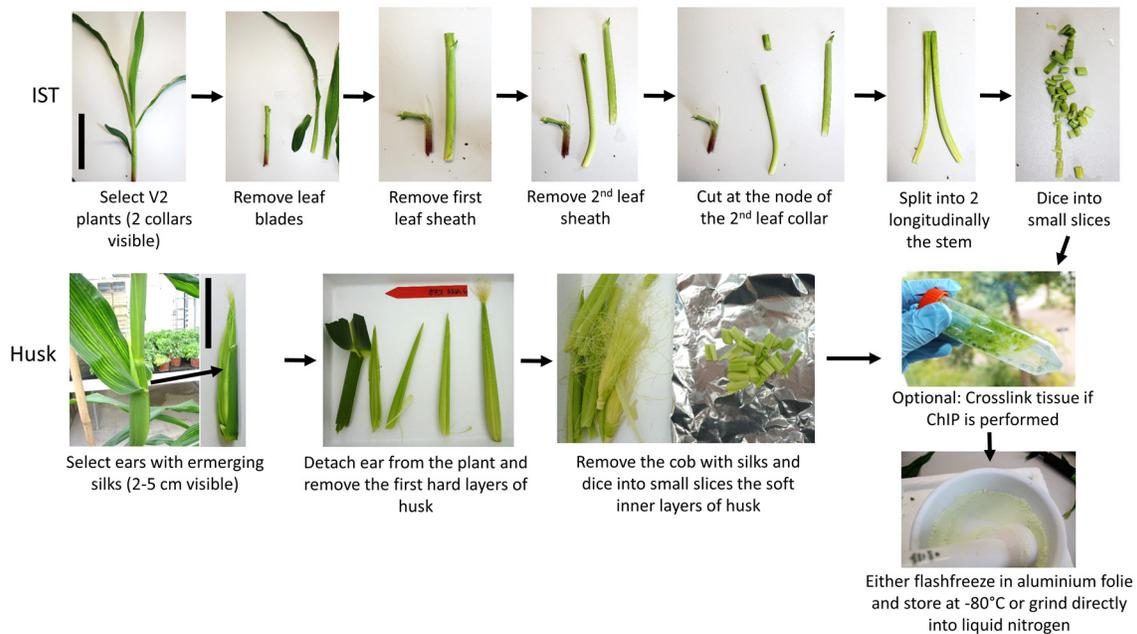


Figure 8: Harvest of the IST and husk tissue for nuclei extraction. Scale bars are indicated with vertical black lines (≈10 cm).

10.3 DNase-seq

10.3.1 Nuclei preparation

Nuclei were extracted from the IST of a dozens of V2 seedlings and from the husk of three plants following the derived protocol of (Steinmüller and Apel, 1986). Two pools of IST and two pools of three husk tissues were used as biological replicates (**Figure 9 a**). Five grams of tissue were grinded into liquid nitrogen, homogenized with 25 ml of cold nuclei isolation buffer [20mM Tris-HCl pH 8, 250mM sucrose, 5mM MgCl₂, 5mM KCl, 40% glycerol (v/v), 0.25% Triton X-100 (v/v), 0.5mM EGTA pH 8, 5mM ethylenediaminetetraacetic acid (EDTA) pH 8, 0.1mM phenylmethylsulfonyl fluoride (PMSF), 0.1% β-mercaptoethanol (v/v), 1:1000 dilution of Proteinase Inhibitor Cocktail (SIGMA)] into an ice-cold 50 ml tube. The lysate was homogenized gently by rotation at 4°C until thawed (ca. 30 min). The lysate was then filtered into a successive layers of 60 μm and 20 μm nylon meshes into an ice-cold 35 ml round-bottom tube (Beckman Coulter) and centrifuged at 6000 x g in JA-25.50 Rotor (Beckman Coulter) 15 min at 4°C. Supernatant was discarded by pipetting and the pellet was resuspended in 15 ml of ice-cold nuclei isolation buffer using a 1-ml cut-off tip and centrifuged at 6000 x g 12 min at 4°C. The pellet was resuspended in 10 ml of ice-cold nuclei isolation buffer and centrifuged at 6000 x g 12 min at 4°C. The pellet was finally resuspended in 1 ml of ice-cold nuclei storage buffer [20% glycerol (v/v), 20mM Tris pH 7.5, 5mM MgCl₂, 1mM dithiothreitol (DTT)]. An aliquot of 20 μl was taken and stained with 4',6-diamidino-2-phenylindole (DAPI; 1 μg of 1 mg/ml solution was added). Nuclei were observed with a confocal microscope under ultraviolet (UV) light for quality check. The nuclei suspensions were flash-frozen in liquid nitrogen and stored at -80°C.

10.3.2 DNase I digestion

The nuclei suspension was thawed on ice while preparing the buffers for DNase I digestion. One undigested control and four concentrations of DNase I recombinant (Roche) were used (50, 100, 150, and 200 U/ml). A volume of 2.5 ml of DNase I buffer [50mM Tris pH 8, 250mM sucrose, 100mM KCl, 0.1mM CaCl₂, 5mM MgCl₂, 50 μg/ml bovine serum albumin (BSA), 0.1% β-mercaptoethanol (v/v)] was prepared per sample and kept on ice. One ml of nuclei suspension was split in 5 x 200 μl into 1.5 ml tubes using 200 μl cut-off tips. The tubes were centrifuged at 1500 x g 5 min at 4°C and the supernatant was discarded. A volume of 100 μl of 100mM EDTA was added to the undigested control followed by 600 μl of phenol/chloroform/isoamyl alcohol 25:24:1 (hereafter P/C/IAA) (Carl Roth) and set aside at Room Temperature (RT). The other pellets were resuspended in 475 μl of cold DNase I buffer by rubbing the tubes against a plastic tube rack vigorously. The tubes were then placed on a heating rack set at 25°C. The DNase I dilutions were prepared by mixing recombinant DNase I (10 U/ μl, Roche) with DNase I dilution buffer [20mM

Tris pH 7.5, 50mM NaCl, 1mM DTT, 100 µg/ml BSA, 50% glycerol (v/v)] to get a final volume of 25 µl at 0, 50, 100, 150, and 200 U/ml of DNase I. The DNase I solution was added to the nuclei solution, the tube was inverted quickly five times for homogenization and let for precisely 10 min at 25°C. After 10 min, 100 µl of 100mM EDTA were added to stop the reaction, followed by 600 µl of P/C/IAA. All samples including the undigested control were shaken using a Tissue Lyser (Qiagen) set at 8 Hz for 5 min at RT. A second P/C/IAA extraction was performed and about 600 µl of the supernatant was recovered by careful pipetting. To degrade remaining RNAs, a volume of 1 µl of RNase A (1 mg/ml) was added to the supernatant and incubated at 37°C for 10 min. Nucleic acid precipitation was performed by adding 600 µl of isopropanol, 50 µl of 7.5M ammonium acetate (final 0.3M), and 2 µl of 10 mg/ml glycogen solution (20 µg/ml final). The tubes were inverted several times and centrifuged at 16,000 x g 30 min at 4°C. Two washing steps with 70% ethanol were performed and the dried pellet was finally resuspended into 30 µl of 10mM Tris-HCl pH 8.5 buffer. The concentration was measured photospectrometrically and the entire sample was mixed with 5 µl of creosol red dye loading dye [1.75M sucrose (60%), 5mM creosol red, pH 8] and loaded on a 1.5% agarose gel containing ethidium bromide. Gel visualization under UV light indicated which digestion to select. In our hands, the samples digested with 50 U/ml of DNase I were chosen as the DNA was only partially digested (**Figure 9 b**). The fraction between 100-300 bp was extracted with a scalpel and processed for gel purification (NucleoSpin Gel, Macherey Nagel). The DNA was finally resuspended in 15 µl of 10mM Tris-HCl pH 8.5 buffer.

10.3.3 Naked DNA control

Extraction of the genomic DNA (gDNA) from the inner layer of three husks pooled together was performed using DNeasy Plant Mini kit (Qiagen) and following manufacturer's instructions. A total amount of 1.7 µg of gDNA was digested with 50 U/ml of DNase I recombinant (Roche) following a similar protocol than for chromatin (see DNase I digestion) (**Figure 9 a, b**).

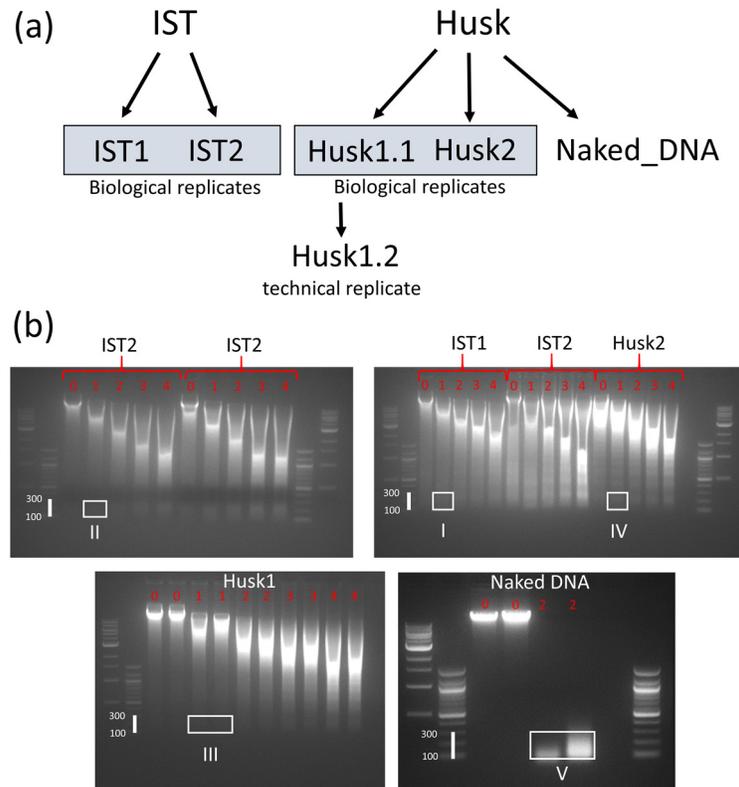


Figure 9: (a) Representation of the different libraries generated from the tissues IST and husk. (b) Agarose gels with ethidium bromide after electrophoresis containing the DNA derived from DNase I digestions performed to generate the five libraries: IST1, IST2, Husk1, Husk2, and Naked DNA, labelled I, II, III, IV, and V, respectively. The concentration of DNase I used were 0, 50, 100, 150, 200 U/ml, labelled 0, 1, 2, 3, and 4, respectively. The part of the gel used for extraction and to generate libraries are indicated by white rectangles. The white vertical bars on the ladders indicate the fraction of the gel selected (100 to 300 bp).

10.3.4 DNA quantification

The concentration of the DNA for each sample extracted from the gel after DNase I treatment was measured by fluorometry using the Quant-iT PicoGreen kit (Invitrogen). Two dilutions of the DNA were prepared by adding 2 μ l of the DNA solution in 398 μ l of Tris-EDTA buffer (TE) (10mM Tris-HCl, 1mM EDTA buffer) (1:200 dilution) and mixing 100 μ l of the 1:200 dilution with 100 μ l TE buffer (1:400 dilution). Each dilution was poured in triplicates (50 μ l each) on a black 96-wells plate (Corning, product #3694). A calibration curve spanning from 0 to 500 ng/ml (0, 12.5, 25, 50, 100, and 500 ng/ml) was made using the λ phage DNA (provided in the kit). Each calibration solution was poured in triplicate on the plate (50 μ l per well). The PicoGreen reagent was then prepared according to manufacturer's protocol (1:200 dilution of the PicoGreen reagent in TE) and mixed by pipetting in each well containing DNA or only TE buffer as a blank control. The plate was incubated 5 min at RT in the dark before measurement. The fluorescence of each well was then

measured on a Synergy 4 Hybrid Multi-Mode Microplate Reader (BioTek) using a wavelength of 480 nm for excitation and 520 nm for emission. The values of fluorescence were averaged by triplicate and subtracted from the blank background fluorescence. The corrected data were used to generate a standard curve of fluorescence in function of the DNA concentration and determine the concentration of DNA samples for the two dilutions. After multiplying by the dilution factors, DNA concentration values obtained ranged from 1 to 3 ng/ μ l.

10.3.5 Library preparation and sequencing

The DNA solutions were diluted to 1 ng/ μ l in a total volume of 10 μ l to construct libraries with the Ovation Ultralow DR Multiplex kit (NuGEN) following manufacturer's protocol with indicated modifications. The first library prepared (Husk1.1) was amplified with 18 cycles. Considering the important amount of DNA obtained and the presence of three distinct bands that could be linked to over amplification, libraries for IST1, IST2, Husk1.2 (technical replicate of Husk1.1), and Husk2 were amplified 15 times (**Figure 10**). After the purification step following the amplification and described in the protocol, an additional purification and size selection on gel was performed, as recommended by our genome center. To do so, 30 μ l of the eluted DNA was mixed with 5 μ l of CR loading dye and loaded on a 2% agarose gel containing ethidium bromide for separation by electrophoresis. DNA between 200-400 bp was extracted from the gel using a scalpel (**Figure 10**). DNA was then purified from agarose gel using PCR clean-up and gel extraction kit (Macherey-Nagel). Importantly, the melting of the gel in the commercial buffer should be performed at RT. Instead, a warming step at 50°C for speeding up gel melting after gel selection was performed for the library IST1, IST2, Husk1.2, and Husk2. Consequently, our genome center recommended to perform a new amplification step with Illumina adapters since fragments of the libraries may have been partially denatured during the gel melting at 50°C. Two additional cycles were therefore performed, making a total of 18 cycles as for the library Husk1.1 (made separately). At last, only 15 cycles were performed for Naked DNA sample (**Figure 10**). Different adapters provided by the NuGen kit were used for each library. The libraries were sent to sequencing on Illumina HiSeq 2500 platform to generate 100 bp single-end reads. The sequencing was performed in the Genome Center of the Max Planck Institute for Plant Breeding Research (Cologne, Germany). The libraries were sequenced in different runs according to the sequencing performance of each library (**Table 1**). Libraries IST1, Husk1.2, and Husk2 performed poorly during sequencing and needed to be rerun three times to obtain a proper amount of reads for analysis.

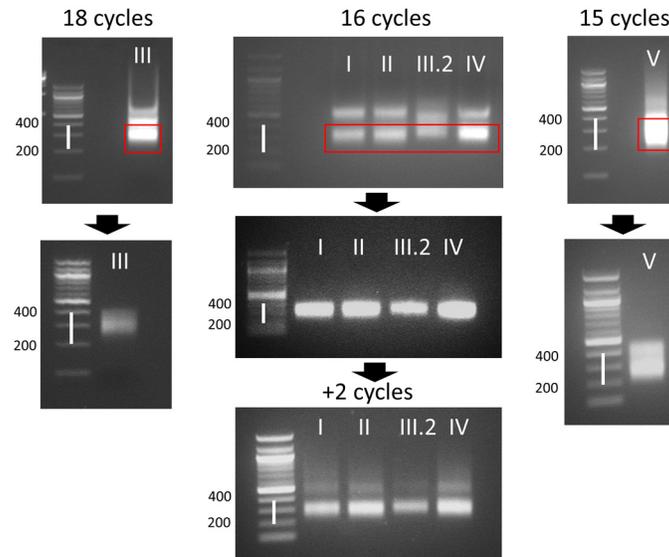


Figure 10: Libraries after amplification were separated on agarose gel (2%) by electrophoresis and the fraction between 200 and 400 bp was extracted (red rectangles). Gel purification was performed and an aliquot was separated again on agarose gel by electrophoresis to verify whether DNA was recovered and if the proper size is obtained. Libraries IST1, IST2, Husk1.1, Husk1.2, and Naked DNA are labelled with Roman numerals I, II, III, III.2, and IV, respectively. White bars indicate the fraction between 200 and 400 bp.

Table 1: Number of runs performed for each library and resulting number of reads obtained. The Roman numerals between parentheses indicate the library code used in **Figure 9 b** and **Figure 10**.

Library	Nb runs	Total reads
IST1 (I)	3	30.812.785
IST2 (II)	1	33.388.567
Husk1 (III)	1	34.911.866
Husk1.2 (III.2)	3	20.910.736
Husk2 (IV)	3	23.749.057
Naked DNA (V)	1	49.818.293

10.4 ChIP-qPCR

ChIP and following qPCR were performed exactly as described in (Haring et al., 2007) using 10 μ l of Histone 3 lysine 27 acetylation (H3K27ac) antibodies (Abcam, #ab4729), 2 μ l of H3 core antibodies (Abcam #ab1791), 10 μ l of H3 lysine 4 monomethylation (H3K4me1) antibodies (#ab8895), 10 μ l of H3K9ac (Abcam, #ab10812), and 20 μ l of rabbit serum (no antibody control, Sigma-Aldrich no. R9133).

Primers used for qPCR (Haring et al., 2010):

Control primers:

Ty1-copia type retrotransposon reverse transcriptase (copia)

F: 5'-CGATGTGAAGACAGCATTCCT-3'

R: 5'-CTCAAGTGACATCCCATGTGT-3'

MAc1 (J01238.1) (actin)

F: 5'-TTTAAGGCTGCTGCTACTGCTGTAGA-3'

R: 5'-CACTTTCTGCTCATGGTTTAAGG-3'

Primers at the *bl* locus:

5'R

F: 5'-CTCGGGGTCAAATGGACGG-3'

R: 5'-GCGGTCACAACCTTTTCAGAT-3'

R3

F: 5'-CAAGATCCATTGAACATCTTGTCC-3'

R: 5'-CATGTGTGAGGGTGATGCTGCG-3'

R6

F: 5'-GTTGTGACTGCAGTGTTAGGTAG-3'

R: 5'-CAAATTAGAGGGCTCCAAGAGGTC-3'

e

F: 5'-CTGGCGGCACTAAAAAACG-3'

R: 5'-TGTGCCACCTTTATTGTGAGTT-3'

f

F: 5'-CAACTGCTATGCGACTGATTGAT-3'

R: 5'-CCTGCTGTCCTTTCTTGTCTGA-3'

g

F: 5'-AGTACGTAACCTGCAAC-3'

R: 5'-AACTCAACGTACGTCACAAC-3'

j

F: 5'-ACACGATGGCCGGCAAT-3'

R: 5'-CCGCCGTTGGGTATGA-3'

k

F: 5'-CAAAGAGGCCGACTCGAC-3'

R: 5'-CAATGTCTTTCATATAACAGATCTGATACG-3'

l

F: 5'-TCCATCTCACCTCTCATTGTATCTTT-3'

R: 5'-AAATCGTGTATATGGTCGTTTAAAACA-3'

5'b

F: 5'-GGTGTGCACACCATTAATTGA-3'

R: 5'-CGATATTTTGGTGAAAAGTTC-3'

UTR

F: 5'-CTCTTCTGATCTTCTTCACCGTCTC-3'

R: 5'-ATTCCCGGGCGGCCGCGCCTTACTAATCCTTC-3'

ex3

F: 5'-AGGAAGGCGTAGGTCATGCAGAT-3'

R: 5'-ACCAGCTGCTCATGCAGAGGA-3'

10.5 ChIP-seq

The ChIP procedure was based on the original protocol from (Haring et al., 2007) with minor modifications and was performed by Blaise Weber (Dr. Maike Stam's lab, University of Amsterdam). In short, plant samples (five IST from V2 plants or 3 g of inner leaves from husk) were fixed with formaldehyde. Chromatin was extracted, and sonicated. The soluble fraction was then immunoprecipitated using 10 µl of H3K9ac antibodies (Abcam, #ab10812) or 20 µl of rabbit serum (no antibody control, Sigma-Aldrich no. R9133) in combination with magnetic beads (Diagenode, kch-802). Immunoprecipitated DNA was then recovered, decrosslinked and column-purified (Qiagen, 28104). For each ChIP-seq library, three ChIP samples were pooled yielding about 50 ng of DNA prior to adapter ligation and PCR amplification. Conversion of ChIP samples in sequencing libraries and PCR amplification was performed as indicated by the manufacturer using the KAPA Hyperprep kit (KAPA, KK8500). Efficiency of the conversion process was assessed by comparing the input DNA to the output library on the Agilent High Sensitivity D1000 ScreenTape System. Efficient conversion corresponded to a visible 100 bp shift in the fragment size and unbiased increase in DNA concentration. Adapters were home-made based on Illumina's True Seq Universal adapter sequences and each library was labeled with a specific index. For this, HPLC purified primers were ordered (IDT, phosphate group at the 5' end of the Indexed oligo, phosphorothiorate bond between the C and the T at the 3' end of the TrueSeq Universal primer), diluted to a concentration of 100µM in 10mM Tris-HCl pH 8 0.1mM EDTA. Each indexed primer was mixed with an equimolar amount of the Universal TrueSeq primer and the combination was annealed in a thermocycler, heated at 95°C for 5 min, and cooled down to 4°C at a rate of 0.1°C/s.

10.6 RNA-seq

The inner layers of the husk of three different plants and the IST of three different seedlings were harvested between Zeitgeber Time (ZT) 9 and 11 and pooled together, making one biological replicate. The samples were directly flash-frozen in liquid nitrogen (**Figure 8**). The same experiment was led in two different locations: Max Planck Institute for Plant Breeding Research, Cologne, Germany (MPI), and University of Amsterdam, the Netherlands (UvA) (**Figure 11**).

Three biological replicates per tissue were grinded in fine powder in liquid nitrogen. In a 2-ml microcentrifuge tube, 100 mg of powder was mixed vigorously on a vortexer with 1 ml of TRIzol (Thermoscientific) until homogenized. The samples were incubated 5 min at RT. A volume of 200 μ l of chloroform/isoamyl alcohol 24:1 (Sigma-Aldrich) was added and the tubes were homogenized on a vortexer for 15 s and incubated 3 min at RT. Tubes were centrifuged at 12,000 x g 15 min at 4°C. The top aqueous phase was transferred into a new 1.5 ml microcentrifuge tube and 500 μ l isopropanol was added for a 10 min incubation at RT. After 10 min precipitation at RT, the supernatant was transferred in two times in an RNeasy MINI spin column (Qiagen RNeasy kit) and centrifuged 15 s at 8000 x g. Flow-through was discarded and 700 μ l of RW1 buffer (Qiagen) was added. Two washing steps were performed using 500 μ l RPE buffer. RNA was eluted in 50 μ l RNase-free water and the concentration was assessed photospectrometrically. DNase I treatment (DNA-free kit, Ambion) was performed using a dilution of the RNA to 200 ng/ μ l as recommended by the manufacturer. The RNA was then pooled again and the volume was adjusted to 300 μ l with RNase-free water. A volume of 300 μ l of P/C/IAA was added and the mix was vortexed. Tubes were centrifuged at 13,000 x g 5 min at 4°C and the supernatant was transferred in a new tube for a second P/C/IAA extraction. The supernatant was transferred in a new tube and precipitation was performed by adding 560 μ l of 100% ethanol, 28 μ l of sodium acetate 3M, and 1 μ l of glycogen 10mg/ml. The mix was centrifuged at 13,000 x g 15 min at 4°C and the pellet was subsequently washed two times with 70% ethanol and finally resuspended in 20 μ l of RNase-free water after all ethanol has evaporated. Concentration was measured and 1 μ g of RNA was loaded on MOPS gel to check for RNA quality (p. 388 volume 1, Green and Sambrook, 2012). RNA concentration was adjusted to 4 μ g in a final volume of 10 μ l of RNase-free water. A total of 500 ng of total RNA was treated with the Ribo-Zero rRNA Removal Kit, Plant Leaf (Epicentre) to specifically remove ribosomal RNAs. RNA-seq libraries were prepared with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs) following manufacturer's protocol. Quality and quantity was assessed at all steps by capillary electrophoresis (Agilent Bioanalyser and Agilent TapeStation). Illumina Sequencing-by-synthesis libraries were quantified by fluorometry, immobilized and processed onto a flow cell with a cBot (Illumina) followed by sequencing-by-synthesis with TruSeq v3 chemistry on a HiSeq 2500. About 15 to 20 million single-end reads of 100 bp were obtained for each libraries, generating a total of \approx 100 million reads per tissues (three biological replicates multiplied by two locations).

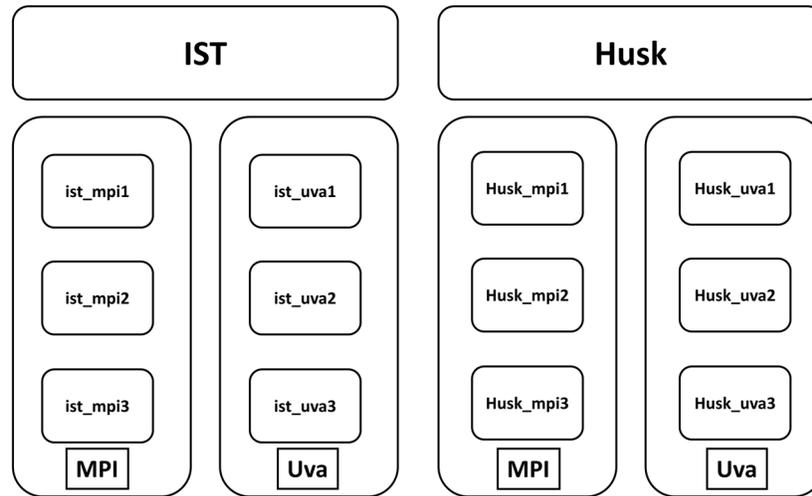


Figure 11: Experimental design of the RNA-seq experiment. Three biological replicates for each tissue were prepared in two different locations: MPI and UvA. Each biological replicate contains a pool of three different plants.

10.7 Bioinformatic analyses

10.7.1 Uniqueome

To assess the fraction of the genome that is uniquely mappable, i.e. not repetitive, Uniqueome program was used (Koehler et al., 2011). The reference genome B73 AGPv3 release 24 fasta sequences for each chromosome were downloaded from Gramene database (<ftp://ftp.ensemblgenomes.org/>) and used as input. The pipeline indicated in Supplementary_File_2.pdf from (Koehler et al., 2011) was used with the parameters filter=0 (non-exhaustive search), and mode=88,2 (length of the tag and number of mismatches allowed). Bedgraph files were generated for each chromosome to display the uniqueome in a genome browser.

10.7.2 Read quality processing

The quality of the reads from all datasets (DNase-seq, ChIP-seq, RNA-seq) was first assessed using FastQC (<http://www.bioinformatics.babraham.ac.uk/>). Reads derived from DNase-seq and ChIP-seq were filtered to remove artifacts and reads of low quality using FastX toolkit (Gordon et al., 2010 http://hannonlab.cshl.edu/fastx_toolkit/). Reads were trimmed based on quality and reads of less than 70 bp were removed using PRINSEQ (Schmieder and Edwards, 2011). For reads from RNA-seq, quality filtering was performed using Trimmomatic (Bolger et al., 2014).

Script for fastq file processing:

```
$fastx_artifacts_filter -Q 33 -i <input.fastq> -o <output.fastq>
$perl prinseq-lite.pl -fastq <input.fastq> -trim_qual_right 20 -trim_qual_left 20 \
```

```
-min_len 70 -out_good stdout > <output.fastq>
$fastq_quality_filter -q 20 -p 80 -Q 33 -v -i <input.fastq> -o <output.fastq>
```

10.7.3 Mapping

Reads derived from DNase-seq and ChIP-seq were mapped with Burrow-Wheeler-Algorithm (BWA) to the reference *Z. mays* reference genome B73 AGPv3 release 24 (Li and Durbin, 2009). For RNA-seq, reads were mapped using TopHat2, which uses the Bowtie2 aligner (Kim et al., 2013).

Script for BWA mapping:

```
$bwa bwsw <reference.genome> <input.fastq> > <output.sam> #map the reads
$samtools view -b -S <input.sam> -o <output.bam> #converts sam to bam
$samtools sort <input.bam> <output.bam> # sort bam
$samtools view -q 20 <input.bam> -b > <output.bam> #select uniquely mapped reads
```

Script TopHat2 mapping:

```
$tophat2 -o /path/outdir --library-type fr-unstranded <reference_genome> <input.fastq>
```

10.7.4 Peak calling

The definition of significantly enriched regions for DNase-seq and ChIP-seq was performed using Model-based Analysis of ChIP-Seq (MACS2) (Zhang et al., 2008). The peak calling was first run on individual libraries in order to compare the similarities between replicates. To perform the comparison between libraries, the bedgraph files generated (with the option --bdg in MACS) were compared using WigCorrelate, which provides the Pearson product-moment correlation coefficient (Bewick et al., 2003). For the final list of candidates, the biological replicates were pooled before peak calling in one bam file using bedtools (Quinlan and Hall, 2010).

Script for peak calling:

```
$macs2 callpeak \
--broad \ #allows clustering of nearby enriched regions
--broad-cutoff 0.001 \ #set a specific pvalue for broad regions
--qvalue 0.001 \ #defines a cutoff to call significant regions
--gsize 8e8 \ #mappable size of the genome defined by uniqueome: 800 Mb (38% of the
2000 Mb assembled)
--bdg \ #generates a fragment pileup profile
--SPMR \ #Signal Per Million Reads. In combination with --bdg, provides a normalized read
count for visualization in a genome browser
--tsize 100 \ #tag size, single-end reads of 100 bp were obtained from sequencing
-t <input.bam> \ #DNase-seq or ChIP-seq bam file (pooled biological replicates)
```

```
--control <control.bam> \ #bam file from Naked DNA (DNase-seq) or no-antibody control  
(ChIP-seq)  
--down-sample \ #scale down the bigger sample
```

10.7.5 RNA-seq analysis

RNA-seq analyses was performed with the help of Erwin Datema (Keygene, Wageningen) and in collaboration with Rurika Oka. Once mapped using TopHat2, reads were assembled and counted with Cufflinks and the different libraries were merged with Cuffmerge (Trapnell et al., 2012). To assess differential expression of the genes across replicates and tissues, Cuffdiff was performed for each location (MPI and UvA) following the commands described in (Trapnell et al., 2012). In parallel, mapped reads were processed for read count using HTSeq (Anders et al., 2015). A home-made script was used to merge the count files from the different libraries. The merged file were then processed in the R bioconductor package DESeq to test differential expression by integrating the effect of the location in a generalized linear model (Anders and Huber, 2010). The significantly Differentially Expressed Genes (DEGs) between IST and husk detected by the two pipelines were compared to assess the overlap. To do so, a list of significant DEGs with a fold change greater than two were selected from DESeq, Cuffdiff MPI, and Cuffdiff UvA outputs. The overlap of the DEGs in each subset and the significance of the overlap considering the $\approx 15,000$ DEGs from the initial datasets was calculated using the R package geneOverlap from Bioconductor (Fisher's exact test) (Shen, 2013). Venn diagrams were generated using VennDiagram R package (Chen and Boutros, 2011). To generate a principal component analysis before fitting the models in DESeq, the plotPCA function from the DESeq package was used.

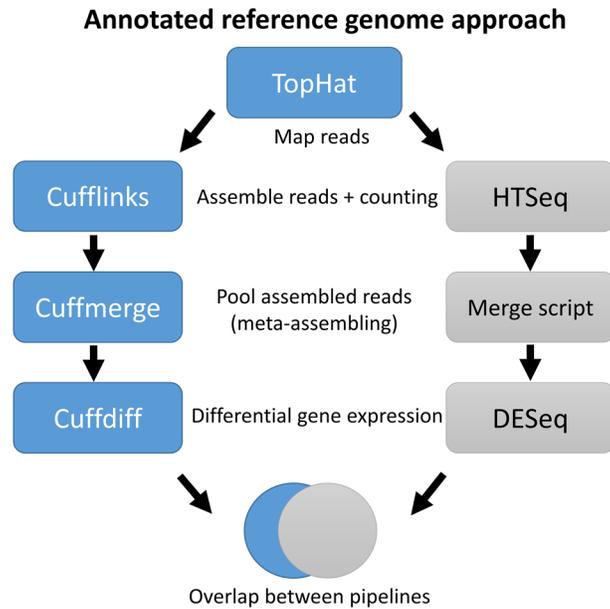


Figure 12: RNA-seq pipelines used. Reads were mapped with TopHat2 and then processed either in the Tuxedo pipeline (Cufflink, Cuffmerge, Cuffdiff) or in the DESeq pipeline (HTSeq, merging, DESeq). The overlap of significant DEGs could be compared between the two pipelines.

11 Results

11.1 Mappability of *Z. mays* genome

About 85% of the *Z. mays* genome is constituted of repetitive sequences (Schnable et al., 2009). Repetitive sequences make unique mapping of short reads difficult and reduce therefore the mappability of the genome (Treangen and Salzberg, 2011). To take into account the repetitive nature of the *Z. mays* genome in our analyses, we calculated the mappable portion of the genome, called hereafter uniqueome. To do so, we used the ISAS (Imagenix, USA) alignment algorithms (Koehler et al., 2011). Shortly, the ISAS algorithms perform an all-against-all alignment of the genome, which provides the uniqueness of each bp according to a defined read length and a number of mismatches allowed (see details in materials and methods). We used single-end reads of 100 bp for the sequencing of DNase-seq, ChIP-seq and RNA-seq libraries. According to this read length, we found that uniqueome of *Z. mays* constitutes $\approx 38\%$ of the genome with ≈ 800 Mb uniquely mapped out of ≈ 2000 Mb golden path length (reference assembly). We see that although TEs represent $\approx 70\%$ of the assembled genome (**Figure 13 a**), only 50% are found in the uniqueome (**Figure 13 b**). Conversely, genic regions represent a bigger part of the uniqueome with $\approx 20\%$ compared to $\approx 7\%$ of the genome (**Figure 13 b**). These results are consistent with the repetitive nature of TEs compared to genes (Treangen and Salzberg, 2011). According to the estimation of *Z. mays* uniqueome with our read size, only $\approx 38\%$ of evenly distributed 100 bp reads along the

genome would theoretically be uniquely mapped reads. This is important to interpret the data generated since our view on the *Z. mays* genome is restricted to uniquely mappable regions.

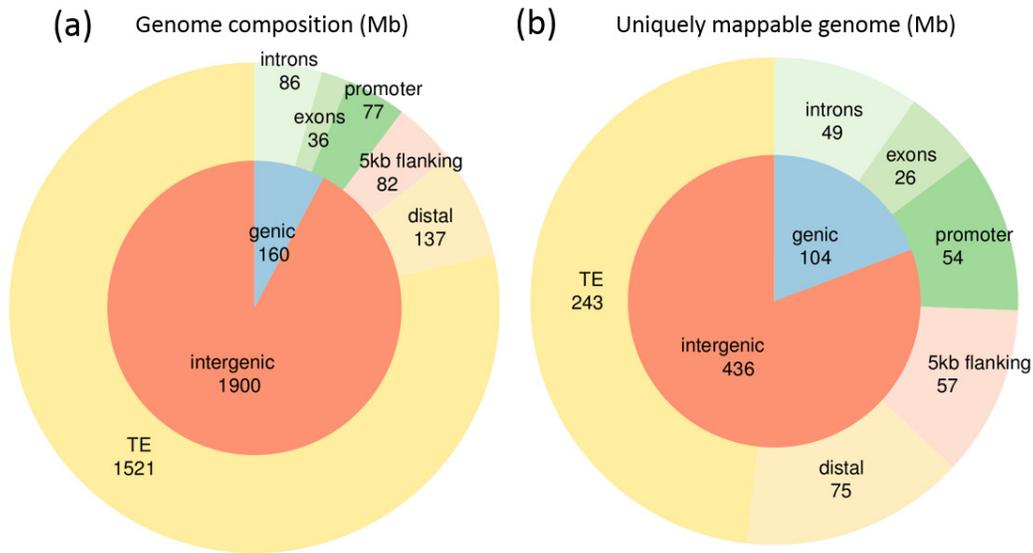


Figure 13: (a) Distribution of the different genomic features of *Z. mays* (in Mb). (b) Fraction of the genome uniquely mappable within different genomic features (in Mb) (pie charts from Rurika Oka).

11.2 Chromatin accessibility profile in *Z. mays*

11.2.1 DNase I assay

To profile chromatin accessibility in *Z. mays* and predict enhancer location in a tissue-specific manner, we performed DNase I-sequencing (DNase-seq) using two types of tissues: Inner Stem Tissue (IST) of V2 seedlings and husk tissue which come from pre- and post-flowering stages, respectively (**Figure 14 a**). These two tissues were chosen for two reasons: (1) they provide a view on the chromatin and the transcriptome at two distinct developmental stages; (2) they are soft tissues from which nuclei preparation is easy to proceed. The method used to obtain DNase-seq libraries was based on the isolation of the 100-300 bp fraction of digested DNA after mild DNase I-chromatin digestion (Wang et al., 2012). This method is faster and more straightforward compared to the initial protocol for DNase-seq (Boyle et al., 2008) (see details in material and methods). In total, five libraries were sequenced with two biological replicates per tissue, one technical replicate for husk, and one library made from naked DNA (**Figure 14 b**). Naked DNA was obtained by digesting deproteinized DNA extracted from husk. This sample provides a control for the cleavage bias of DNase I endonuclease and enables normalization of the DNase-seq profiles for samples derived from digested chromatin (He et al., 2014). The technical replicate allowed to assess the variation due to sequencing.

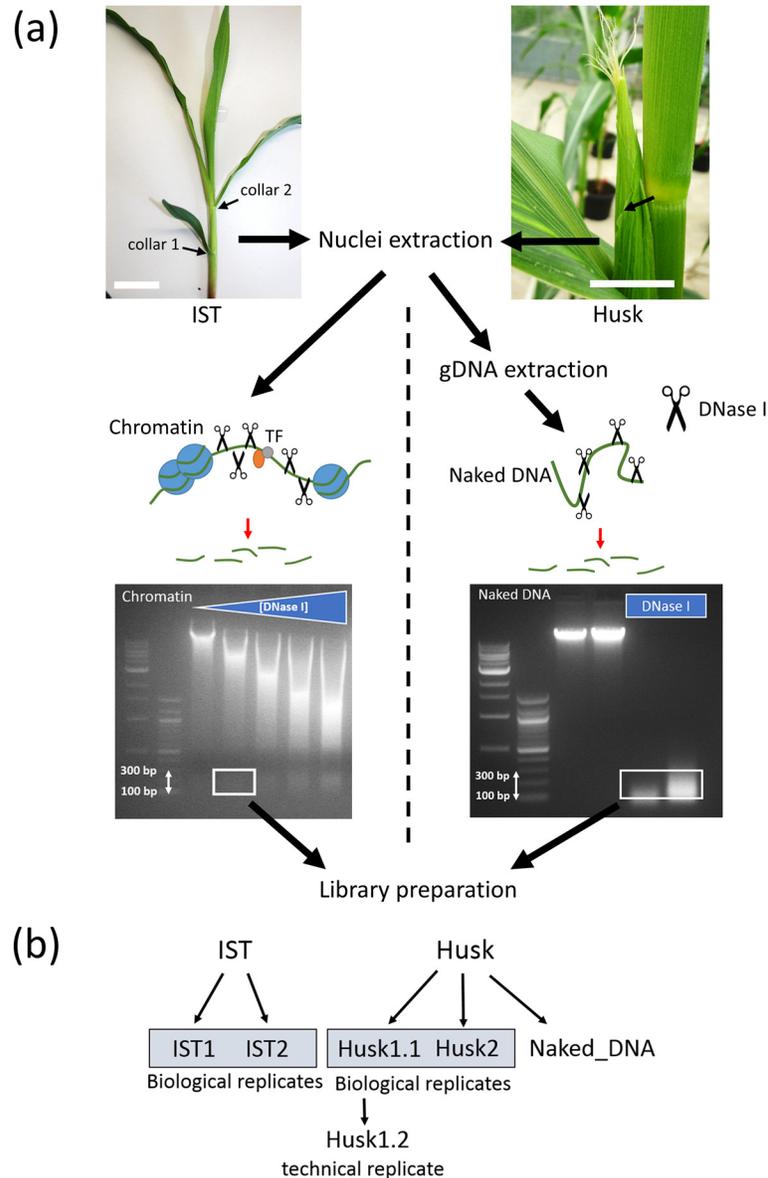


Figure 14: (a) Pipeline for DNase I assay. Nuclei are extracted from Inner Stem Tissue (IST) taken from V2 seedlings (two collars visible) and husk, and separated into two fractions. The scale bar in white represents ≈ 5 cm. On the left panel, one fraction of nuclei is digested directly with different concentrations of DNase I, which preferentially cuts at unprotected DNA, i.e. not bound by nucleosomes (in blue) or TFs (in orange and grey). The DNA is then extracted and separated on agarose gel by electrophoresis. The appearance of the smears of digested DNA allows to determine which sample should be processed considering that over digestion results in increased background while under digestion results in limited amount of extractible DNA usable for library preparation. In this case, the second lane indicates a partial digestion (gel on the left panel) in which most of the DNA is still in large fragments and seems appropriate for gel extraction (white rectangle). On the right panel, naked DNA is extracted from nuclei digested with the concentration of DNase I selected for the chromatin on the left panel. The digested naked DNA is separated on agarose gel by electrophoresis. Fraction of DNA ranging between 100 and 300 bp (white rectangle) for the selected concentration of DNase I is extracted from gel for both chromatin and naked DNA for subsequent library preparation. (b) Representation of the different libraries generated from the tissues IST and husk.

11.2.2 DNase-seq analysis

DNase-seq libraries were sequenced to obtain between 20 and 50 million 100 bp single-end reads. The pipeline for bioinformatics analysis of DNase-seq is summarized in **Figure 15**. The overall quality of the reads for each library was good with most of the bases having a Phred quality score (Q) above 20. A Phred quality score is a measure of the quality of each nucleotide within a read defining a probability of identification accuracy with P being the error probability such as $Q = -10 \log_{10} P$. For instance, $Q = 20$ represents a base call accuracy of 99% (i.e. $P=1\%$) (Ewing et al., 1998). The duplication level in libraries was also assessed. Duplicated reads can originate from a real enrichment but can also be linked to technical artifacts such as PCR over amplification or optical duplicates during sequencing. The duplication level was calculated with the initial reads before mapping. Apart from library IST1, which had a rather high duplication level with only 60% reads left after deduplication, remaining libraries had an acceptable level of duplication with between 70 and 80% of reads remaining after deduplication (**Table 2**). Duplicated reads are removed automatically during peak calling. The reads were then mapped to B73 *Z. mays* reference genome and filtered to keep only uniquely mapped reads. Most of the reads were mapped for the different libraries ($\approx 90\%$) except for naked DNA (65%) (**Table 2**). The difference in mapping can be due to the fact that only ≈ 2000 Mb of the genome out of an estimated size of ≈ 3200 Mb is assembled ($\approx 57\%$)². Our data suggest that accessible chromatin regions are mostly located in assembled regions. In contrary, reads from naked DNA represent randomly distributed fragments of DNA across the genome, including poorly assembled regions. This is what we observed with a smaller fractions of the reads mapped for the naked DNA library. In comparison, 70% of uniquely mapped reads were obtained from DNase-seq in rice (Zhang et al., 2012a) considering that rice genome contains an estimated repetitive fraction of 50% (Nagano et al., 1999).

The removal of reads mapped at multiple locations, also called multireads, introduces an important bias in subsequent analyses as enhancers located in mappable but repetitive regions are discarded. The choice of discarding multireads is based on a balance between the loss of accessible regions located in repetitive elements and the reduction of false positives (Pepke et al., 2009). We chose to increase specificity by using only uniquely mapped reads instead of increasing sensitivity by keeping multireads. Uniquely mapped reads constituted between 50% and 62% of mapped reads (**Table 2**). The fraction of multireads is similar for all samples including naked DNA ($\approx 40\text{-}50\%$), indicating that both accessible and closed chromatin regions can be intrinsically repetitive in the *Z. mays* genome. To confirm this, we assessed the overlap between multireads and genic regions and found that between 15% and 32% of multireads overlapped with genic regions for the samples derived from digested chromatin while only 4% of multireads from naked DNA did (**Table 3**).

² http://plants.ensembl.org/Zea_mays/Info/Annotation/#assembly (consulted on March 2016)

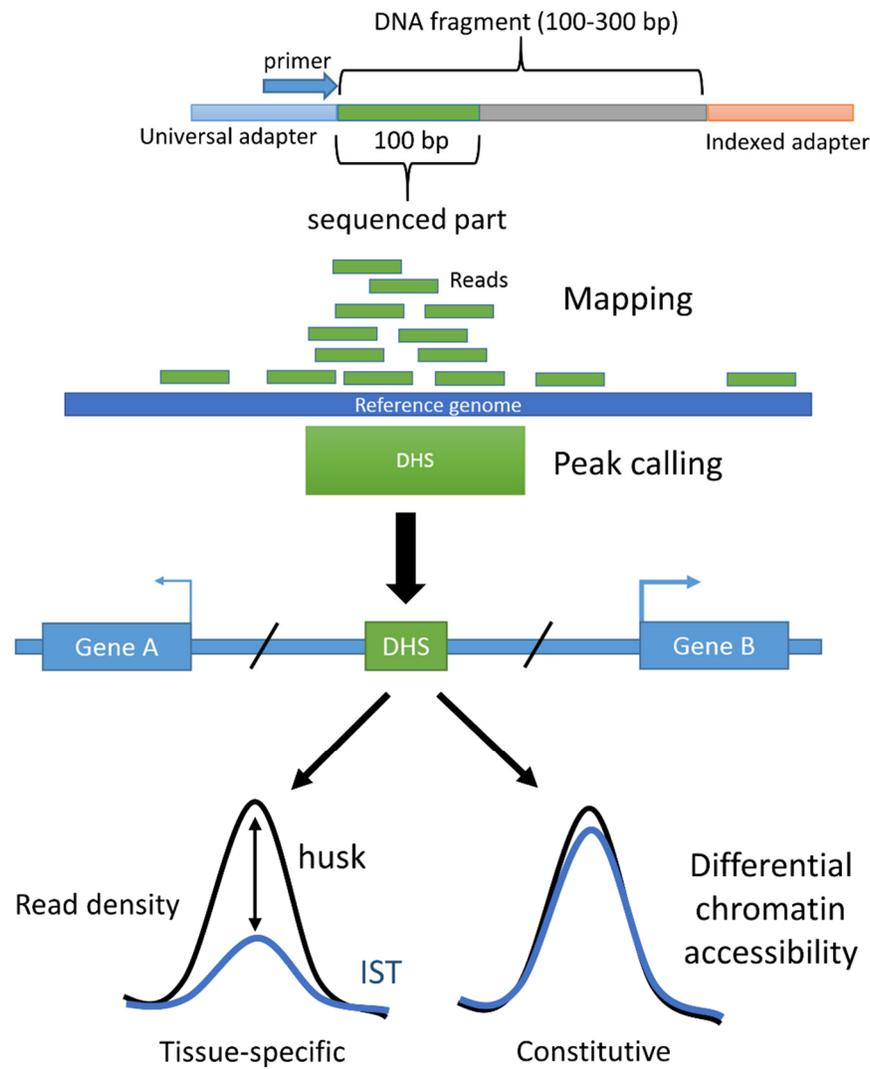


Figure 15: Simplified bioinformatics pipeline. The Illumina single-end reads provide 100 bp reads which can be mapped on the reference genome B73. Peak calling defines regions with significant enrichment in read density, called DHSs, compared to the background. Differences of chromatin accessibility of intergenic DHSs are assessed between tissues to determine putative “tissue-specific” enhancers. “Constitutive” indicates here DHSs for which chromatin accessibility does not vary between IST and husk.

Table 2: Table summarizing DNase-seq data obtained for the five libraries. Initial reads in million (M) designate the amount of reads obtained from the sequencing. The % of mapped reads represent the fraction of initial reads that could be mapped. The % of uniquely mapped reads represent the fraction of mapped reads that could be mapped at only one location in the genome. The number of usable reads is the resulting number of reads after discarding reads that are unmapped or mapped at different locations in the genome. The % of reads remaining after deduplication provides an estimation of the level of duplication due to PCR over amplification or sequencing bias in each libraries.

	IST1	IST2	Husk1	Husk1.2	Husk2	Naked DNA
Initial reads (M)	30,8	33,4	34,9	20,9	23,7	49,8
% mapped reads	93%	94%	95%	96%	95%	65%
% uniquely mapped reads	51%	59%	60%	62%	61%	62%
Usable reads (M)	14,5	18,4	19,8	12,4	13,7	20,2
% remaining reads after deduplication	60%	70%	74%	79%	77%	80%

Table 3: Overlapping of the multireads with genic regions (values in bp).

	IST1	IST2	Husk1	Husk1.2	Husk2	Naked DNA
Multireads	14.207.076	12.998.860	13.234.065	7.616.024	8.802.601	12.400.042
Overlap with genic regions	4.533.660	3.203.095	2.466.846	1.131.516	1.407.104	470.765
%	31,9%	24,6%	18,6%	14,9%	16,0%	3,8%

11.2.3 DHSs are enriched at promoters and coding regions

Uniquely mapped reads were used for peak calling in order to define significantly enriched regions hereafter called DNase I Hypersensitive Sites (DHS). Peak calling can be performed by several algorithms also used for ChIP-seq analyses (Koohy et al., 2014). Similarly to ChIP-seq, DNase-seq sequencing data generated by size-selection provide densities of reads that should be higher than the background at the regions of interest, i.e. target protein-bound regions for ChIP-seq and DHSs for DNase-seq. We selected Model-based Analysis of ChIP-Seq (MACS2) for peak calling and identifying DHSs genome-wide (analysis performed in collaboration with Rurika Oka) (Zhang et al., 2008). The distribution of the read density and the DHSs defined for each tissue and for the naked DNA control can be exemplified in **Figure 16**. As expected, a clear enrichment in accessible chromatin was found in genic regions, especially at TSSs, which colocalize with most of identified DHSs in this example. As expected, read density for naked DNA did not show strong peaks as for chromatin sample but rather a diffuse signal (**Figure 16**). The fraction of the genome in the uniqueome is concentrated mostly in genic and promoter regions, and therefore, all DHSs present outside of these regions are undetectable in this example (**Figure 16**).

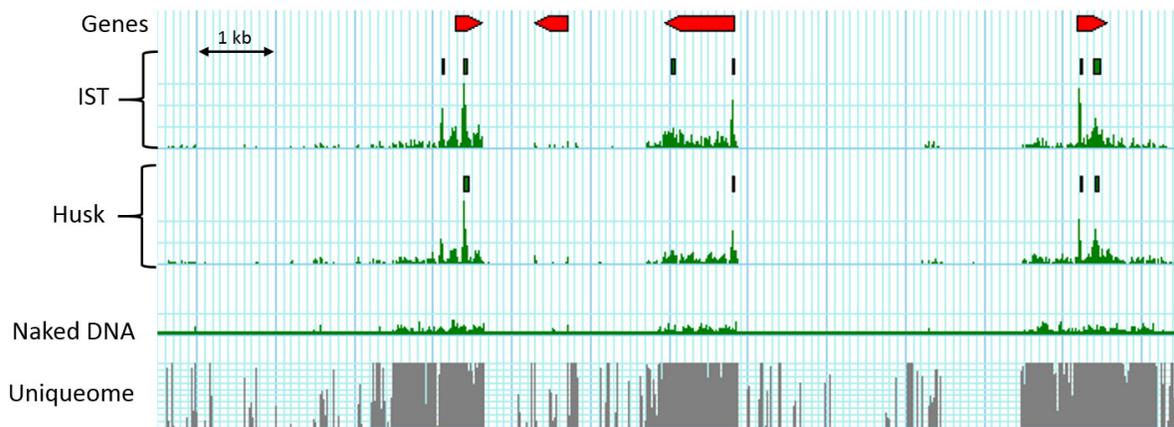


Figure 16: Example of DNase-seq profile. Genes are represented in red, DHSs are represented by green rectangles and the read density is indicated below in green (signal per million reads). The uniqueome track (in grey) represents the mappable fraction of the genome. In this example, DHSs are overlapping mainly with the TSSs and 3' ends of genes.

To evaluate DHSs distribution, we defined promoters as the region from 1 kb upstream to 1 kb downstream of the TSSs of annotated genes (Hesselberth et al., 2009) (**Figure 17 a**). The distribution of DHSs genome-wide is clearly biased toward promoter and genic regions with about 50 to 60% of identified DHSs in husk and IST, respectively (**Figure 17 c, d**). Although intergenic regions account for $\approx 80\%$ of the *Z. mays* uniqueome (**Figure 17 b**), they contain only $\approx 50\%$ of defined DHSs (**Figure 17 c, d**). These results are consistent with DNase-seq data in rice, with $\approx 45\%$ of the identified DHSs locating in intergenic regions (Zhang et al., 2012a). In *A. thaliana*, only 15% of DHSs were located in intergenic regions, which can be explained by the compact nature of its genome (Zhang et al., 2012b).

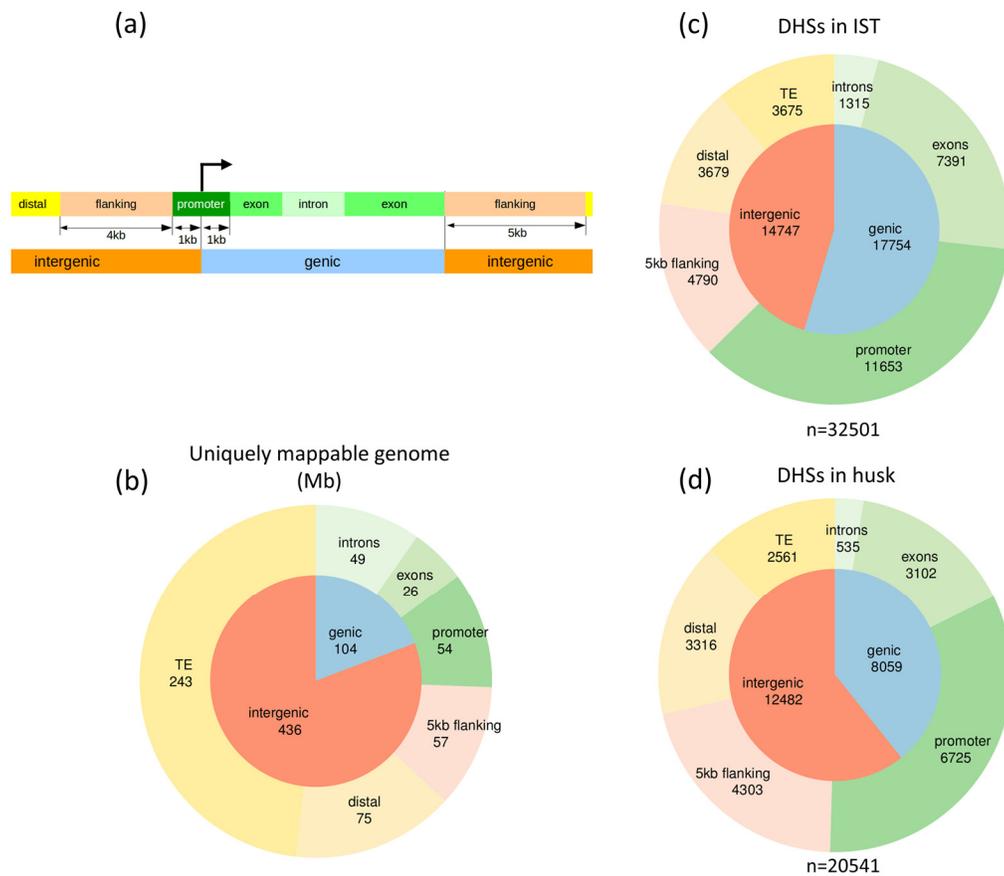


Figure 17: (a) Definition of the different genomic features used to classify DHSs. (b) Distribution of the different genomic features in the estimated uniqueome (in Mb). DHSs distribution in different genomic features for IST (c) and husk (d). The total number of DHSs is indicated under each pie charts (figures from Rurika Oka).

11.3 Active histone mark H3K9ac profile in *Z. mays*

11.3.1 Validation of the histone mark H3K9ac at the *b1* locus

To improve enhancer prediction based on DHSs, we wanted to integrate the profile of an enhancer-specific histone mark. H3K4me1 is associated with both active and poised enhancers while H3K27ac and H3K9ac are associated with active enhancers in animals (Heintzman et al., 2007; Creighton et al., 2010; Karmodiya et al., 2012). We tested these three histone marks at the *b1* locus in *Z. mays* plants containing an active *b1* allele, called *B-I*. In *B-I*, a hepta-repeat located 100 kb upstream of *b1* TSS acts as enhancer (Stam et al., 2002). The hepta-repeat is inactive in young seedlings while it becomes active in certain tissue of mature plants such as husk (**Figure 7**) (Stam et al., 2002). It was previously shown that H3K9ac and H3K14ac active histone marks were enriched at the hepta-repeat when *b1* is highly expressed (Haring et al., 2010). We wanted to test more precisely the level of enrichment for H3K9ac alone and verify whether H3K27ac also correlates with active enhancers as it was observed in mammals (Creighton et al., 2010; Karmodiya et al.,

2012). We performed ChIP with the tissues used for DNase-seq: IST and husk for *bl* inactive and active stages, respectively (**Figure 7**). As controls, we used the constitutively expressed gene *actin1*, which was previously shown to be enriched in H3K9ac, H3K14ac, and H3K4me2; and a Ty1-copia retrotransposon element, which was previously shown to be depleted for these histone marks (Haring et al., 2007). We found that both H3K9ac and H3K27ac were associated with *bl* expression at both the enhancer (R3-R6) and the coding region of *bl* (**Figure 18**). We verified for each ChIP and each region tested the level of background by using rabbit serum as no-antibody control. The percentage of input and quantity of DNA of the no-antibody control was for each case negligible (data not shown). H3K9ac and H3K27ac were also enriched in non-repetitive regions (e, g, and l, **Figure 18 a, b**), which were previously shown to be involved in the multi-loop structure formed in husk, in which *bl* is highly expressed (Louwers et al., 2009) (**Figure 7**). In contrary, H3K4me1 enrichment was low at the enhancer of *bl* in both tissues while it was relatively high at the coding regions of *bl* in husk (**Figure 18 b**). In *A. thaliana*, H3K4me1 was distributed mainly in transcribed regions but did not correlate with transcriptional activation (Zhang et al., 2009). Our results suggest that H3K9ac and H3K27ac are enriched at the active hepta-repeat and the coding region of *bl*, when transcriptionally active. H3K4me1 did not seem to correlate with the hepta-repeat activity and was therefore not retained for further analyses. Although H3K27ac and H3K9ac marks provided similar trends, we found that the latter was a better indicator of the activity of the hepta-repeat with a stronger differential signal between active and inactive enhancer states (**Figure 18 b**).

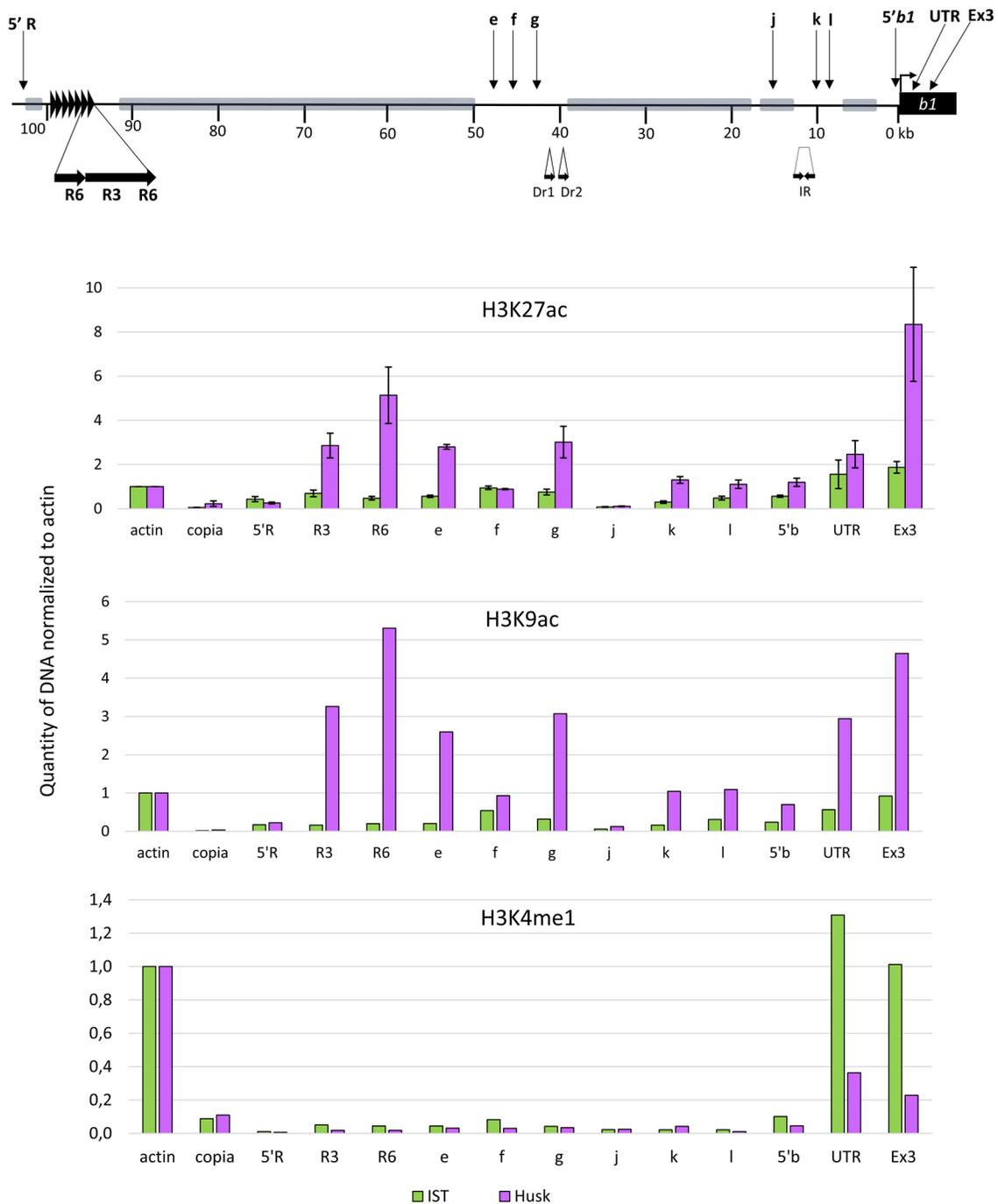


Figure 18: ChIP analysis at *b1* for the histone marks H3K27ac, H3K9ac, and H3K4me1. **(a)** The *b1* locus is described with the location of the primers used for ChIP-qPCR. The seven black arrows located about 100 kb upstream of *b1* represent the seven direct repeats of the enhancer element hepta-repeat. Dr1 and Dr2 indicate Direct repeats and IR indicates Inverted Repeats about 40 and 10 kb upstream of *b1*, respectively. Gray bars indicate repetitive regions (adapted from Louwers *et al.*, 2009). **(b)** Relative amounts of DNA normalized to actin are shown for the three antibodies tested: H3K27ac, H3K9ac, and H3K4me1 for IST (green) and husk (purple). The regions tested are *actin1* (actin) and a Ty1-copia retrotransposon element (copia) as positive and negative controls, respectively. Several regions of the *b1* locus were

examined: a nucleosome-depleted region upstream of hepta-repeat (5'R); two regions (repeated seven times) of the hepta-repeat (R3-R6); six regions located in non-repetitive regions between the hepta-repeat and *bl* (e, f, g, h, k and l); a repetitive region (j); a promoter region of *bl* (5'b); a region of the UTR (UTR); and at last a part of the exon 3 of *bl* (Ex3). For H3K27ac ChIP, three biological replicates were performed (apart from the regions j and k, which were tested with only two biological replicates). Error bars represent Standard Errors of the Mean (SEM) for the biological replicates. ChIP for H3K9ac and H3K4me1 histone marks was performed once. The signals levels for the no-antibody control were negligible for all primer sets used and are not shown.

11.3.2 Genome-wide H3K9ac enrichment profile in *Z. mays*

Since H3K9ac was a good predictor of the enhancer activity at the *bl* locus, ChIP-seq was performed to obtain a genome-wide distribution of H3K9ac enrichment. The combination of accessible chromatin (DNase-seq) and active histone mark (ChIP-seq) data should lead to a better prediction of enhancers than either method alone (Shlyueva et al., 2014). To be able to compare DNase-seq and ChIP-seq data, we used the same inbred line B73 and the same tissues: IST and husk. Two biological replicates were obtained per tissue with two samples each, one for H3K9ac and the other for the no-antibody control (Figure 19). ChIP-seq experiment and bioinformatic analyses were performed by Blaise Weber and Rurika Oka, respectively (Group of Dr Maike Stam, University of Amsterdam). As for DNase-seq, the peak calling software MACS2 was used to identify enriched regions for H3K9ac. We found that, as for DHSs, the H3K9ac histone mark was especially enriched in promoter regions with about 50% of H3K9ac enrichment peaks for both tissues (Figure 20 a, b, c). In total, we found $\approx 14,000$ H3K9ac enrichment peaks located in intergenic regions, an amount relatively similar to the number of DHSs found in intergenic regions ($\approx 12,500-15,000$) (Figure 20 d). Conversely, 1.5 and 3.5 times more H3K9ac peaks than DHSs were found in genic region in IST and husk, respectively (Figure 20 d). Globally, $\approx 27\%$ and $\approx 37\%$ of DHSs overlapped with H3K9ac peaks in IST and husk respectively. This is more than expected if the two features were not related according to a Fisher's exact test (p-values obtained $< 1\%$).

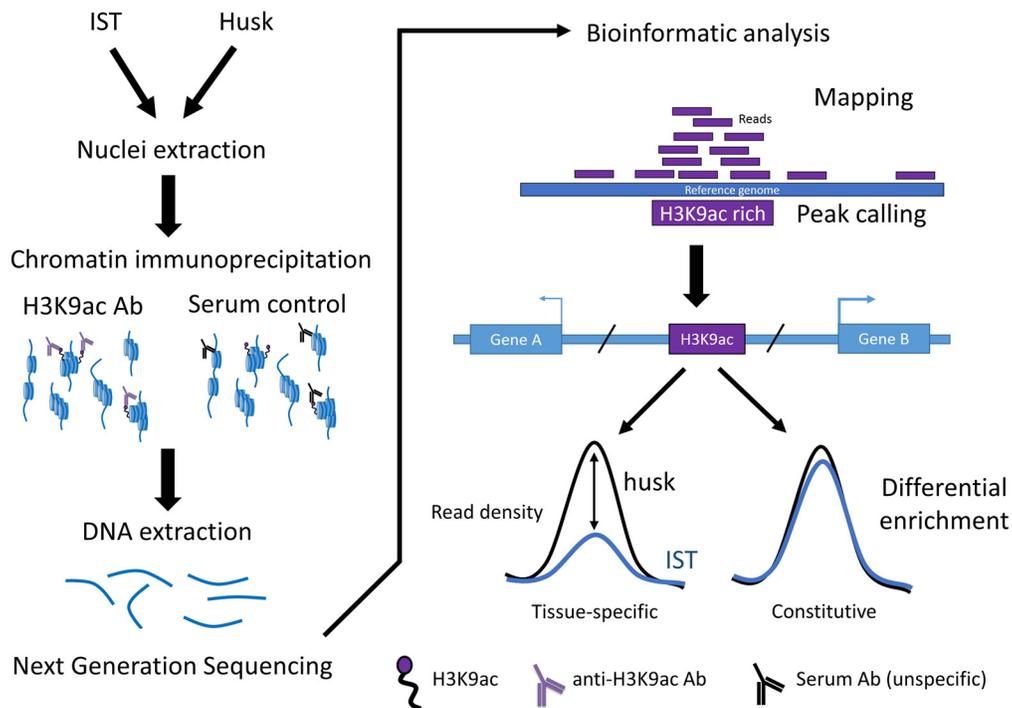


Figure 19: Simplified pipeline for ChIP-seq. Crosslinked chromatin is extracted from IST and husk nuclei to be precipitated with either specific antibodies (Ab) (e.g. anti-H3K9ac) or rabbit serum (no-antibody control) which contains unspecific antibodies. DNA fragments interacting with nucleosomes containing H3K9ac are purified and sequenced. The resulting reads are mapped onto the genome and regions with read density significantly enriched above background (compared to the read density provided by the no-antibody control) are delimited using a peak calling algorithm. Differential enrichment between tissues can be calculated to distinguish tissue-specific H3K9ac enrichment peaks from constitutive ones.

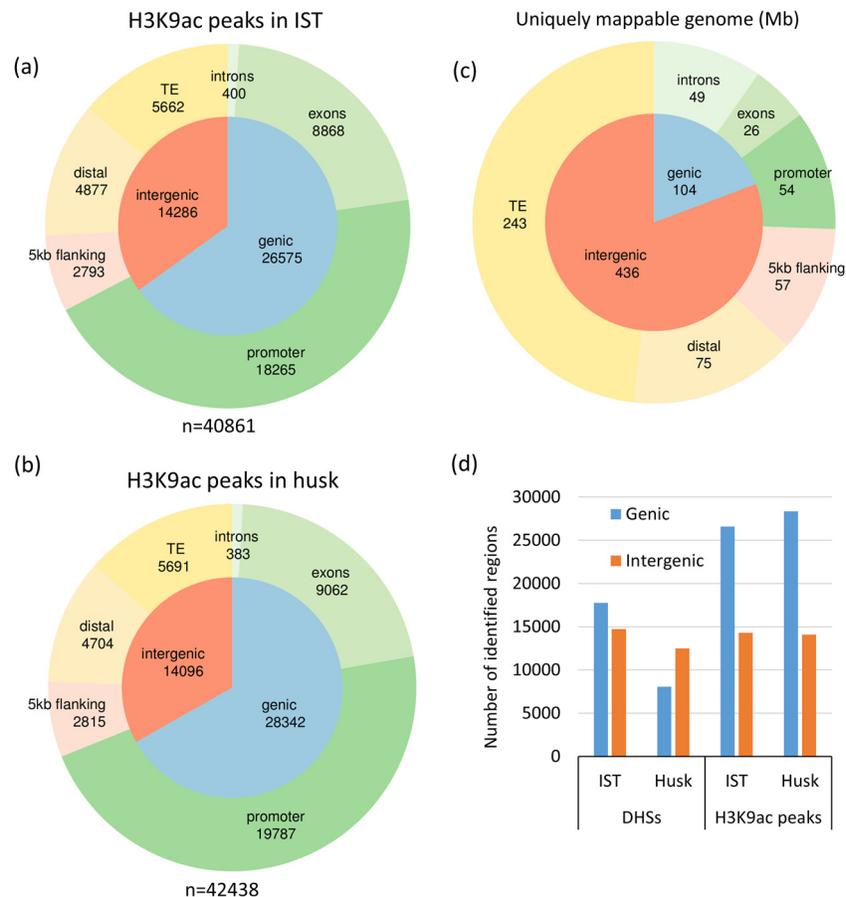


Figure 20: H3K9ac enriched peaks distribution in different genomic features for IST (a) and husk (b). (c) Distribution of the different genomic features in the estimated uniqueome (in Mb). (d) Comparison of the distribution of DHSs and H3K9ac peaks in genic and intergenic regions. Pie charts from Rurika Oka.

11.4 Transcriptome analysis

In addition to chromatin accessibility and H3K9ac enrichment profiles in the two tissues used, we wanted to find DEGs between the tissues and connect the chromatin features at putative enhancers with the expression of nearby genes. We therefore performed RNA-seq on IST and husk. Six biological replicates were performed per tissue with three biological replicates in two locations: MPI in Cologne, Germany (MPI) and University of Amsterdam, the Netherlands (UvA). The growth conditions in the two locations were slightly different with one hour of difference for the photoperiod (16h at the MPI and 15h at the UvA). We assumed that stage- and tissue-specific genes should not be strongly affected by the different growth conditions received. The six biological replicates were sequenced to generate ≈ 15 -19 million reads each, giving a total of ≈ 100 million single-end reads of 100 bp per tissue. Several approaches can be used to characterize transcriptomes and the outcome of the analysis will vary according to the chosen package (Seyednasrollah et al., 2013). We selected two of the most commonly used packages in transcriptomics to detect DEGs: (1) Cuffdiff which is usually used within the Tuxedo pipeline (Trapnell et al., 2012) and (2) DESeq

which enables multi-factor designs for statistical analysis (Anders et al., 2013). We first assessed the distribution of the variance in the read count performed for each RNA-seq libraries using the DESeq pipeline. To do so, we performed a principal component analysis with variance stabilizing transformation from the fitted dispersion-mean relation. As expected, most of the variance found is mainly explained by the factor “tissue type”, although the factor “location” seem to explain an important part of the variance for husk (**Figure 21**). The number of significant DEGs was between 15,000 and 19,000 according to the approach used. Since the number of DEGs represented almost all genes expressed in the tissues, we selected significant DEGs with a fold change in expression greater than two. We could narrow down our list to about 2800-8400 DEGs and overlapped these obtained with the different methods (**Figure 22 a, b, c, d**). Although many DEGs did not overlap between the different subsets of DEGs, Fisher’s exact tests indicated that the overlaps were in each case significantly different than a random overlap considering the $\approx 15,000$ DEGs found by the different methods (**Figure 22 a, b, c, d**). These results indicate that our RNA-seq datasets are providing robust data considering the location of growth or the package used for analysis. In order to identify tissue-specific enhancers and their target genes, we decided to select the overlapping subset of 5326 DEGs found with Cuffdiff outputs for the two locations (**Figure 22 c**).

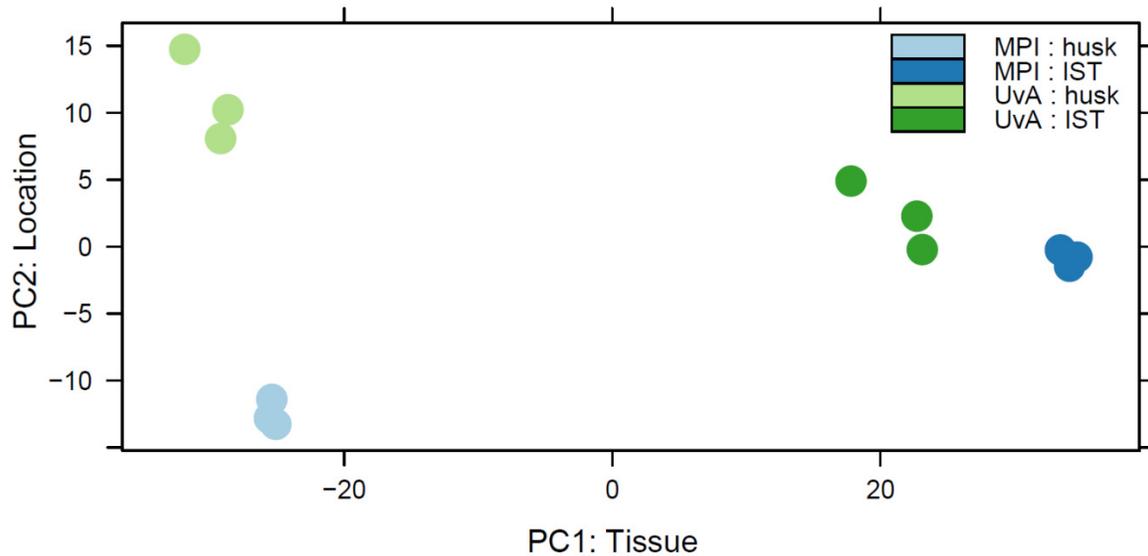


Figure 21: Principal component analysis displaying the weight of each factor (tissue type and location) on the distribution of the variance. Each point represents one RNA-seq library with three biological replicates per condition (one color code). The factor “tissue type” explains most of the variance (PC1: first component). The factor “location” describes a part of the variance for the husk tissue between locations while it does not for the IST (PC2: second component). The three biological replicates are clustered together indicating that the variances represented by the library factor is minimal. These data are derived from the DESeq pipeline before the fitting of the models (see materials and methods).

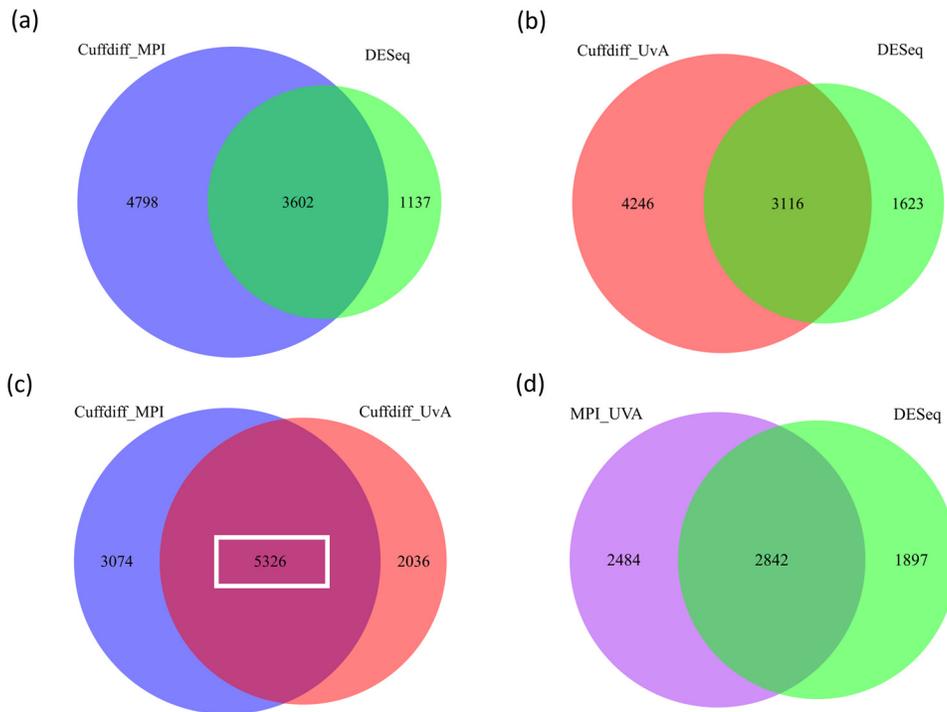


Figure 22: Venn diagrams indicating the overlap of significant DEGs with a fold change greater than two for the three RNA-seq analyses performed: **(a)** overlap between MPI Cuffdiff and DESeq DEGs, **(b)** overlap between UvA Cuffdiff and DESeq DEGs, **(c)** overlap between Cuffdiff MPI and Cuffdiff UvA DEGs, **(d)** overlap between the overlap of Cuffdiff MPI and UvA (MPI_UvA), and DESeq DEGs. The significance of the overlaps were tested against a pool of 15,000 DEGs (average of DEGs in the three datasets) with a Fisher's exact test. All overlaps were significant with odds ratios greater than four. The DEGs overlapping in **c** (white rectangle) were retained for further analyses.

11.5 Data validation at known enhancers

11.5.1 Differential gene expression

To assess whether the combined data of DHSs, H3K9ac enrichment, and gene expression in IST and husk can predict enhancer location and activity, we used as control the known *tb1* upstream enhancer, the single repeat of *b1* present in B73, and the putative enhancer of *ZmRap2.7*, called *Vgt1*. Although *b1* hepta-repeat was shown to act as enhancer in *B-I* plants, no data showed whether the remaining single repeat present in B73 has any enhancer activity. However, a single repeat from the hepta-repeat could drive β -glucuronidase (*GUS*) reporter gene expression when placed in front of the cauliflower mosaic virus 35S minimal promoter in stable transgenic *Z. mays* plants (Belele et al., 2013), indicating that the single repeat present in B73 may have an enhancer activity. Previous RNA-seq data showed that these three genes were differentially expressed in husk and IST in B73: *tb1* and *b1* were more expressed in the developing ear (which contains husk), and *ZmRap2.7* is more expressed in seedling (**Figure 23 a**) (Davidson et al., 2011). Consistently with previous

results, our data suggest that the three genes are significantly differentially expressed between IST (young seedling immature leaves) and husk (which is a part of the ear) (**Figure 23 b**).

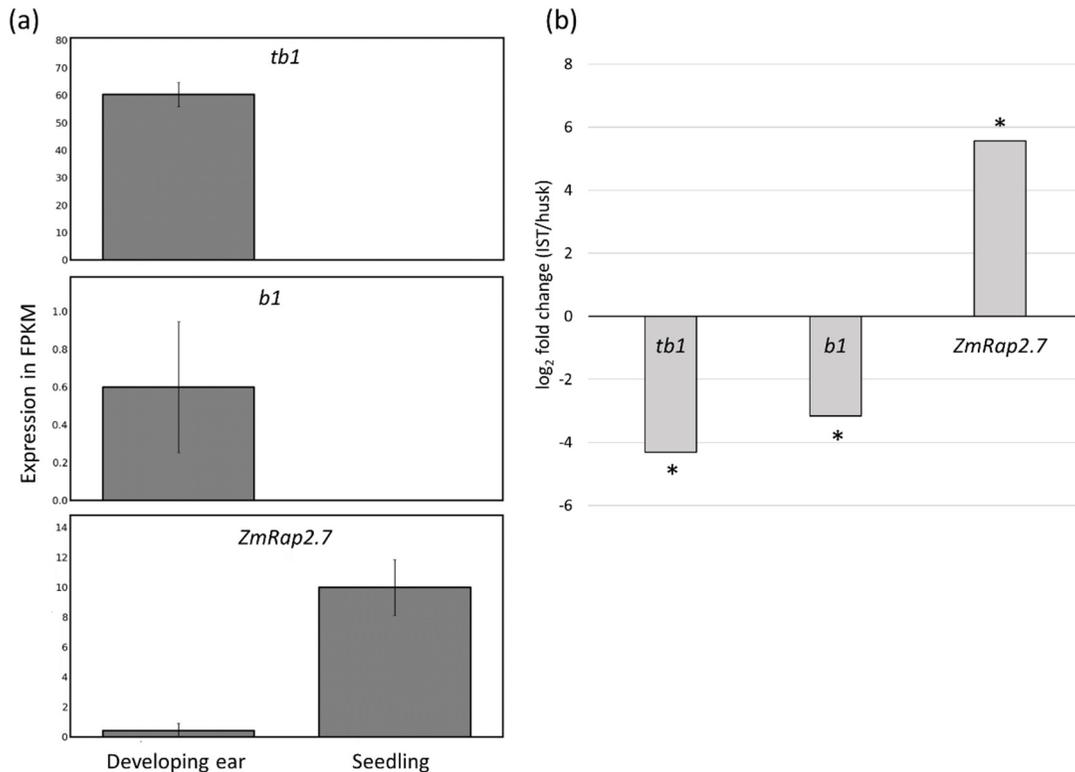


Figure 23: Differential expression of *tb1*, *b1*, and *ZmRap2.7* in two tissues. **(a)** Previously published RNA-seq data from Davidson et al. 2011. The bar chart was generated using qTeller tool (www.qteller.com) with “seedling leaves field” and “developing ear” as selected tissues from “Davidson 2011” datasets. The expression values are given in Fragments Per Kilobase of exon model per million fragments Mapped (FPKM). **(b)** Differential expression of *tb1*, *b1*, and *ZmRap2.7* from our RNA-seq dataset given in \log_2 (fold change IST/Husk). The values were derived from DESeq analysis on the pooled data from MPI and UvA locations after fitting the models. Asterisks indicate a significant \log_2 (fold change) (p-value adjusted < 0.05).

11.5.2 Chromatin accessibility and H3K9ac enrichment profiles

After verifying that *b1*, *tb1*, and *ZmRap2.7* were among significant DEGs in our RNA-seq dataset, we assessed whether DHSs and H3K9ac enrichment peaks were present at the enhancers at the three loci. Indeed, the three enhancer regions contained DHSs and showed an enrichment of H3K9ac in at least one tissue (**Figure 24 a, b, c**).

In the case of *tb1*, the enhancer region is ≈ 10.8 kb long and contains an *Hopscotch* TE able to increase expression of a reporter gene in transient expression assays (Clark et al., 2006; Studer et al., 2011) (**Figure 24 a**). Chromatin accessibility and H3K9ac enrichment cannot be assessed at the *Hopscotch* element due to its repetitive nature (**Figure 24 a**). Nevertheless, the adjacent 5' region to the *Hopscotch* showed an increase in chromatin accessibility and H3K9ac enrichment specifically in husk (**Figure 24 a**). Another region located at the 5' end of the enhancer region showed a high

chromatin accessibility and H3K9ac enrichment that seemed to be stable across the two tissues. It may well be that most changes in accessible chromatin and H3K9ac enrichment in the active tissue husk occurs within the *Hopscotch* element. In addition to the previously defined enhancer region, we found two regions located ≈ 40 kb upstream of *tb1* TSS and ≈ 9 kb downstream of *tb1* 3' end that display an enrichment in H3K9ac in both tissues but more markedly in husk (**Figure 24 a**). These elements may participate in the regulation of *tb1* expression, although they do not show an increase in chromatin accessibility.

In the case of *b1*, DHSs were located at the single repeat ≈ 60 kb upstream of *b1*, but also in regions ≈ 30 kb and ≈ 54 kb upstream of *b1* TSS, called hereafter *A* and *B*, respectively (**Figure 24 b**). In the *B-I* allele, *A* and *B* are located ≈ 8 kb and ≈ 47 kb upstream of the TSS, respectively (see regions e and f for *B* and region l for *A* in **Figure 18 a**). H3K9ac enriched peak at the repeat was found only in husk while H3K9ac enriched peaks are present at *A* in both tissues but do not overlap (**Figure 24 b**). Enrichment in chromatin accessibility is visible at *B* with a husk-specific DHS. The regions *A* and *B* in the *B-I* allele were shown to be involved in the looping formation upon *b1* activation in husk (Louwers et al., 2009).

In the case of *ZmRap2.7*, a DHS was found within the *Vgt1* region in both tissues (**Figure 24 c**). This DHS of ≈ 370 bp encompasses the Conserved Non-coding Sequence 4 (CNS4), previously defined by *Z. mays-sorghum* comparison (Salvi et al., 2007). One H3K9ac enriched peak is located upstream of *Vgt1* in IST but no significant H3K9ac enrichment peak was found within *Vgt1* itself. Also, another H3K9ac enrichment peak was found in husk but the read density indicates that the variation across tissues was not strong and the enrichment rather weak (**Figure 24 c**). The initial described region of *Vgt1* was defined by standard positional cloning as a ≈ 2 kb region (Salvi et al., 2007). Our data narrow down the region to a putative regulatory element within *Vgt1* to a DHS of ≈ 400 bp. Conversely to *tb1* and *b1* enhancers, chromatin accessibility in *Vgt1* seems to be similar in both tissues, indicating that the tissue-specific activity of *Vgt1*, if any, could not be predicted with changes in chromatin accessibility.

In addition to chromatin changes at enhancers or other identified intergenic regions, the transcriptional activity of *tb1* and *ZmRap2.7* was clearly associated with an increase in both chromatin accessibility and H3K9ac (**Figure 24 a, c**). However, no clear changes were observed at the TSS of *b1* (**Figure 24 b**). Our data showed that increase in chromatin accessibility and H3K9ac enrichment mapped at known enhancers. Furthermore, the differential increase in these two chromatin features correlated at a certain extent with the expected activity of the enhancers in the tissue in which their target genes are more expressed. Finally, we identified new putative *cis*-regulatory elements at *tb1* and *ZmRap2.7* loci.

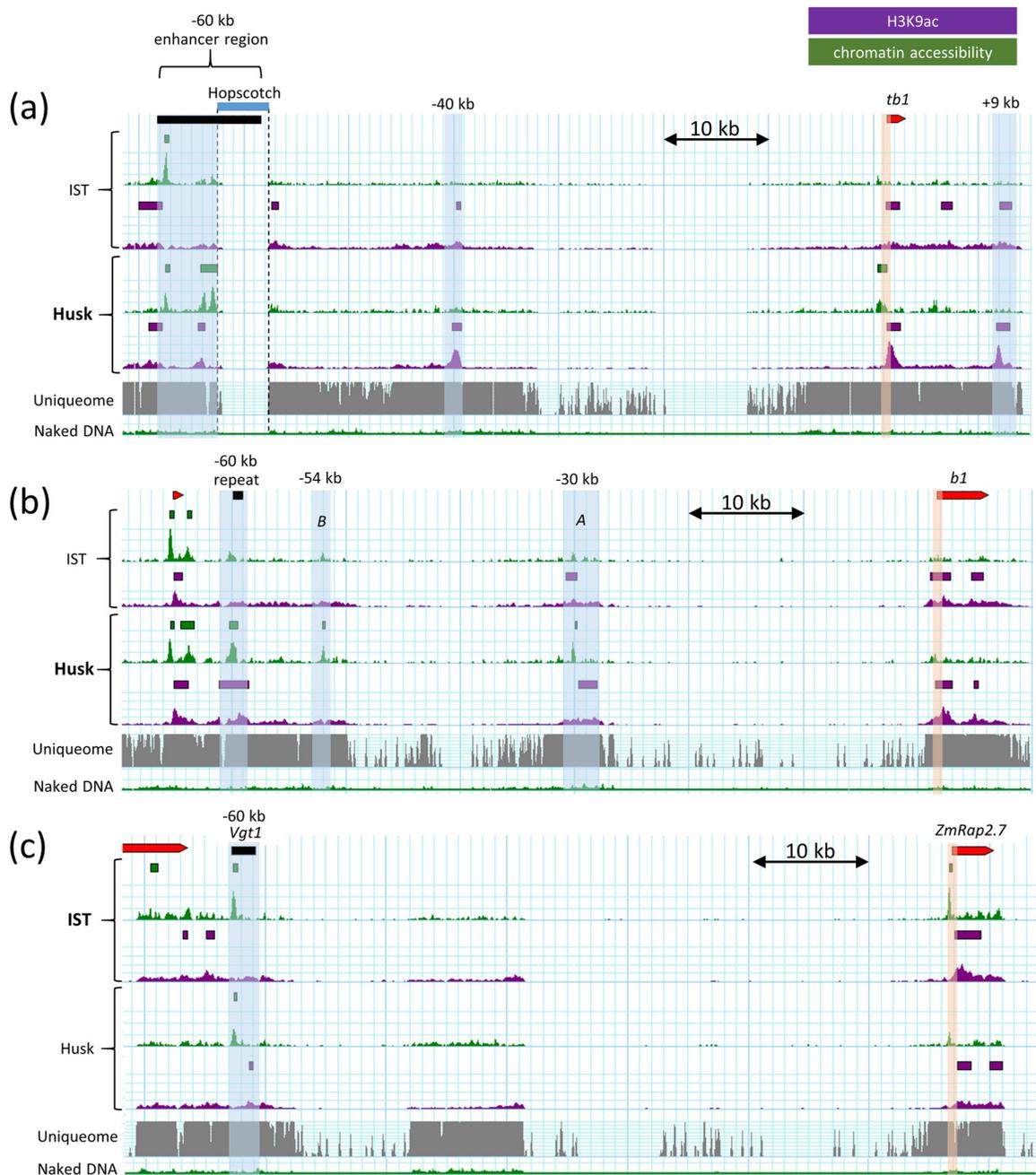


Figure 24: Chromatin accessibility and H3K9ac enrichment profiles at *tb1* (a), *b1* (b), and *ZmRap2.7* (c) loci. Tissue in which each gene is the most expressed is indicated in bold. Filled red and black rectangles indicate protein-coding genes and previously defined enhancer regions, respectively. Green and purple rectangles represent DHSs and H3K9ac islands, respectively. Read density is indicated in green and purple for DNase-seq and ChIP-seq data, respectively (signal per million reads). Uniqueome track (in grey) represents the mappable fraction of the genome. Naked DNA track indicates the background signal derived from digestion of naked DNA by DNase I (signal per million reads). Known and putative enhancers are indicated by blue shadows (distances indicated in - kb from TSS or in + kb from 3' end of the gene), TSS regions are indicated by orange shadows, and putative enhancers are indicated by blue rectangles. The *Hopscotch* TE insertion in the *tb1* enhancer region is indicated by a blue box in a.

11.6 Genome-wide identification of enhancers

11.6.1 Data validation

The cross validation of our data with three known enhancers in plants showed that chromatin accessibility and/or H3K9ac enrichment correlated with enhancer regions and TSSs of the target genes. We overlapped data from DNase-seq and ChIP-seq to identify intergenic regions that are both DHSs and enriched in H3K9ac as putative enhancers. First, we verified whether our three datasets correlated with each other according to prior knowledge, i.e. if chromatin accessibility and H3K9ac enrichment were correlated with gene expression. To do so, we plotted the read signal for both DNase-seq and ChIP-seq data at genes classified by their expression level. The **Figure 25** shows that TSSs, and at a lesser extent, 3' ends of genes were more accessible and enriched in H3K9ac when highly expressed. Although the correlation between chromatin accessibility and gene expression was consistent (**Figure 25 a, b**), H3K9ac enrichment for bins containing the genes the most expressed was not consistent. For instance, genes in bin 4 and 5 showed stronger enrichment of H3K9ac at TSS than genes from bin 6 which are more expressed (**Figure 25 c, d**). This can be due to the fact that the most highly expressed genes (bin 6) showed a strong increase in chromatin accessibility (**Figure 25 a, b**), potentially reducing the H3K9ac signal due to loss of nucleosomes around the TSS. Important standard errors of the means were reported for genes in bins 2, 3, and 4 in the chromatin accessibility profiles (**Figure 25 a, b**) and can be linked to the binning of the genes and their length. We concluded that both chromatin accessibility and H3K9ac enrichment were correlated with gene expression.

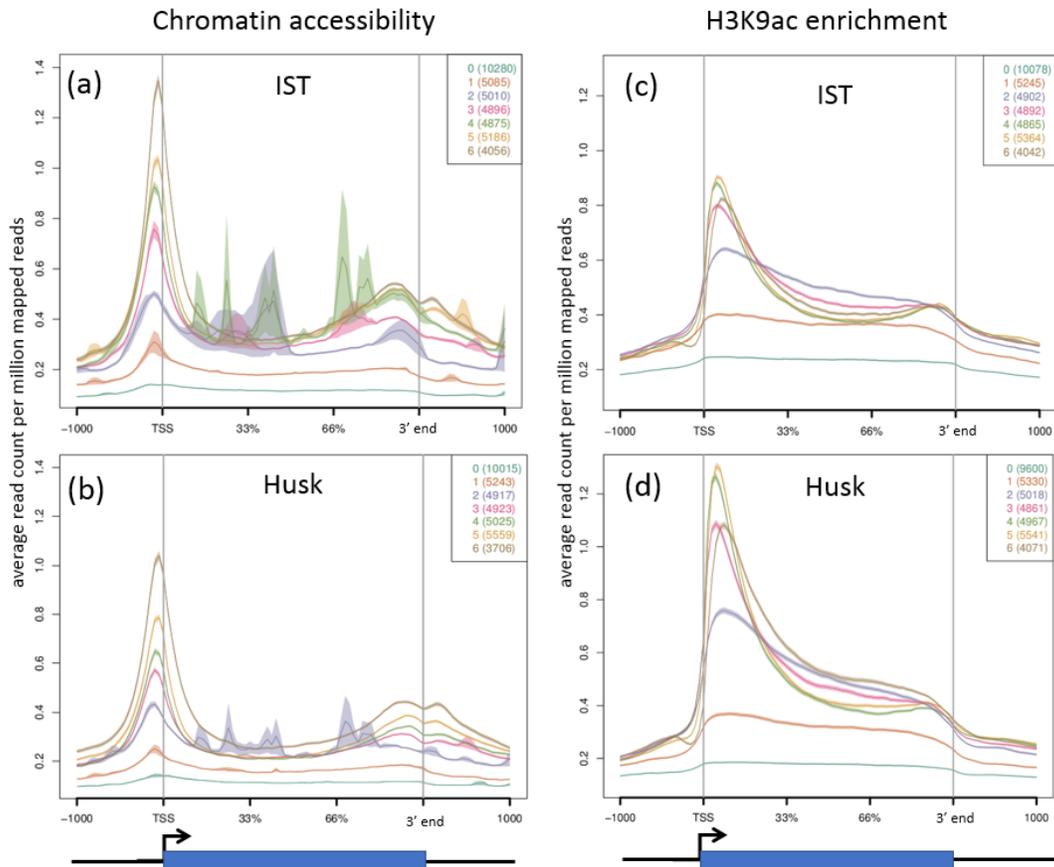


Figure 25: Average signal in reads per million mapped reads (RPM) for chromatin accessibility and H3K9ac enrichment at genes (blue box) and their flanking regions for IST **(a, c)** and husk **(b, d)**. Genes were binned by expression profiles ranging from 0 (not expressed) to 6 (highly expressed). The number of genes in each bin is indicated between parentheses. Shaded areas represent SEM. TSS are represented by black arrows (plots from Rurika Oka).

We then looked at the relationship between H3K9ac enrichment and chromatin accessibility at intergenic DHSs (excluding promoter regions) and found that H3K9ac enrichment was biased to one side of the DHSs (**Figure 26 a, b**) in the same manner as for TSSs (**Figure 26 c, d**). Although unidirectionality of H3K9ac enrichment downstream of TSS of genes was previously reported (Karmodiya et al., 2012), it was not clearly shown at enhancers.

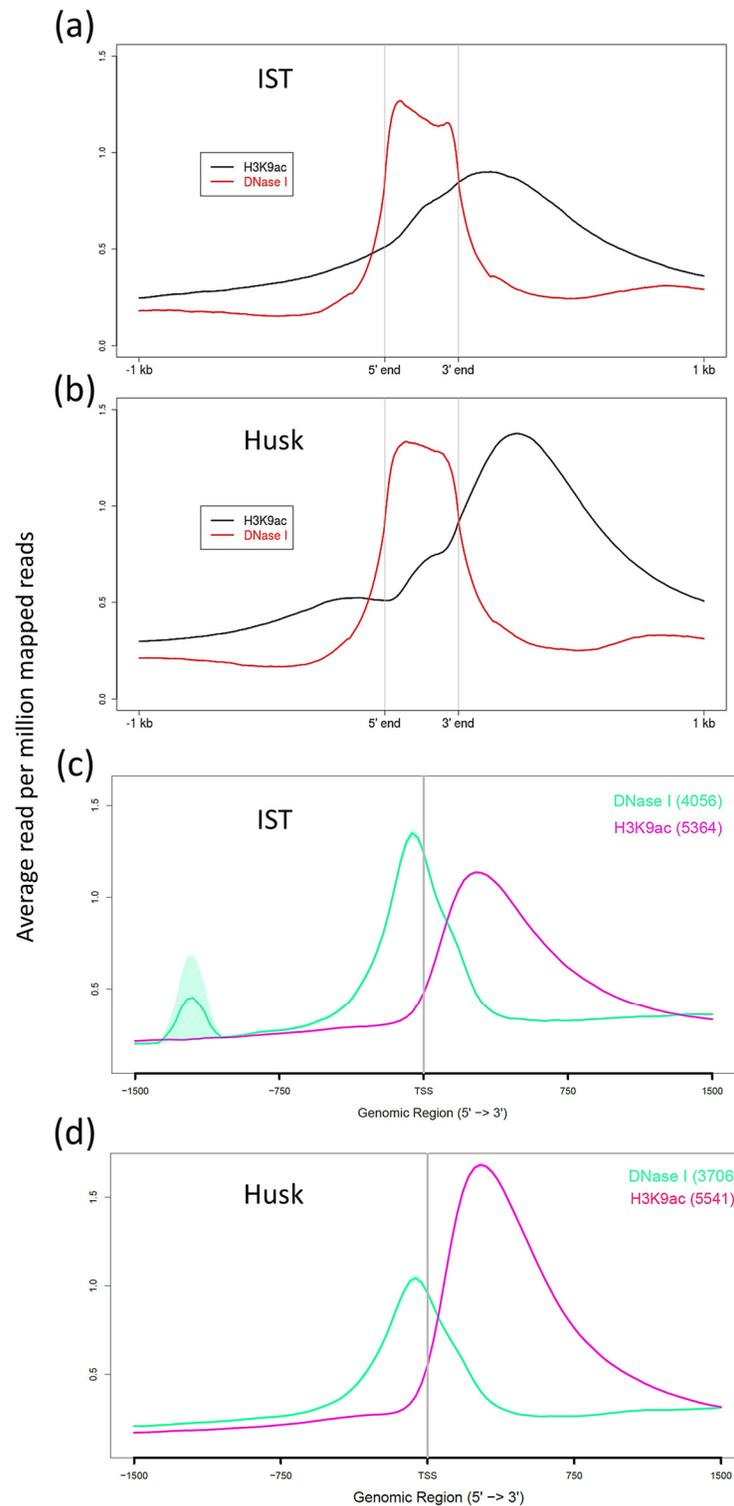


Figure 26: Average signal in RPM for chromatin accessibility (DNase I) and H3K9ac enrichment at intergenic DHSs (excluding promoter regions) and their 1 kb flanking regions for IST **(a)** and husk **(b)** and at TSS regions for IST **(c)** and husk **(d)**. The 3' end of DHSs **(a, b)** were defined as the end with the highest H3K9ac enrichment value at 300 bp away from DHSs end. The figures **c** and **d** are showing the same information than in **Figure 25** but only the genes of bin 6 for DNase I and bin 5 for H3K9ac are shown to highlight the switch between chromatin accessibility and H3K9ac enrichment at DHSs (plots from Rurika Oka).

11.6.2 Enhancer prediction

We defined putative enhancers in intergenic regions containing both DHSs and H3K9ac enrichment peaks (**Figure 27 a**). Although this filtering may discard many enhancers that contain non-overlapping features, we assumed to increase the likelihood that selected candidates may have *cis*-regulatory functions. We identified 1911 and 1990 candidates in IST and husk, respectively. Among these, 739 were shared among tissues, making a total of 3162 candidates. We calculated for each candidate the largest difference in chromatin accessibility and H3K9ac enrichment in IST and husk independently, sorted in descending order and ranked the values for each feature, summed the rank values of both features, sorted in ascending order the sum values and rank them so that every candidate in each tissue has a unique rank (**Figure 27 b**). The candidates at the top of the list are showing the strongest differential level in chromatin accessibility and H3K9ac enrichment (**Figure 27 b**). From this list, we could find both the enhancers of *b1* and *tb1*, ranked as candidates Husk #93 and Husk #188, respectively. However, we could not find *Vgt1*, which showed no overlap between H3K9ac enrichment peaks and DHSs.

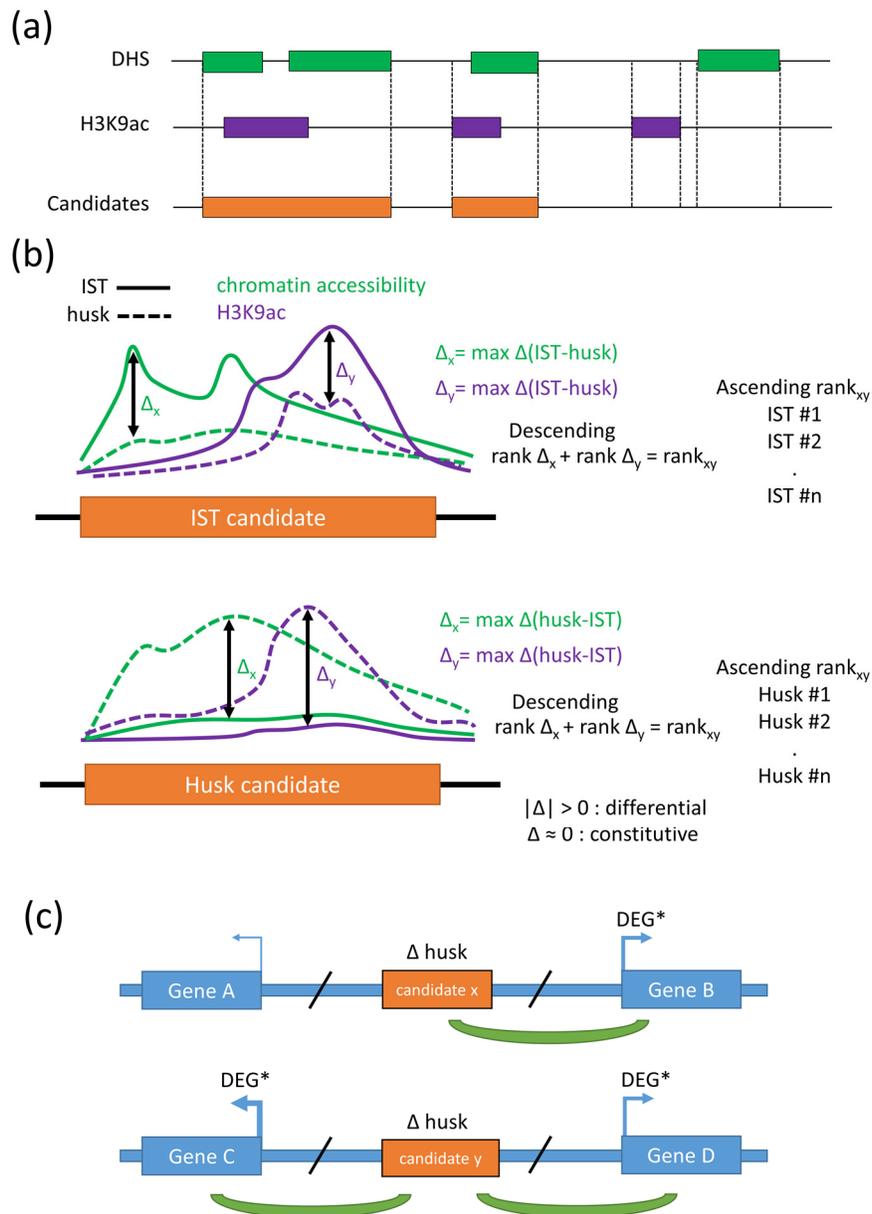


Figure 27: (a) Putative enhancers were defined as region containing overlapping DHSs and H3K9ac enrichment peaks. We excluded candidates overlapping with genic regions. (b) Maximum differential levels of chromatin accessibility (Δ_x) and H3K9ac enrichment (Δ_y) were measured between IST and husk for candidates identified in IST husk separately. The values of Δ_x and Δ_y were sorted by descending order and ranked for each candidate in each tissue. The ranks were then summed for each candidate and the rank_{xy} values were sorted in ascending order. For instance, husk candidates with high rank showed high chromatin accessibility and high H3K9ac enrichment in husk, middle ranked husk candidates showed small or no differences with IST, and low ranked husk candidates showed enrichment of chromatin accessibility and H3K9ac enrichment in IST. The first 100 candidates were selected for further analysis. (c) For each putative enhancer from the top 100 list, flanking genes were identified and linked to the candidate if they were significant DEGs with a fold change greater than two (labelled DEG*). If both flanking genes were DEGs*, they were both linked to the putative enhancer.

To associate tissue-specific putative enhancers with their target genes, we assessed the differential expression of genes located directly downstream and upstream of the top 100 ranked candidates of each tissue (**Supplementary Table 6** and **Supplementary Table 7**). Although enhancers can act across several genes, we assumed that adjacent genes of the putative enhancers are most likely to be regulated (Mendenhall et al., 2013; Ghavi-Helm et al., 2014; Kvon et al., 2014). We selected only significant DEGs with a fold change greater than two (**Table 4**). For instance, *tb1*, *bl*, and *ZmRap2.7* expression levels were found to vary from a fold change of 16 to 64 (**Figure 23 b**), indicating that putative tissue-specific enhancers may have an increasing effect on gene expression over the twofold change. However, it is not excluded that some tissue-specific enhancers may have a more subtle effect on the expression of their target gene. We could obtain a final list of 34 and 31 genes flanking a subset of the top 100 candidates in IST and husk, respectively (**Table 4** and **Supplementary Figure 35**). This indicated firstly that genes with a high fold change are more likely to be close to a putative enhancer as the odds ratio for both subsets of genes were greater than two and the p-values associated to a Fisher's exact test were lesser than 5%. In term of expression pattern across tissues, we found that 21/34 genes for IST candidates and 21/31 genes for husk candidates were more expressed in the tissue of their putative enhancer (**Table 5** and **Supplementary Figure 35**). This seems to indicate a bias toward a higher expression of the gene near their "active" putative enhancers but it could not be tested statistically. Within this subset, we considered genes being more expressed in the tissue for which the putative enhancer was ranked. This selection yielded a list of 21-22 enhancers linked to 21 genes in for each tissue (**Table 5**).

Table 4: From all significant DEGs from MPI and UvA calculated by Cuffdiff, only the overlapping DEGs between locations were selected. DEGs with a fold change greater than two were selected from the overlap between the two locations. The overlap of the selection with the flanking genes of the top 100 putative enhancers in each tissue was performed. From these two lists, DEGs more expressed in the tissue in which the flanking putative enhancer was highly ranked (top 100 list) were kept, representing 21 genes for each tissue.

Location	MPI	UvA
DEG q-value < 0.05	16967	14614
DEG fold change > 2	8435	7391
\cap between locations	5326	
Top 100 IST putative enhancer flanking genes	34	
Top 100 husk putative enhancer flanking genes	31	
More expressed in IST for IST putative enhancers	22	21
More expressed in husk for husk putative enhancers	21	21

Table 5: List of the candidate enhancers and their putative target genes. A total of 22 and 21 candidate enhancers from the top 100 lists were associated with 21 genes in IST and husk, respectively. Rows in grey indicate multiple putative candidate enhancers associate to the same gene. Rows in orange indicate a single enhancer linked to two genes. The candidate husk #93 (H93), in bold, corresponds to the enhancer of *bl* and is associated with the gene *bl*. Gene names are given according to the Gramene annotation system.

IST		Husk	
Candidate	Gene	Candidate	Gene
V1	GRMZM2G089448	H4	GRMZM2G029153
V6	GRMZM2G085945	H5	GRMZM2G312661
V10	GRMZM2G034727	H13	GRMZM2G089698
V17	GRMZM2G310465	H16	GRMZM2G022538
V20	GRMZM2G121151	H18	GRMZM2G043498
V32	GRMZM2G024948	H18	GRMZM2G043396
V33	GRMZM2G035278	H20	GRMZM2G022538
V34	GRMZM5G879749	H22	GRMZM2G339122
V35	GRMZM2G017319	H24	GRMZM5G860241
V37	GRMZM2G176355	H27	GRMZM2G093826
V39	GRMZM2G159237	H37	GRMZM2G055575
V47	GRMZM5G884137	H38	GRMZM2G159500
V50	GRMZM2G442000	H53	GRMZM2G039399
V56	GRMZM2G033074	H57	GRMZM2G065394
V59	GRMZM2G107377	H58	GRMZM5G806638
V60	GRMZM2G383735	H66	GRMZM2G394321
V63	GRMZM2G085945	H69	GRMZM2G379005
V77	GRMZM2G160730	H75	GRMZM2G438202
V87	GRMZM2G040706	H76	GRMZM2G006964
V94	GRMZM2G115635	H84	GRMZM2G405090
V95	GRMZM2G142751	H93	GRMZM2G172795 (<i>bl</i>)
V96	GRMZM2G116185	H99	GRMZM2G168917

12 Discussion

We described the first genome-wide profile of chromatin accessibility for IST and husk in *Z. mays* using DNase-seq. In addition, we found that H3K9ac was associated with the active enhancer of *b1*, and therefore generated a genome-wide profile of H3K9ac enrichment using ChIP-seq. We also performed transcriptome analyses in the two tissues using RNA-seq for two main purposes: (1) to assess the correlation between chromatin accessibility, H3K9ac enrichment, and gene expression; and (2) to link DEGs to putative tissue-specific enhancers. Chromatin accessibility and H3K9ac enrichment were both associated with gene expression and could predict the location, and partially, the activity of the previously known or putative enhancers of *b1*, *tb1*, and *ZmRap2.7*. We then performed a genome-wide prediction of putative enhancers based on the overlap between DHSs and H3K9ac enrichment peaks. We defined a list of ≈ 2000 putative enhancers for each tissue examined. We decided to focus on tissue-specific enhancers to be able to link DEGs to flanking putative enhancers. A list of 21 putative enhancers for each tissue was generated based on their differential (1) chromatin accessibility, (2) H3K9ac enrichment, and (3) expression of the flanking genes.

Several caveats should be considered in our analysis. We could not define chromatin accessibility and H3K9ac enrichment in the repetitive fraction of the *Z. mays* genome since most reads could not map uniquely there. Therefore, putative enhancers located in repetitive elements could not be found. For instance, no chromatin features could be defined for the *Hopscotch* insertion in *tb1* enhancer but we could nevertheless retrieve the 5' flanking region of the *Hopscotch*, which displayed high chromatin accessibility and H3K9ac enrichment. In addition, the decision of selecting only candidates with overlapping DHSs and H3K9ac enrichment peaks led to the loss of putative enhancers. For instance, *Vgt1*, the putative enhancer of *ZmRap2.7*, contained a clear DHS but did not present an enrichment in H3K9ac, resulting in its removal from our list. However, the confirmed enhancer of *b1* and a part of the enhancer of *tb1* could be retrieved, supporting our method. We applied a stringent selection to narrow down our list of putative enhancers based on their differential chromatin accessibility and H3K9ac enrichment between tissues. This allowed us to link a subset of putative enhancers with flanking DEGs. The 21-22 candidates obtained for each tissue are currently cloned to be fused to the cauliflower mosaic virus minimal promoter 35S and a reporter gene (either *GUS* or *luciferase*) to be tested in transient reporter assays using particle bombardment. The stringency of selection applied can be lowered to predict more putative enhancers, depending on the downstream analyses to be performed. For instance, constitutive enhancers in the two tissues can be selected by considering stable chromatin accessibility and H3K9ac enrichment, and be linked to flanking genes stably expressed. Considering that DHSs and H3K9ac enrichment peaks do not always overlap, filtering can be applied using either of these feature to obtain more putative enhancers. Also, it would be interesting to see whether our putative

enhancers are enriched in TFBSs. Recently, about 1340 novel TFBSs were predicted in *Z. mays* and can be used as database (Yu et al., 2015).

In plants, enhancer prediction based on genome-wide profiling of chromatin features was only performed *A. thaliana* (Zhu et al., 2015). We found $\approx 20,000$ to $\approx 30,000$ DHSs for each tissue in *Z. mays*, while $\approx 40,000$ DHSs were found in leaf and flower tissues in *A. thaliana* (Zhang et al., 2012b). However, the count of DHSs depends strongly on the algorithm and the parameters used during the peak calling (Koohy et al., 2014). For a more meaningful comparison, we looked at DNase-seq data generated in rice (Zhang et al., 2012a), which has a genome closer to *Z. mays*, albeit about seven times smaller (Bruggmann et al., 2006). Genomic distribution of DHSs in *Z. mays* were rather comparable with what was obtained in rice (Zhang et al., 2012a) (**Figure 28**). DHSs are strongly enriched in genic and promoter regions, though they represent only $\approx 8\%$ and $\approx 30\%$ of *Z. mays* and rice genome, respectively³. We can notice that less DHSs seem to map in introns compared to rice, although the average size of introns in *Z. mays* is longer with ≈ 600 bp compared to ≈ 400 bp in rice (Wendel et al., 2002; Haberer et al., 2005) (**Figure 28**). The discrepancy could be linked either to a real biological difference, what we doubt, or more likely to technical differences linked to the parameters used to bin DHSs in the different genomic features. Introns aside, we see that despite the different methods used for DHSs calling in *Z. mays*, we retrieve similar trends than observed in rice.

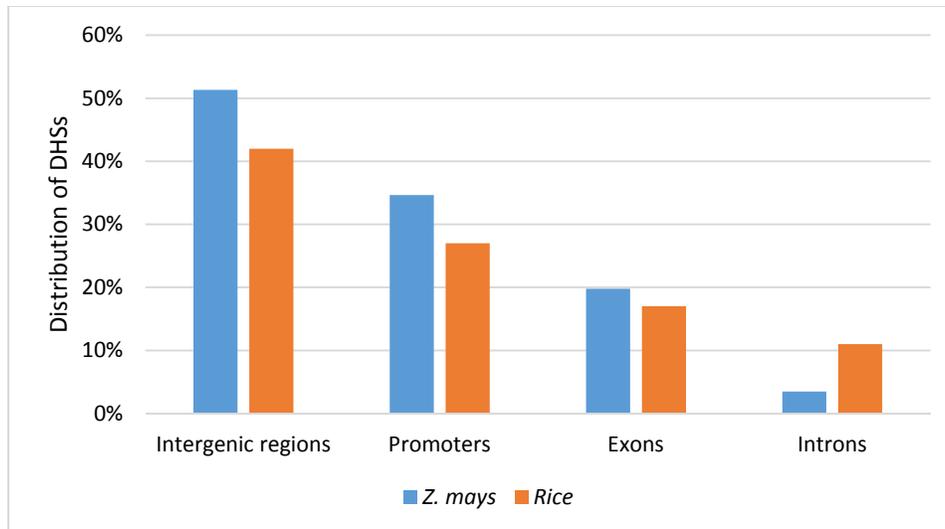


Figure 28: Distribution of DHSs in *Z. mays* and rice. DHSs in promoters include all DHSs between 1 kb upstream and 1kb downstream of TSSs of genes in *Z. mays*, and 1 kb upstream of rice genes. DHSs were collected from IST and husk in *Z. mays* (this thesis) and rice seedlings (Zhang et al., 2012a). Figure implemented from (Zhang et al., 2014).

³ Values estimated from protein-coding gene annotations and golden path lengths available for the two species on the Gramene database (<http://ensembl.gramene.org/> consulted on March 2016).

From the DHSs identified in *A. thaliana*, $\approx 10,000$ intergenic regions were predicted as putative enhancers, based only on chromatin accessibility (Zhu et al., 2015). We found $\approx 15,000$ and $\approx 12,000$ intergenic DHSs in IST and husk, respectively. We then overlapped these DHSs with H3K9ac enrichment peaks to obtain a final set of ≈ 2000 putative enhancers for each tissue. We therefore strongly reduced the potential number of putative enhancers by taking into account H3K9ac. In *A. thaliana*, 10 out of 14 predicted enhancers tested were validated by the reporter assay in stable transgenic lines (Zhu et al., 2015). We cannot yet comment on the predictive success of our method but it will be interesting to assess whether the consideration of H3K9ac enrichment to determine putative enhancers is improving the prediction. Unfortunately, the generation of stable transgenic lines in *Z. mays* is more time consuming and costly than in *A. thaliana*. As quicker alternative to begin with, we will test our putative enhancers in transient reporter assays using particles bombardment on IST and husk. However, stable transgenic lines allow to assess enhancer activity in different tissues and at different stages, whereas transient expression relies mainly on few types of tissue that can be easily transformed (Kirienko et al., 2012). Ultimately, validated enhancers in transient reporter assays should be integrated into stable transgenic lines to assess whether the expression pattern generated by the enhancer is restricted to a specific tissue or is also active elsewhere.

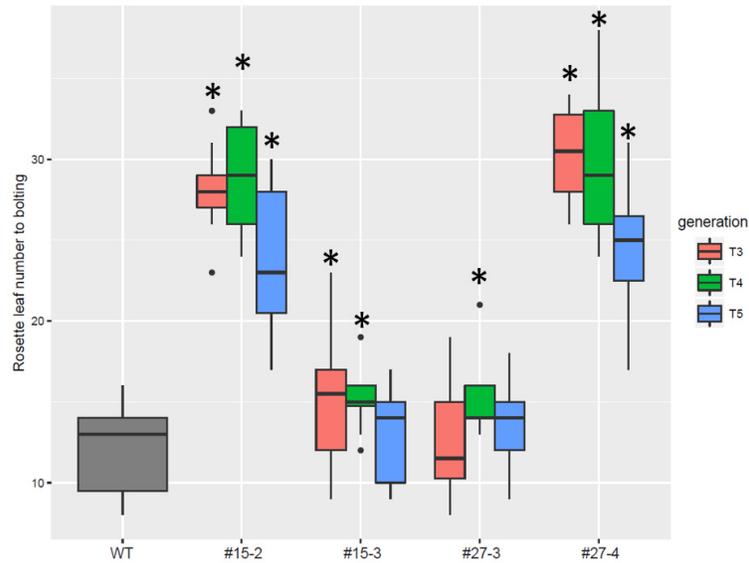
We also found in our data an interesting pattern that could highlight a difference between enhancers in plants and animals. We observed that H3K9ac enrichment at intergenic DHSs was, as for TSSs, mostly unidirectional. In mammals, enhancers were shown to be transcribed bidirectionally and this bidirectionality was also reflected by the symmetric distribution of the histone marks H3K4me1 and H3K27ac (Kim et al., 2010; Andersson et al., 2014). Also, the unidirectionality of H3K9ac at TSSs was shown to be related to transcriptional directionality of protein-coding genes (Bornelöv et al., 2015). Consequently, we would have expected bidirectional enrichment of H3K9ac at intergenic DHSs under the assumption that most of the DHSs are indeed located at enhancers that are transcribed. We still need to assess the level of transcription occurring at intergenic DHSs from our RNA-seq data. To support the idea that our observation was not an artefact, H3K27ac enrichment at intergenic DHSs also displayed unidirectionality in *A. thaliana*, although less clearly than what we saw with H3K9ac (Zhu et al., 2015). This unidirectionality may be partly explained by the presence of TSSs of unannotated genes but we suggest rather that plant-specific features are responsible of this pattern. We are currently integrating published DNA methylation data (Regulski et al., 2013) to investigate the relationship between DNA methylation and H3K9ac distribution at intergenic DHSs. We can already say that DNA methylation shows a unidirectional pattern opposite to H3K9ac at intergenic DHSs but further analyses are still required.

13 Conclusion

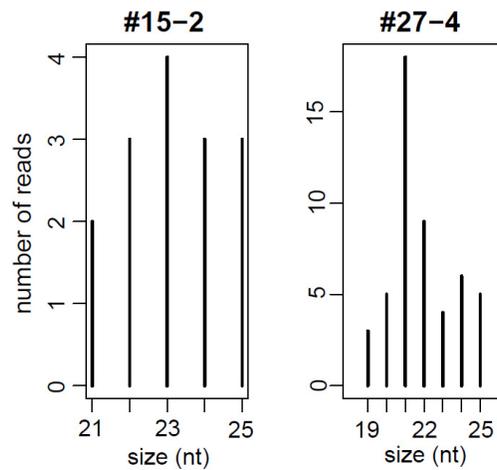
The two projects carried out in this PhD thesis had as main objective to improve our understanding of enhancers in plants. Although few enhancers are described in plants, it is likely that they play an important role in the fine-tuning of gene expression. The two approaches used in the projects were rather different, with one approach based on induced-chromatin modifications at a single locus to affect the function of known and putative enhancers, whereas the other was based on genome-wide chromatin feature profiling. Both approaches are complementary since predictions from genome-wide work-flows require proper validation methods to ascertain the role of an enhancer in a given regulatory network.

In the first part of this thesis, we found that IRs were useful to alter the function of an enhancer in its endogenous location and assess the effect on *FT*. We furthermore could test new regulatory elements that also seem to be involved in *FT* regulation. In the second part of the thesis, we identified putative enhancers based on predictive chromatin features in *Z. mays*. We could determine a set of putative enhancers that will be validated in the future. In the junction between the two projects, transgenic lines of *Z. mays* containing an IR targeting *Vgt1* were generated and are currently being phenotyped to see whether an early flowering can be detected. This experiment would confirm whether *Vgt1* is indeed an enhancer, and constitute an additional proof of concept of the usefulness of IRs in enhancer characterization. We expect that additional knowledge on enhancers will improve our understanding of complex gene regulatory networks in plants and maybe provide new markers in breeding programs.

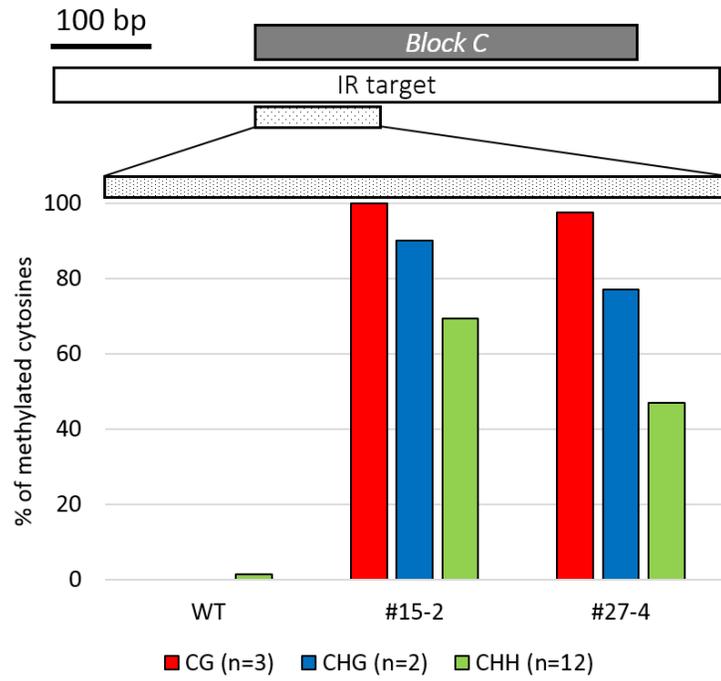
14 Supplementary data



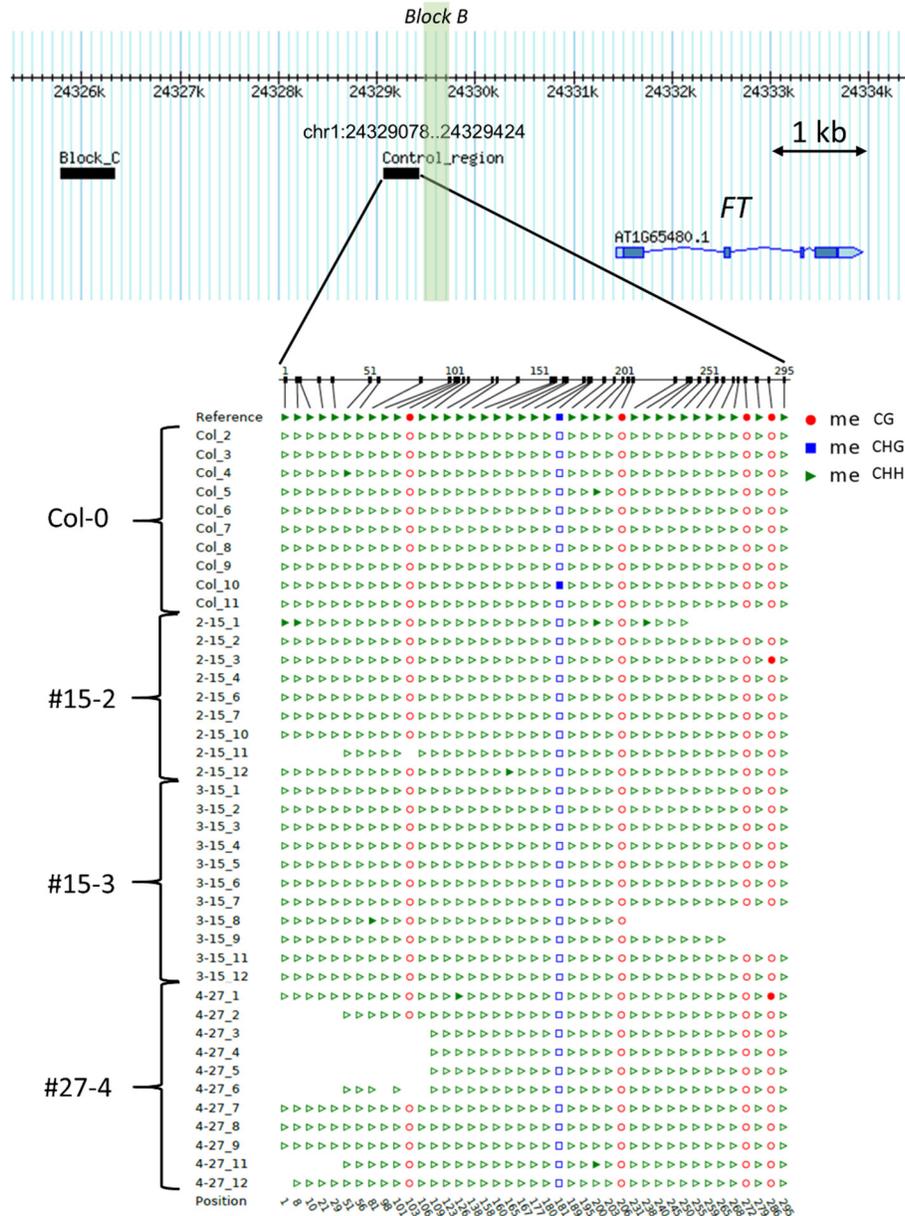
Supplementary Figure 29: Flowering time in LD conditions for three generations. Four independent experiments with two carried out in greenhouse and two carried out in growth chamber conditions are pooled here, explaining the important variation since plants in greenhouse always flowered earlier than in growth chamber. At least 12 plants were assessed by modality. For statistical ANOVA followed by post-hoc Dunnett's tests was performed for each generation independently (one-sided, WT as control group). Asterisks indicate a significant difference with WT (α -risk=5%).



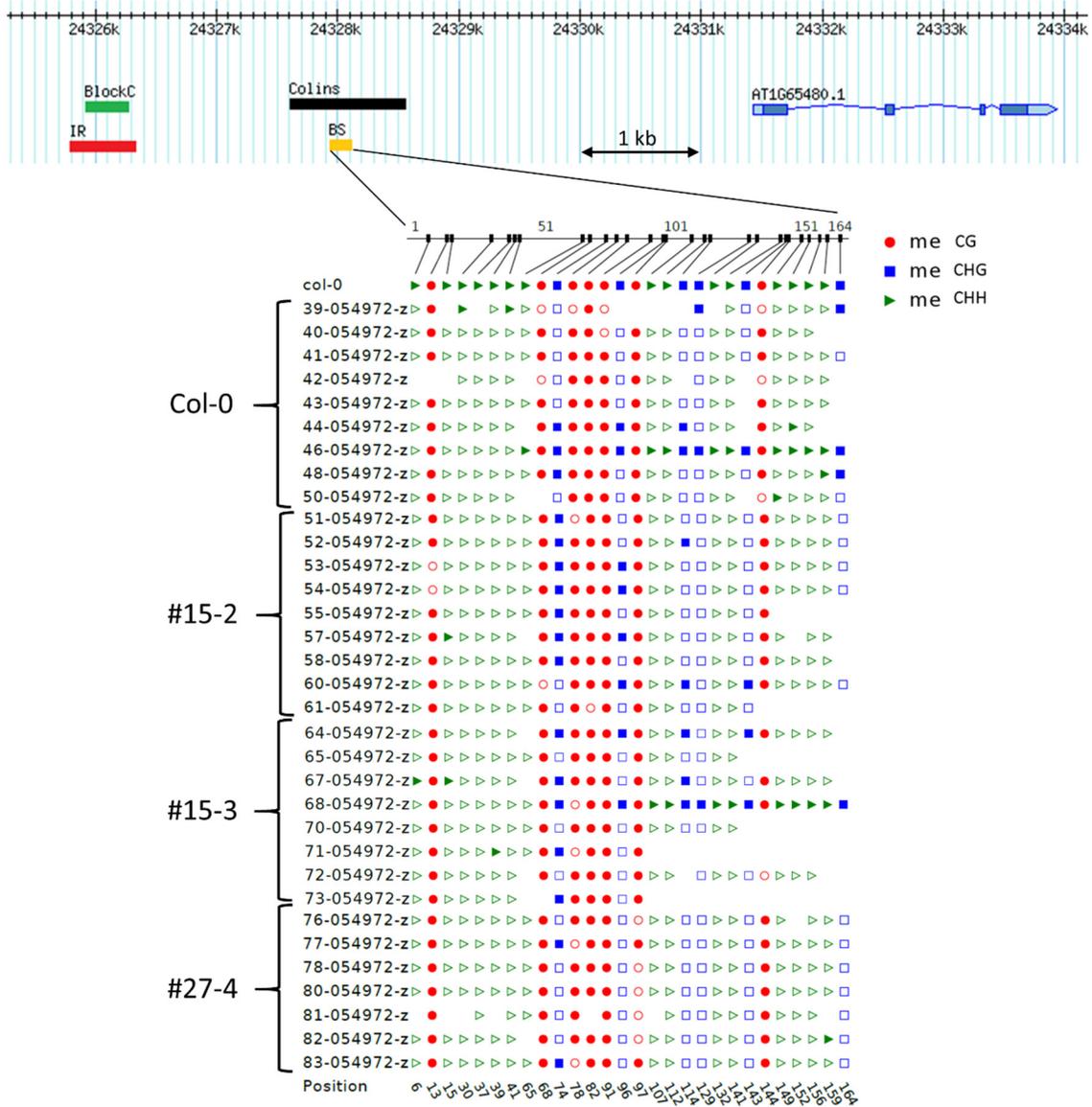
Supplementary Figure 30: Distribution of read sizes for both transgenic lines #15-2 and #27-4 at the region of the *WRKY* intron used as separator in the IR construct. Few reads were found compared to the IR target region (at *Block C*).



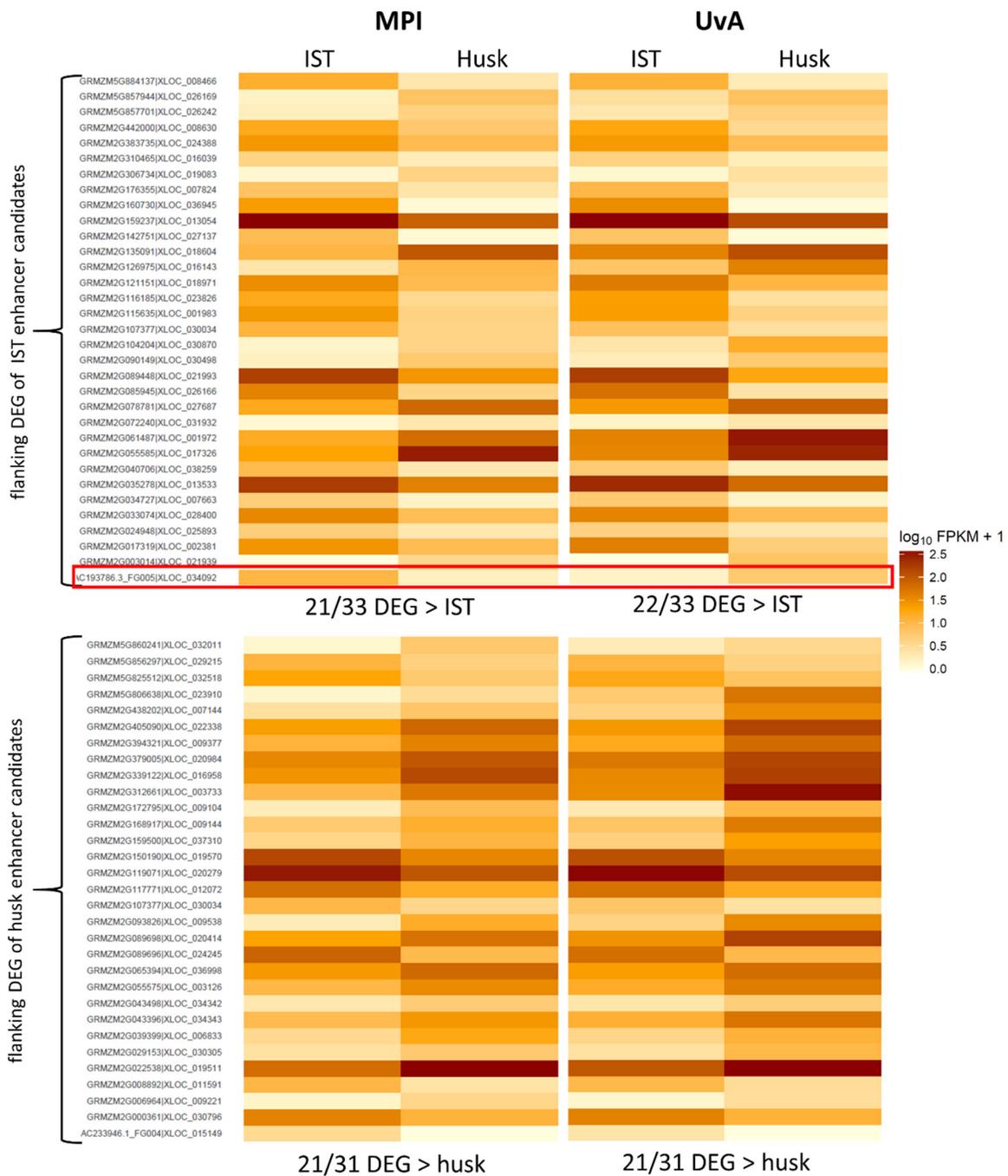
Supplementary Figure 31: DNA methylation level at the 5' end of *Block C* for the two transgenic lines #15-2 and #27-4 at generation T3 (10 to 13 clones were examined). Methylation contexts and number of cytosines in each context are indicated in the legend.



Supplementary Figure 32: DNA methylation status at the control region (black box) for Col-0, transgenic lines #15-2 and #27-4, and non-transgenic sibling #15-3 at generation T3. The vertical green box represents the location of *Block B*. Each line represents a different clone and triangles, circles and squares indicate cytosines in CHH, CG, and CHG contexts, respectively. The cytosine is methylated when the shape is filled. Few cytosines were methylated at this region, indicating that DNA methylation from *Block C* did not spread downstream. Missing shapes indicate sequencing errors.



Supplementary Figure 34: DNA methylation status from a part of Col-0 insertion (insertion and sequenced region indicated by black and yellow rectangles, respectively) for the transgenic lines #15-2 and #27-4, and the non-transgenic sibling #15-3 at generation T3. *Block C* and IR indicated by a green and red rectangles, respectively. Each line represents a different clone and triangles, circles and squares indicate cytosines in CHH, CG, and CHG contexts, respectively. The cytosine is methylated when the shape is filled. The region was methylated in almost all CG context, at a lesser extent in CHG context, and surprisingly not in CHH context. Missing shapes indicate sequencing errors.



Supplementary Figure 35: Heatmap of the gene expression (in \log_{10} FPKM+1) of the 33 and 31 flanking DEGs for IST and husk putative enhancers, respectively. The red rectangle indicates the gene with an inverse expression pattern between the two locations. This gene was discarded from the analysis. Generated with CummeRbund R package.

Supplementary Table 6: List of the top 100 putative enhancers in IST based on differential chromatin accessibility and enrichment in H3K9ac. Genes upstream and downstream are indicated (Gramene annotation).

name	chr	start	end	upstream gene	downstream gene
V1	5	4922525	4924615	GRMZM2G089448	GRMZM2G089425
V2	1	82642192	82643589	GRMZM2G162251	GRMZM2G039880
V3	7	127327511	127329681	GRMZM2G072240	GRMZM2G072669
V4	7	159268080	159269659	GRMZM2G107562	GRMZM5G816386
V5	8	157425582	157426751	GRMZM2G356714	AC193786.3_FG005
V6	5	211602880	211605644	GRMZM2G085945	GRMZM5G857944
V7	7	133211119	133212598	GRMZM2G307588	GRMZM2G419739
V8	7	41012355	41014418	GRMZM2G148056	GRMZM2G154178
V9	8	20677991	20680150	GRMZM2G173124	AC195174.2_FG007
V10	10	39779508	39780630	GRMZM2G088487	GRMZM2G034727
V11	2	7852874	7853997	GRMZM2G180870	GRMZM2G169013
V12	1	223685945	223689425	GRMZM6G332242	GRMZM2G032222
V13	1	273296296	273300805	GRMZM2G396114	GRMZM2G481194
V14	2	17843714	17846216	GRMZM2G129261	GRMZM2G117060
V15	9	122631026	122632486	GRMZM2G152908	GRMZM2G152105
V16	9	93723994	93724933	GRMZM5G833760	GRMZM2G055243
V17	3	47028949	47029834	GRMZM2G310465	GRMZM2G011101
V18	7	156711267	156712797	GRMZM2G050933	GRMZM2G129973
V19	7	108074229	108074841	GRMZM2G043240	GRMZM2G700188
V20	4	177350986	177356420	GRMZM2G115013	GRMZM2G121151
V21	3	10465871	10467633	GRMZM2G036340	GRMZM2G348238
V22	10	144532047	144533354	GRMZM2G027056	GRMZM2G117531
V23	10	63604485	63605427	GRMZM2G009353	GRMZM2G368902
V24	4	142414013	142414398	GRMZM2G459645	GRMZM2G135091
V25	3	47847943	47850568	GRMZM2G133802	GRMZM2G090374
V26	6	91472299	91474944	AC205886.3_FG001	GRMZM2G089640
V27	5	206652082	206655648	GRMZM2G053396	GRMZM2G140635
V28	7	13061875	13062975	GRMZM6G285883	GRMZM2G144653
V29	7	152037588	152040162	GRMZM2G105750	GRMZM2G134545
V30	4	185487168	185488995	GRMZM2G131516	GRMZM2G306734
V31	7	160653094	160653640	GRMZM2G058197	GRMZM2G104204
V32	5	196808624	196809505	GRMZM2G024948	GRMZM2G010673
V33	3	4522067	4523336	AC155622.2_FG004	GRMZM2G035278
V34	5	213583593	213586022	GRMZM5G879749	GRMZM5G879749
V35	1	255382191	255383491	GRMZM2G017047	GRMZM2G017319
V36	2	200416085	200416636	AC190788.2_FG004	GRMZM2G133421
V37	10	74659387	74660538	GRMZM2G176355	GRMZM2G438602
V38	2	175034103	175034948	GRMZM5G827257	AC203957.3_FG004
V39	2	212151158	212152257	GRMZM2G148194	GRMZM2G159237
V40	2	180825759	180826504	GRMZM2G176141	GRMZM2G036547
V41	7	4844851	4846529	GRMZM2G167892	GRMZM2G102968

V42	3	211612266	211613326	GRMZM2G074821	GRMZM2G055585
V43	5	186668785	186670220	GRMZM2G540538	GRMZM2G074604
V44	5	185699151	185700994	GRMZM2G028830	GRMZM2G332843
V45	2	109559589	109560490	GRMZM2G158043	GRMZM2G140299
V46	9	96413813	96416705	GRMZM2G040382	GRMZM2G139973
V47	10	137593713	137594458	GRMZM5G884137	AC209206.3_FG004
V48	8	165101074	165104063	GRMZM2G005236	GRMZM2G005374
V49	2	188273862	188275106	GRMZM2G389379	GRMZM2G009344
V50	10	146374467	146375756	GRMZM2G008730	GRMZM2G442000
V51	2	183525131	183526173	GRMZM2G019901	GRMZM2G115388
V52	8	159034407	159035945	GRMZM2G037585	AC233953.1_FG004
V53	2	41492082	41493469	GRMZM5G869350	GRMZM2G110279
V54	3	200667995	200668787	GRMZM5G800586	GRMZM2G023285
V55	2	185050408	185051237	GRMZM2G001660	GRMZM5G820822
V56	6	78349907	78350996	GRMZM2G033074	GRMZM2G113640
V57	3	229218311	229219400	GRMZM2G001415	GRMZM2G001457
V58	3	121136448	121137408	GRMZM2G152421	GRMZM2G154317
V59	7	50030692	50032026	GRMZM2G107377	GRMZM2G061996
V60	5	14530537	14532020	GRMZM2G383735	GRMZM2G176748
V61	3	65191691	65192664	GRMZM2G126975	GRMZM2G037209
V62	4	41872479	41873216	GRMZM2G124530	GRMZM2G009214
V63	5	211605873	211607729	GRMZM2G085945	GRMZM5G857944
V64	1	217692872	217694042	GRMZM2G061487	GRMZM2G347280
V65	8	76033967	76036202	GRMZM5G822426	GRMZM2G015692
V66	5	85693905	85694223	GRMZM2G060857	GRMZM2G063452
V67	4	3040553	3041840	GRMZM2G167549	GRMZM2G172491
V68	6	153411533	153413061	GRMZM2G359892	AC209257.4_FG003
V69	9	126282535	126283209	GRMZM2G170843	GRMZM2G123709
V70	8	168617820	168619387	GRMZM2G112836	GRMZM2G003963
V71	2	72174964	72176681	GRMZM2G096048	GRMZM2G019056
V72	3	134799288	134800914	GRMZM2G005205	GRMZM2G005419
V73	8	173484792	173485752	GRMZM2G170079	GRMZM2G428179
V74	7	123520437	123521070	GRMZM5G824944	GRMZM2G127537
V75	5	128111515	128112738	GRMZM2G103864	GRMZM2G071089
V76	1	153716286	153716843	GRMZM2G321750	GRMZM2G056039
V77	9	96605233	96610368	GRMZM2G140082	GRMZM2G160730
V78	7	160818645	160820314	GRMZM2G104204	GRMZM2G004694
V79	7	131612131	131613062	GRMZM2G018044	GRMZM2G090149
V80	5	215004940	215007122	GRMZM5G857701	GRMZM2G150876
V81	2	43980857	43982285	GRMZM2G092107	GRMZM2G442523
V82	2	166580633	166581291	GRMZM2G466543	GRMZM2G476810
V83	6	119256734	119257402	GRMZM2G172787	GRMZM2G104511
V84	2	196102413	196103361	GRMZM2G113618	GRMZM2G073942
V85	7	133980311	133981943	GRMZM2G001645	GRMZM2G101711
V86	2	155254653	155256077	GRMZM2G010363	GRMZM2G021299

V87	9	70154222	70155307	GRMZM2G003488	GRMZM2G040706
V88	7	171053569	171054714	GRMZM2G036113	GRMZM2G031298
V89	4	146389143	146389832	GRMZM2G157458	GRMZM2G071448
V90	7	173007006	173008196	GRMZM5G807560	GRMZM2G053554
V91	5	188986859	188988129	GRMZM2G170061	GRMZM2G040924
V92	5	110495944	110497388	GRMZM2G001889	GRMZM2G052821
V93	6	157034481	157036782	GRMZM2G701221	GRMZM2G078781
V94	1	219232030	219233084	GRMZM2G115635	GRMZM2G097342
V95	6	112779464	112781530	GRMZM2G142751	
V96	5	206259875	206260650	AC212112.4_FG002	GRMZM2G116185
V97	5	3209930	3211082	GRMZM2G071071	GRMZM2G003014
V98	3	196032411	196034461	GRMZM2G105863	GRMZM2G324886
V99	2	10375044	10375486	GRMZM2G005155	GRMZM2G005107
V100	5	200120399	200121779	AC233960.1_FG002	GRMZM2G108829

Supplementary Table 7: List of the top 100 putative enhancers in husk based on differential chromatin accessibility and enrichment in H3K9ac. Genes upstream and downstream are indicated (Gramene annotation).

name	chr	start	end	upstream gene	downstream gene
H1	6	160523568	160526882	GRMZM2G432801	GRMZM2G043943
H2	5	49863807	49867322	GRMZM2G477846	GRMZM2G177828
H3	3	184291189	184295375	GRMZM2G702036	GRMZM2G027375
H4	7	113180385	113182315	GRMZM2G114706	GRMZM2G029153
H5	1	54615864	54619723	GRMZM2G312661	GRMZM2G106283
H6	8	166082603	166084645	GRMZM2G387076	GRMZM2G341159
H7	2	84577290	84578239	GRMZM2G117771	GRMZM2G351514
H8	8	170220562	170222705	AC218972.3_FG004	GRMZM5G805026
H9	9	145316095	145318731	GRMZM2G069726	GRMZM2G152919
H10	9	1084647	1085171	GRMZM2G063765	GRMZM2G154678
H11	4	235655241	235659420	GRMZM2G150190	GRMZM2G037185
H12	4	87301781	87302733	GRMZM2G132591	GRMZM2G114778
H13	4	86294569	86299553	GRMZM2G370863	GRMZM2G089698
H14	10	77270828	77274725	GRMZM2G178014	GRMZM2G110145
H15	4	63196241	63197412	GRMZM2G119071	GRMZM2G094500
H16	4	231162586	231165168	GRMZM2G038588	GRMZM2G022538
H17	7	162301442	162303830	GRMZM2G167932	AC185108.3_FG010
H18	8	169610169	169612697	GRMZM2G043498	GRMZM2G043396
H19	3	164566686	164569125	GRMZM2G023257	GRMZM2G077809
H20	4	231186809	231188994	GRMZM2G038588	GRMZM2G022538
H21	1	22461951	22463837	GRMZM2G077127	GRMZM2G457178
H22	3	184412453	184414995	GRMZM2G330159	GRMZM2G339122
H23	5	127393679	127394476	GRMZM2G000812	GRMZM2G144581
H24	7	133922648	133927432	GRMZM5G860241	GRMZM2G001645
H25	9	151440783	151443523	AC149829.2_FG006	GRMZM2G087901
H26	7	172905830	172908211	GRMZM2G042347	GRMZM2G177050

H27	2	62657701	62658930	GRMZM2G093826	GRMZM2G041312
H28	4	237505594	237508779	GRMZM2G111319	GRMZM2G410963
H29	5	27123281	27125549	GRMZM2G174145	GRMZM2G341036
H30	4	11843302	11846147	GRMZM2G017804	GRMZM2G017624
H31	1	39143530	39146013	GRMZM2G160506	GRMZM2G141799
H32	8	36886795	36887521	GRMZM2G143777	GRMZM5G813809
H33	1	70599960	70602487	GRMZM2G524711	GRMZM2G142390
H34	9	143068359	143070625	AC205419.3_FG001	GRMZM2G104626
H35	9	8433549	8435310	GRMZM2G160964	GRMZM2G088053
H36	5	6613275	6616726	GRMZM2G089696	GRMZM2G057623
H37	1	5576284	5577481	GRMZM2G055575	GRMZM2G177508
H38	9	133073363	133074597	GRMZM2G159500	GRMZM2G102382
H39	3	206681400	206685458	GRMZM2G169654	AC233946.1_FG004
H40	2	61919691	61921779	GRMZM2G165063	GRMZM2G122239
H41	5	13813049	13816305	GRMZM2G033489	GRMZM2G033612
H42	10	16034132	16036460	GRMZM5G831224	GRMZM2G181251
H43	1	33676046	33679614	GRMZM2G172537	GRMZM2G165944
H44	4	149574577	149575437	GRMZM5G882482	GRMZM2G522468
H45	5	204922159	204924616	GRMZM2G375116	GRMZM2G089454
H46	10	53806761	53807428	GRMZM2G336859	GRMZM2G170499
H47	2	1540926	1543118	GRMZM2G084958	GRMZM2G084327
H48	6	131727813	131729291	GRMZM2G094666	GRMZM2G441565
H49	4	216775776	216777800	GRMZM2G138427	GRMZM2G138881
H50	7	93185799	93186847	GRMZM2G495846	GRMZM2G154621
H51	2	97470489	97471309	GRMZM5G834532	GRMZM2G095464
H52	1	219663466	219664743	GRMZM2G127396	AC206198.3_FG001
H53	10	117485061	117487380	GRMZM2G176443	GRMZM2G039399
H54	7	166308999	166310761	GRMZM2G097593	GRMZM2G056732
H55	6	134843050	134844946	GRMZM2G061469	GRMZM2G170646
H56	2	6358819	6361005	AC191113.2_FG002	GRMZM5G866678
H57	9	103107829	103109298	GRMZM2G146573	GRMZM2G065394
H58	5	210336001	210337545	GRMZM5G806638	GRMZM2G113418
H59	6	117654870	117656560	GRMZM2G168304	GRMZM2G382711
H60	7	24706671	24707961	GRMZM2G127412	GRMZM2G170542
H61	1	230894944	230896564	GRMZM2G069976	GRMZM2G048434
H62	7	167629097	167630356	GRMZM5G825512	GRMZM2G056407
H63	9	133564366	133566577	GRMZM2G317584	GRMZM2G043983
H64	7	152040293	152045212	GRMZM2G105750	GRMZM2G134545
H65	6	60862307	60863129	GRMZM2G098594	GRMZM2G456023
H66	2	41858857	41861295	GRMZM2G094532	GRMZM2G394321
H67	4	128310066	128310980	GRMZM2G008748	GRMZM2G024647
H68	3	53008264	53011186	GRMZM2G033226	GRMZM2G111593
H69	4	172801069	172806654	GRMZM2G379005	GRMZM2G162052
H70	4	151759326	151761258	GRMZM2G144638	GRMZM2G150714
H71	3	172178588	172181768	GRMZM2G135743	GRMZM2G180509

H72	3	168801412	168807578	GRMZM2G156861	GRMZM2G156861
H73	2	132056565	132057969	GRMZM2G124715	GRMZM2G416491
H74	6	132378636	132381494	GRMZM2G024073	GRMZM2G118951
H75	10	141244949	141246714	GRMZM2G073581	GRMZM2G438202
H76	2	28361341	28363038	GRMZM2G008892	GRMZM2G006964
H77	7	49562484	49563453	GRMZM2G037743	GRMZM2G107377
H78	5	21372449	21375773	GRMZM5G823318	GRMZM5G819464
H79	8	101567142	101568297	GRMZM2G149808	AC234152.1_FG007
H80	3	25320124	25324050	GRMZM2G390236	GRMZM2G340342
H81	7	174593146	174595503	GRMZM2G113011	GRMZM2G056772
H82	8	5254938	5256023	GRMZM2G146012	GRMZM2G017329
H83	1	13676117	13678768	GRMZM2G004641	GRMZM2G417455
H84	5	27876425	27878632	GRMZM2G405090	GRMZM2G159399
H85	9	59483080	59484015	GRMZM2G376416	GRMZM2G115698
H86	3	49080535	49081357	GRMZM2G074742	GRMZM2G066870
H87	7	155628395	155630000	GRMZM2G042047	GRMZM2G000361
H88	7	160108304	160110019	GRMZM5G887631	GRMZM2G121543
H89	1	8424667	8426005	GRMZM2G018706	GRMZM2G157510
H90	9	3603291	3604018	AC212353.4_FG004	GRMZM2G098643
H91	3	167418764	167420237	GRMZM2G093119	GRMZM2G044055
H92	4	91408158	91409858	GRMZM2G081538	GRMZM2G316275
H93	2	18978912	18981518	GRMZM2G036996	GRMZM2G172795
H94	7	19684705	19686933	GRMZM2G111216	AC235546.1_FG001
H95	6	151133120	151137206	GRMZM5G856297	GRMZM2G125976
H96	4	88094624	88095829	GRMZM2G114778	GRMZM2G131699
H97	2	86666620	86667296	GRMZM2G000431	GRMZM2G052268
H98	3	178094901	178097600	GRMZM2G127949	GRMZM2G047851
H99	2	22441341	22442547	GRMZM2G016649	GRMZM2G168917
H100	7	105767172	105768053	GRMZM2G100090	GRMZM2G011030

15 Literature

- Abendroth, L.J., Elmore, R.W., Boyer, M.J., and Marlay, S.K.** (2011). Corn Growth and Development Iowa State Univ.
- Adams, J.M., Harris, A.W., Pinkert, C.A., Corcoran, L.M., Alexander, W.S., Cory, S., Palmiter, R.D., and Brinster, R.L.** (1985). The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature* **318**: 533–538.
- Adrian, J., Farrona, S., Reimer, J.J., Albani, M.C., Coupland, G., and Turck, F.** (2010). cis-Regulatory Elements and Chromatin State Coordinately Control Temporal and Spatial Expression of FLOWERING LOCUS T in Arabidopsis. *Plant Cell Online* **22**: 1425–1440.
- Anders, S. and Huber, W.** (2010). Differential expression analysis for sequence count data. *Genome Biol.* **11**: R106.
- Anders, S., McCarthy, D.J., Chen, Y., Okoniewski, M., Smyth, G.K., Huber, W., and Robinson, M.D.** (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* **8**: 1765–1786.
- Anders, S., Pyl, P.T., and Huber, W.** (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.
- Andersson, R. et al.** (2014). An atlas of active enhancers across human cell types and tissues. *Nature* **507**: 455–461.
- Andersson, R., Sandelin, A., and Danko, C.G.** (2015). A unified architecture of transcriptional regulatory elements. *Trends Genet.* **31**: 426–433.
- Andrés, F. and Coupland, G.** (2012). The genetic basis of flowering responses to seasonal cues. *Nat. Rev. Genet.* **13**: 627–639.
- Aran, D., Sabato, S., and Hellman, A.** (2013). DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol.* **14**: R21.
- Athma, P., Grotewold, E., and Peterson, T.** (1992). Insertional mutagenesis of the maize P gene by intragenic transposition of Ac. *Genetics* **131**: 199–209.
- Athma, P. and Peterson, T.** (1991). Ac induces homologous recombination at the maize P locus. *Genetics* **128**: 163–173.
- Banerji, J., Rusconi, S., and Schaffner, W.** (1981). Expression of a β -globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**: 299–308.
- Bannister, A.J. and Kouzarides, T.** (2011). Regulation of chromatin by histone modifications. *Cell Res.* **21**: 381–395.
- Baucom, R.S., Estill, J.C., Chaparro, C., Upshaw, N., Jogi, A., Deragon, J.-M., Westerman, R.P., SanMiguel, P.J., and Bennetzen, J.L.** (2009). Exceptional Diversity, Non-Random Distribution, and Rapid Evolution of Retroelements in the B73 Maize Genome. *PLoS Genet* **5**: e1000732.

- Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., James Kent, W., and Haussler, D.** (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441**: 87–90.
- Belele, C.L., Sidorenko, L., Stam, M., Bader, R., Arteaga-Vazquez, M.A., and Chandler, V.L.** (2013). Specific Tandem Repeats Are Sufficient for Paramutation-Induced Trans-Generational Silencing. *PLOS Genet.* **9**: e1003773.
- Ben-Naim, O., Eshed, R., Parnis, A., Teper-Bamnolker, P., Shalit, A., Coupland, G., Samach, A., and Lifschitz, E.** (2006). The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. *Plant J.* **46**: 462–476.
- Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E.** (2008). Genome-Wide Association of Histone H3 Lysine Nine Methylation with CHG DNA Methylation in *Arabidopsis thaliana*. *PLoS ONE* **3**: e3156.
- Bewick, V., Cheek, L., and Ball, J.** (2003). Statistics review 7: Correlation and regression. *Crit. Care* **7**: 451–459.
- Bianchi, M., Crinelli, R., Giacomini, E., Carloni, E., and Magnani, M.** (2009). A potent enhancer element in the 5'-UTR intron is crucial for transcriptional regulation of the human ubiquitin C gene. *Gene* **448**: 88–101.
- Blattler, A., Yao, L., Witt, H., Guo, Y., Nicolet, C.M., Berman, B.P., and Farnham, P.J.** (2014). Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biol.* **15**.
- Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K.D., Ovcharenko, I., Pachter, L., and Rubin, E.M.** (2003). Phylogenetic Shadowing of Primate Sequences to Find Functional Regions of the Human Genome. *Science* **299**: 1391–1394.
- Bolger, A.M., Lohse, M., and Usadel, B.** (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
- Bonner, J.T.** (2000). *First Signals: The Evolution of Multicellular Development.* (Princeton University Press).
- Bonner, J.T.** (1998). The origins of multicellularity. *Integr. Biol. Issues News Rev.* **1**: 27–36.
- Bornelöv, S., Komorowski, J., and Wadelius, C.** (2015). Different distribution of histone modifications in genes with unidirectional and bidirectional transcription and a role of CTCF and cohesin in directing transcription. *BMC Genomics* **16**.
- Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E.** (2008). High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell* **132**: 311–322.
- Bratzel, F. and Turck, F.** (2015). Molecular memories in the regulation of seasonal flowering: from competence to cessation. *Genome Biol.* **16**.
- Britten, R.J.** (1997). Mobile elements inserted in the distant past have taken on important functions. *Gene* **205**: 177–182.

- Britten, R.J. and Davidson, E.H.** (1969). Gene Regulation for Higher Cells: A Theory. *Science* **165**: 349–357.
- Bruggmann, R. et al.** (2006). Uneven chromosome contraction and expansion in the maize genome. *Genome Res.* **16**: 1241–1251.
- Cao, S., Kumimoto, R.W., Gnesutta, N., Calogero, A.M., Mantovani, R., and Holt, B.F.** (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 Online*: tpc.113.120352.
- Chandler, V. and Alleman, M.** (2008). Paramutation: Epigenetic Instructions Passed Across Generations. *Genetics* **178**: 1839–1844.
- Chen, H. and Boutros, P.C.** (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12**: 35.
- Chung, B.Y., Simons, C., Firth, A.E., Brown, C.M., and Hellens, R.P.** (2006). Effect of 5'UTR introns on gene expression in Arabidopsis thaliana. *BMC Genomics* **7**: 120.
- Clark, R.M., Wagler, T.N., Quijada, P., and Doebley, J.** (2006). A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat. Genet.* **38**: 594–597.
- Clough, S.J. and Bent, A.F.** (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**: 735–743.
- Coe, E.H.** (1959). A REGULAR AND CONTINUING CONVERSION-TYPE PHENOMENON AT THE B LOCUS IN MAIZE*. *Proc. Natl. Acad. Sci. U. S. A.* **45**: 828–832.
- Coe, E.H.** (2001). The origins of maize genetics. *Nat. Rev. Genet.* **2**: 898–905.
- Creyghton, M.P. et al.** (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci.* **107**: 21931–21936.
- Dalla-Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R.C., and Croce, C.M.** (1982). Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. *Proc. Natl. Acad. Sci.* **79**: 7824–7827.
- Davidson, E.H.** (2001). Chapter 1 - Regulatory Hardwiring: A Brief Overview of the Genomic Control Apparatus and its Causal Role in Development and Evolution. In *Genomic Regulatory Systems* (Academic Press: San Diego), pp. 1–23.
- Davidson, E.H.** (2006). Chapter 2 - cis-Regulatory Modules, and the Structure/Function Basis of Regulatory Logic. In *The Regulatory Genome*, E.H. Davidson, ed (Academic Press: Burlington), pp. 31–86.
- Davidson, R.M., Hansey, C.N., Gowda, M., Childs, K.L., Lin, H., Vaillancourt, B., Sekhon, R.S., de Leon, N., Kaeppler, S.M., Jiang, N., and Buell, C.R.** (2011). Utility of RNA Sequencing for Analysis of Maize Reproductive Transcriptomes. *Plant Genome* **4**: 191–203.

- Daxinger, L., Kanno, T., Bucher, E., Winden, J. van der, Naumann, U., Matzke, A.J.M., and Matzke, M.** (2009). A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. *EMBO J.* **28**: 48–57.
- Dekker, J., Rippe, K., Dekker, M., and Kleckner, N.** (2002). Capturing chromosome conformation. *Science* **295**: 1306–1311.
- Deng, S. and Chua, N.-H.** (2015). Inverted-repeat RNAs targeting FT intronic regions promote FT expression in Arabidopsis. *Plant Cell Physiol.*: pcv091.
- Deng, S., Dai, H., Arenas, C., Wang, H., Niu, Q.-W., and Chua, N.-H.** (2014). Transcriptional Silencing of Arabidopsis Endogenes by Single-Stranded RNAs Targeting the Promoter Region. *Plant Cell Physiol.* **55**: 823–833.
- Dermitzakis, E.T. and Clark, A.G.** (2002). Evolution of Transcription Factor Binding Sites in Mammalian Gene Regulatory Regions: Conservation and Turnover. *Mol. Biol. Evol.* **19**: 1114–1121.
- Doebley, J. and Stec, A.** (1991). Genetic analysis of the morphological differences between maize and teosinte. *Genetics* **129**: 285–295.
- Doebley, J. and Stec, A.** (1993). Inheritance of the morphological differences between maize and teosinte: comparison of results for two F2 populations. *Genetics* **134**: 559–570.
- Doebley, J., Stec, A., and Hubbard, L.** (1997). The evolution of apical dominance in maize. *Nature* **386**: 485–488.
- Du, J., Johnson, L.M., Groth, M., Feng, S., Hale, C.J., Li, S., Vashisht, A.A., Gallego-Bartolome, J., Wohlschlegel, J.A., Patel, D.J., and Jacobsen, S.E.** (2014). Mechanism of DNA Methylation-Directed Histone Methylation by KRYPTONITE. *Mol. Cell* **55**: 495–504.
- Du, J., Johnson, L.M., Jacobsen, S.E., and Patel, D.J.** (2015). DNA methylation pathways and their crosstalk with histone methylation. *Nat. Rev. Mol. Cell Biol.* **16**: 519–532.
- Ducrocq, S., Madur, D., Veyrieras, J.-B., Camus-Kulandaivelu, L., Kloiber-Maitz, M., Presterl, T., Ouzunova, M., Manicacci, D., and Charcosset, A.** (2008). Key Impact of Vgt1 on Flowering Time Adaptation in Maize: Evidence From Association Mapping and Ecogeographical Information. *Genetics* **178**: 2433–2437.
- Ehrlich, M. and Ehrlich, K.C.** (1993). Effect of DNA methylation on the binding of vertebrate and plant proteins to DNA. In *DNA Methylation*, D.J.-P. Jost and D.H.-P. Saluz, eds, EXS. (Birkhäuser Basel), pp. 145–168.
- Ewing, B., Hillier, L., Wendl, M.C., and Green, P.** (1998). Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Res.* **8**: 175–185.
- Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I.** (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* **32**: W273–W279.
- Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E.M.** (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **advance online publication**.

- Goff, S.A., Klein, T.M., Roth, B.A., Fromm, M.E., Cone, K.C., Radicella, J.P., and Chandler, V.L.** (1990). Transactivation of anthocyanin biosynthetic genes following transfer of B regulatory genes into maize tissues. *EMBO J.* **9**: 2517–2522.
- Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S.** (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**: D1178–D1186.
- Gould, S.J. and Vrba, E.S.** (1982). Exaptation-A Missing Term in the Science of Form. *Paleobiology* **8**: 4–15.
- Green, M.R. and Sambrook, J.** (2012). *Molecular Cloning* (Cold Spring Harbor Laboratory Press).
- Gross, D.S. and Garrard, W.T.** (1988). Nuclease Hypersensitive Sites in Chromatin. *Annu. Rev. Biochem.* **57**: 159–197.
- Grotewold, E., Athma, P., and Peterson, T.** (1991). Alternatively spliced products of the maize P gene encode proteins with homology to the DNA-binding domain of myb-like transcription factors. *Proc. Natl. Acad. Sci.* **88**: 4587–4591.
- Gruntman, E., Qi, Y., Slotkin, R.K., Roeder, T., Martienssen, R.A., and Sachidanandam, R.** (2008). Kismeth: Analyzer of plant methylation states through bisulfite sequencing. *BMC Bioinformatics* **9**: 371.
- Gumucio, D.L., Heilstedt-Williamson, H., Gray, T.A., Tarlé, S.A., Shelton, D.A., Tagle, D.A., Slightom, J.L., Goodman, M., and Collins, F.S.** (1992). Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human gamma and epsilon globin genes. *Mol. Cell. Biol.* **12**: 4919–4929.
- Haag, J.R. and Pikaard, C.S.** (2011). Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat. Rev. Mol. Cell Biol.* **12**: 483–492.
- Haberer, G. et al.** (2005). Structure and Architecture of the Maize Genome. *Plant Physiol.* **139**: 1612–1624.
- Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D.** (2002). Two classes of short interfering RNA in RNA silencing. *EMBO J.* **21**: 4671–4679.
- Hanahan, D.** (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**: 557–580.
- Haring, M., Bader, R., Louwers, M., Schwabe, A., van Driel, R., and Stam, M.** (2010). The role of DNA methylation, nucleosome occupancy and histone modifications in paramutation. *Plant J.* **63**: 366–378.
- Haring, M., Offermann, S., Danker, T., Horst, I., Peterhansel, C., and Stam, M.** (2007). Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. *Plant Methods* **3**: 11.
- He, H.H., Meyer, C.A., Hu, S.S., Chen, M.-W., Zang, C., Liu, Y., Rao, P.K., Fei, T., Xu, H., Long, H., Liu, X.S., and Brown, M.** (2014). Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat. Methods* **11**: 73–78.

- Heintzman, N.D. et al.** (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**: 311–318.
- Heintzman, N.D. et al.** (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**: 108–112.
- Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S., and Mullineaux, P.M.** (2000). pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Mol. Biol.* **42**: 819–832.
- Herz, H.-M., Hu, D., and Shilatifard, A.** (2014). Enhancer Malfunction in Cancer. *Mol. Cell* **53**: 859–866.
- Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., Fields, S., and Stamatoyannopoulos, J.A.** (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat. Methods* **6**: 283–289.
- Hetzl, J., Foerster, A.M., Raidl, G., and Scheid, O.M.** (2007). CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. *Plant J.* **51**: 526–536.
- Hong, R.L., Hamaguchi, L., Busch, M.A., and Weigel, D.** (2003). Regulatory Elements of the Floral Homeotic Gene *AGAMOUS* Identified by Phylogenetic Footprinting and Shadowing. *Plant Cell* **15**: 1296–1309.
- Hubbard, L., McSteen, P., Doebley, J., and Hake, S.** (2002). Expression Patterns and Mutant Phenotype of teosinte branched1 Correlate With Growth Suppression in Maize and Teosinte. *Genetics* **162**: 1927–1935.
- Hwang, W., Oliver, V.F., Merbs, S.L., Zhu, H., and Qian, J.** (2015). Prediction of promoters and enhancers using multiple DNA methylation-associated features. *BMC Genomics* **16**: S11.
- Irish, E.E. and Nelson, T.M.** (1991). Identification of multiple stages in the conversion of vegetative to floral development. *Development* **112**: 891–898.
- Jash, A., Yun, K., Sahoo, A., So, J.-S., and Im, S.-H.** (2012). Looping Mediated Interaction between the Promoter and 3' UTR Regulates Type II Collagen Expression in Chondrocytes. *PLoS ONE* **7**: e40828.
- Jeziorska, D.M., Jordan, K.W., and Vance, K.W.** (2009). A systems biology approach to understanding cis-regulatory module function. *Semin. Cell Dev. Biol.* **20**: 856–862.
- Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Böhmendorfer, G., Gregor, W., Kreil, D.P., Matzke, M., and Matzke, A.J.M.** (2008). A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nat. Genet.* **40**: 670–675.
- Karmodiya, K., Krebs, A.R., Oulad-Abdelghani, M., Kimura, H., and Tora, L.** (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.

- Khurana, E. et al.** (2013). Integrative Annotation of Variants from 1092 Humans: Application to Cancer Genomics. *Science* **342**: 1235587.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L.** (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**: R36.
- Kim, T.-K. et al.** (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**: 182–187.
- Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W.J.J., Koornneef, M., and Kakutani, T.** (2007). Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J.* **49**: 38–45.
- Kirienko, D.R., Luo, A., and Sylvester, A.W.** (2012). Reliable Transient Transformation of Intact Maize Leaf Cells for Functional Genomics and Experimental Study. *Plant Physiol.* **159**: 1309–1318.
- Koch, F. et al.** (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* **18**: 956–963.
- Koehler, R., Issac, H., Cloonan, N., and Grimmond, S.M.** (2011). The uniqueome: a mappability resource for short-tag sequencing. *Bioinformatics* **27**: 272–274.
- Koohy, H., Down, T.A., Spivakov, M., and Hubbard, T.** (2014). A Comparison of Peak Callers Used for DNase-Seq Data. *PLOS ONE* **9**: e96303.
- Kovalchuk, I. and Zemp, F.J.** (2010). *Plant Epigenetics: Methods and Protocols* (Humana Press).
- Kowalczyk, M.S. et al.** (2012). Intragenic Enhancers Act as Alternative Promoters. *Mol. Cell* **45**: 447–458.
- Kumimoto, R.W., Adam, L., Hymus, G.J., Repetti, P.P., Reuber, T.L., Marion, C.M., Hempel, F.D., and Ratcliffe, O.J.** (2008). The Nuclear Factor Y subunits NF-YB2 and NF-YB3 play additive roles in the promotion of flowering by inductive long-day photoperiods in *Arabidopsis*. *Planta* **228**: 709–723.
- Kvon, E.Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., and Stark, A.** (2014). Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature* **512**: 91–95.
- Law, J.A. and Jacobsen, S.E.** (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**: 204–220.
- Lettice, L.A. et al.** (2002). Disruption of a long-range cis-acting regulator for *Shh* causes preaxial polydactyly. *Proc. Natl. Acad. Sci.* **99**: 7548–7553.
- Leung, D. et al.** (2015). Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature* **518**: 350–354.
- Levine, M.** (2010). Transcriptional Enhancers in Animal Development and Evolution. *Curr. Biol.* **20**: R754–R763.

- Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A.** (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69**: 905–914.
- Li, G. et al.** (2012). Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation. *Cell* **148**: 84–98.
- Li, H. and Durbin, R.** (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25**: 1754–1760.
- Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R.** (2008). Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell* **133**: 523–536.
- Liu, L., Adrian, J., Pankin, A., Hu, J., Dong, X., von Korff, M., and Turck, F.** (2014). Induced and natural variation of promoter length modulates the photoperiodic response of FLOWERING LOCUS T. *Nat. Commun.* **5**.
- Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C.** (2002). Endogenous and Silencing-Associated Small RNAs in Plants. *Plant Cell* **14**: 1605–1619.
- Louwers, M., Bader, R., Haring, M., Driel, R. van, Laats, W. de, and Stam, M.** (2009). Tissue- and Expression Level-Specific Chromatin Looping at Maize b1 Epialleles. *Plant Cell Online* **21**: 832–842.
- Ludwig, M.Z., Bergman, C., Patel, N.H., and Kreitman, M.** (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403**: 564–567.
- Ludwig, M.Z. and Kreitman, M.** (1995). Evolutionary dynamics of the enhancer region of even-skipped in *Drosophila*. *Mol. Biol. Evol.* **12**: 1002–1011.
- Ludwig, M.Z., Patel, N.H., and Kreitman, M.** (1998). Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development* **125**: 949–958.
- Lund, G., Prem Das, O., and Messing, J.** (1995). Tissue-specific DNase I-sensitive sites of the maize P gene and their changes upon epimutation. *Plant J.* **7**: 797–807.
- Maston, G.A., Evans, S.K., and Green, M.R.** (2006). Transcriptional Regulatory Elements in the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **7**: 29–59.
- Matzke, M.A., Kanno, T., and Matzke, A.J.M.** (2015). RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. *Annu. Rev. Plant Biol.* **66**: 243–267.
- Matzke, M.A. and Mosher, R.A.** (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **advance online publication**.
- Maurano, M.T. et al.** (2012). Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**: 1190–1195.
- Meehan, R.R., Lewis, J.D., and Bird, A.P.** (1992). Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res.* **20**: 5085–5092.

- Melnyk, C.W., Molnar, A., Bassett, A., and Baulcombe, D.C.** (2011). Mobile 24 nt Small RNAs Direct Transcriptional Gene Silencing in the Root Meristems of *Arabidopsis thaliana*. *Curr. Biol.* **21**: 1678–1683.
- Mendenhall, E.M., Williamson, K.E., Reyon, D., Zou, J.Y., Ram, O., Joung, J.K., and Bernstein, B.E.** (2013). Locus-specific editing of histone modifications at endogenous enhancers. *Nat. Biotechnol.* **advance online publication**.
- Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A., and Matzke, A.J.** (2000). Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* **19**: 5194–5201.
- Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P., and Chambon, P.** (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res.* **9**: 6047–6068.
- Moreno, M.A., Chen, J., Greenblatt, I., and Dellaporta, S.L.** (1992). Reconstititional mutagenesis of the maize P gene by short-range Ac transpositions. *Genetics* **131**: 939–956.
- Nagano, null, Wu, null, Kawasaki, null, Kishima, null, and Sano, null** (1999). Genomic organization of the 260 kb surrounding the waxy locus in a Japonica rice. *Genome Natl. Res. Counc. Can. Genome Cons. Natl. Rech. Can.* **42**: 1121–1126.
- Nagarajan, S., Benito, E., Fischer, A., and Johnsen, S.A.** (2015). H4K12ac is regulated by estrogen receptor-alpha and is associated with BRD4 function and inducible transcription. *Oncotarget* **6**: 7305–7317.
- Neph, S. et al.** (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**: 83–90.
- Norris, J., Fan, D., Aleman, C., Marks, J.R., Futreal, P.A., Wiseman, R.W., Iglehart, J.D., Deininger, P.L., and McDonnell, D.P.** (1995). Identification of a New Subclass of Alu DNA Repeats Which Can Function as Estrogen Receptor-dependent Transcriptional Enhancers. *J. Biol. Chem.* **270**: 22777–22782.
- Okano, Y., Miki, D., and Shimamoto, K.** (2008). Small interfering RNA (siRNA) targeting of endogenous promoters induces DNA methylation, but not necessarily gene silencing, in rice. *Plant J.* **53**: 65–77.
- Pedmale, U.V., Huang, S.C., Zander, M., Cole, B.J., Hetzel, J., Ljung, K., Reis, P.A.B., Sridevi, P., Nito, K., Nery, J.R., Ecker, J.R., and Chory, J.** (2016). Cryptochromes Interact Directly with PIFs to Control Plant Growth in Limiting Blue Light. *Cell* **164**: 233–245.
- Pepke, S., Wold, B., and Mortazavi, A.** (2009). Computation for ChIP-seq and RNA-seq studies. *Nat. Methods* **6**: S22–S32.
- Pfaffl, M.W.** (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**: e45.
- Polak, P. and Domany, E.** (2006). Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* **7**: 133.

- Quinlan, A.R. and Hall, I.M.** (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.
- R Core Team** (2014). R: A Language and Environment for Statistical Computing.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J.** (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**: 279–283.
- Radicella, J.P., Brown, D., Tolar, L.A., and Chandler, V.L.** (1992). Allelic diversity of the maize B regulatory gene: different leader and promoter sequences of two B alleles determine distinct tissue specificities of anthocyanin production. *Genes Dev.* **6**: 2152–2164.
- Rebeiz, M., Jikomes, N., Kassner, V.A., and Carroll, S.B.** (2011). Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proc. Natl. Acad. Sci.* **108**: 10036–10043.
- Regulski, M. et al.** (2013). The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. *Genome Res.* **23**: 1651–1662.
- Reimer, J.J. and Turck, F.** (2010). Genome-Wide Mapping of Protein-DNA Interaction by Chromatin Immunoprecipitation and DNA Microarray Hybridization (ChIP-chip). Part A: ChIP-chip Molecular Methods. In *Plant Epigenetics*, I. Kovalchuk and F.J. Zemp, eds, *Methods in Molecular Biology*. (Humana Press), pp. 139–160.
- Rubinstein, M. and Souza, F.S.J. de** (2013). Evolution of transcriptional enhancers and animal diversity. *Philos. Trans. R. Soc. B Biol. Sci.* **368**: 20130017.
- Russell, W.K. and Stuber, C.W.** (1983). Inheritance of Photosensitivity in Maize. *Crop Sci.* **23**: 935–939.
- Salvi, S. et al.** (2007). Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc. Natl. Acad. Sci.* **104**: 11376–11381.
- Santangelo, A.M., de Souza, F.S.J., Franchini, L.F., Bumaschny, V.F., Low, M.J., and Rubinstein, M.** (2007). Ancient Exaptation of a CORE-SINE Retroposon into a Highly Conserved Mammalian Neuronal Enhancer of the Proopiomelanocortin Gene. *PLoS Genet* **3**: e166.
- Schmieder, R. and Edwards, R.** (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**: 863–864.
- Schnable, P.S. et al.** (2009). The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* **326**: 1112–1115.
- Syednasrollah, F., Laiho, A., and Elo, L.L.** (2013). Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief. Bioinform.:* bbt086.
- Shen, L.** (2013). GeneOverlap: Test and visualize gene overlaps. R package version 1.6.0.
- Sherwood, R.I., Hashimoto, T., O'Donnell, C.W., Lewis, S., Barkal, A.A., van Hoff, J.P., Karun, V., Jaakkola, T., and Gifford, D.K.** (2014). Discovery of directional and

- nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32**: 171–178.
- Shlyueva, D., Stampfel, G., and Stark, A.** (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**: 272–286.
- Sidorenko, L., Li, X., Tagliani, L., Bowen, B., and Peterson, T.** (1999). Characterization of the regulatory elements of the maize P-rr gene by transient expression assays. *Plant Mol. Biol.* **39**: 11–19.
- Sidorenko, L.V. and Peterson, T.** (2001). Transgene-Induced Silencing Identifies Sequences Involved in the Establishment of Paramutation of the Maize p1 Gene. *Plant Cell* **13**: 319–335.
- Sloan, A.E., Sidorenko, L., and McGinnis, K.M.** (2014). Diverse Gene-Silencing Mechanisms with Distinct Requirements for RNA Polymerase Subunits in *Zea mays*. *Genetics* **198**: 1031–1042.
- Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., Erik van Nimwegen, Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., Tiwari, V.K., and Schübeler, D.** (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**: 490–495.
- Stam, M., Bebele, C., Dorweiler, J.E., and Chandler, V.L.** (2002). Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. *Genes Dev.* **16**: 1906–1918.
- Steinmüller, K. and Apel, K.** (1986). A simple and efficient procedure for isolating plant chromatin which is suitable for studies of DNase I-sensitive domains and hypersensitive sites. *Plant Mol. Biol.* **7**: 87–94.
- Stone, J.R. and Wray, G.A.** (2001). Rapid Evolution of cis-Regulatory Sequences via Local Point Mutations. *Mol. Biol. Evol.* **18**: 1764–1770.
- Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J.** (2011). Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat. Genet.* **43**: 1160–1163.
- Styles, E.D. and Ceska, O.** (1981). P and R Control of Flavonoids in Bronze Coleoptiles of Maize. *Can. J. Genet. Cytol.* **23**: 691–704.
- Styles, E.D., Ceska, O., and Seah, K.-T.** (1973). Developmental Differences in Action of R and B Alleles in Maize. *Can. J. Genet. Cytol.* **15**: 59–72.
- Tagle, D.A., Koop, B.F., Goodman, M., Slightom, J.L., Hess, D.L., and Jones, R.T.** (1988). Embryonic ϵ and γ globin genes of a prosimian primate (*Galago crassicaudatus*). *J. Mol. Biol.* **203**: 439–455.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S.** (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* **28**: 2731–2739.
- Taub, R., Kirsch, I., Morton, C., Lenoir, G., Swan, D., Tronick, S., Aaronson, S., and Leder, P.** (1982). Translocation of the *c-myc* gene into the immunoglobulin heavy chain locus in

- human Burkitt lymphoma and murine plasmacytoma cells. *Proc. Natl. Acad. Sci.* **79**: 7837–7841.
- Thurman, R.E. et al.** (2012). The accessible chromatin landscape of the human genome. *Nature* **489**: 75–82.
- Tiwari, S.B. et al.** (2010). The flowering time regulator CONSTANS is recruited to the FLOWERING LOCUS T promoter via a unique cis-element. *New Phytol.* **187**: 57–66.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L.** (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**: 562–578.
- Treangen, T.J. and Salzberg, S.L.** (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* **13**: 36–46.
- Vansant, G. and Reynolds, W.F.** (1995). The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element. *Proc. Natl. Acad. Sci. U. S. A.* **92**: 8229–8233.
- Vera, D.L., Madzima, T.F., Labonne, J.D., Alam, M.P., Hoffman, G.G., Girimurugan, S.B., Zhang, J., McGinnis, K.M., Dennis, J.H., and Bass, H.W.** (2014). Differential Nuclease Sensitivity Profiling of Chromatin Reveals Biochemical Footprints Coupled to Gene Expression and Functional DNA Elements in Maize. *Plant Cell Online*: tpc.114.130609.
- Vernimmen, D., Gobbi, M.D., Sloane-Stanley, J.A., Wood, W.G., and Higgs, D.R.** (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J.* **26**: 2041–2051.
- Vierstra, J., Wang, H., John, S., Sandstrom, R., and Stamatoyannopoulos, J.A.** (2014). Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH. *Nat. Methods* **11**: 66–72.
- Visel, A. et al.** (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**: 854–858.
- Visel, A., Bristow, J., and Pennacchio, L.A.** (2007). Enhancer identification through comparative genomics. *Semin. Cell Dev. Biol.* **18**: 140–152.
- Vlăduțu, C., McLaughlin, J., and Phillips, R.L.** (1999). Fine Mapping and Characterization of Linked Quantitative Trait Loci Involved in the Transition of the Maize Apical Meristem From Vegetative to Generative Structures. *Genetics* **153**: 993–1007.
- Wang, Y.-M., Zhou, P., Wang, L.-Y., Li, Z.-H., Zhang, Y.-N., and Zhang, Y.-X.** (2012). Correlation Between DNase I Hypersensitive Site Distribution and Gene Expression in HeLa S3 Cells. *PLoS ONE* **7**: e42414.
- Weintraub, H. and Groudine, M.** (1976). Chromosomal subunits in active genes have an altered conformation. *Science* **193**: 848–856.
- Wendel, J.F., Cronn, R.C., Alvarez, I., Liu, B., Small, R.L., and Senchina, D.S.** (2002). Intron Size and Genome Size in Plants. *Mol. Biol. Evol.* **19**: 2346–2352.

- Wenkel, S., Turck, F., Singer, K., Gissot, L., Gourrierc, J.L., Samach, A., and Coupland, G.** (2006). CONSTANS and the CCAAT Box Binding Complex Share a Functionally Important Domain and Interact to Regulate Flowering of Arabidopsis. *Plant Cell Online* **18**: 2971–2984.
- Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., and Romano, L.A.** (2003). The Evolution of Transcriptional Regulation in Eukaryotes. *Mol. Biol. Evol.* **20**: 1377–1419.
- Wroblewski, T., Matvienko, M., Piskurewicz, U., Xu, H., Martineau, B., Wong, J., Govindarajulu, M., Kozik, A., and Michelmore, R.W.** (2014). Distinctive profiles of small RNA couple inverted repeat-induced post-transcriptional gene silencing with endogenous RNA silencing pathways in Arabidopsis. *RNA* **20**: 1987–1999.
- Wu, L., Mao, L., and Qi, Y.** (2012). Roles of DICER-LIKE and ARGONAUTE Proteins in TAS-Derived Small Interfering RNA-Triggered DNA Methylation. *Plant Physiol.* **160**: 990–999.
- Yoo, S.K., Chung, K.S., Kim, J., Lee, J.H., Hong, S.M., Yoo, S.J., Yoo, S.Y., Lee, J.S., and Ahn, J.H.** (2005). CONSTANS Activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to Promote Flowering in Arabidopsis. *Plant Physiol.* **139**: 770–778.
- Yu, C.-P. et al.** (2015). Transcriptome dynamics of developing maize leaves and genomewide prediction of cis elements and their cognate transcription factors. *Proc. Natl. Acad. Sci.* **112**: E2477–E2486.
- Zentner, G.E., Tesar, P.J., and Scacheri, P.C.** (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* **21**: 1273–1283.
- Zhang, P., Chopra, S., and Peterson, T.** (2000). A Segmental Gene Duplication Generated Differentially Expressed myb-Homologous Genes in Maize. *Plant Cell* **12**: 2311–2322.
- Zhang, W., Wu, Y., Schnable, J.C., Zeng, Z., Freeling, M., Crawford, G.E., and Jiang, J.** (2012a). High-resolution mapping of open chromatin in the rice genome. *Genome Res.* **22**: 151–162.
- Zhang, W., Zhang, T., Wu, Y., and Jiang, J.** (2012b). Genome-Wide Identification of Regulatory DNA Elements and Protein-Binding Footprints Using Signatures of Open Chromatin in Arabidopsis. *Plant Cell Online*.
- Zhang, W., Zhang, T., Wu, Y., and Jiang, J.** (2014). Open Chromatin in Plant Genomes. *Cytogenet. Genome Res.*
- Zhang, X., Bernatavichute, Y.V., Cokus, S., Pellegrini, M., and Jacobsen, S.E.** (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. *Genome Biol.* **10**: R62.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S.** (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**: R137.
- Zheng, B., Wang, Z., Li, S., Yu, B., Liu, J.-Y., and Chen, X.** (2009). Intergenic transcription by RNA Polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. *Genes Dev.* **23**: 2850–2860.

-
- Zhong, S., Fei, Z., Chen, Y.-R., Zheng, Y., Huang, M., Vrebalov, J., McQuinn, R., Gapper, N., Liu, B., Xiang, J., Shao, Y., and Giovannoni, J.J.** (2013). Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nat. Biotechnol.* **31**: 154–159.
- Zhu, B., Zhang, W., Zhang, T., Liu, B., and Jiang, J.** (2015). Genome-Wide Prediction and Validation of Intergenic Enhancers in Arabidopsis Using Open Chromatin Signatures. *Plant Cell*: tpc.15.00537.
- Zylicz, J.J., Dietmann, S., Günesdogan, U., Hackett, J.A., Cougot, D., Lee, C., and Surani, M.A.** (2015). Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. *eLife* **4**: e09571.

16 Abbreviations

General abbreviations

-	minus, not present
%	percentage
°C	degrees Celsius
μ	micro
3'	three prime end of DNA fragment
35S	promoter of the cauliflower mosaic virus
3C	chromosomal conformation capture
5'	five prime end of DNA fragment
A	Adenine
<i>A. thaliana</i>	<i>Arabidopsis thaliana</i>
<i>A. tumefaciens</i>	<i>Agrobacterium tumefaciens</i>
Ac	Activator
B73	sequenced <i>Z. mays</i> inbred line B73
bp	base pair
BSA	Bovine Serum Albumin
C	Cytosine
CA	Co-Activator
cDNA	complementary DNA
CG	dinucleotide in symmetric context
CHG	trinucleotide in symmetric context with H being A, C, or T
CHH	trinucleotide in asymmetric context with H being A, C, or T
ChIP	Chromatin immunoprecipitation
CNS	Conserved Non-coding Sequence
Col-0	<i>Arabidopsis thaliana</i> accession Columbia
CRE	<i>Cis</i> -Regulatory Element
DEG	Differentially Expressed Gene
DHS	DNase I Hypersensitive Site
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleic triphosphate
<i>Drosophila</i>	<i>Drosophila melanogaster</i>
dsRNA	double-stranded RNA
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
et al.	et aliae [Lat.] and others
FPKM	Fragments Per Kilobase of exon model per million fragments Mapped
G	Guanine
g	gram
gDNA	genomic DNA
GM	½ strength Murashige and Skoog medium
h	hour

H3	histone 3
H3K14ac	histone 3 lysine 14 acetylation
H3K27ac	histone 3 lysine 27 acetylation
H3K27me3	histone 3 lysine 27 trimethylation
H3K4me1	histone 3 lysine 4 monomethylation
H3K4me2	histone 3 lysine 4 dimethylation
H3K9ac	histone 3 lysine 9 acetylation
H4K12ac	histone 4 lysine 12 acetylation
IR	Inverted Repeat
IST	Inner Stem Tissue
k	kilo
kb	kilobase pair
l	liter
LB	Luria Broth
LD	Long Day
M	molar (mol/l)
m	milli
Mb	megabase pair
MD	Mid Day
min	minute
Mite	Miniature inverted-repeat
mol	mole
MPI	Max Planck Institute for Plant Breeding Research
mRNA	messenger RNA
n	nano
NDR	Nucleosome-Depleted Region
nt	nucleotide
P/C/IAA	Phenol/Chloroform/Isoamyl alcohol 25:24:1
PC	Principal Component
PCR	Polymerase Chain Reaction
pH	decimal logarithm of the reciprocal of the hydrogen ion activity
PMSF	phenylmethylsulfonyl fluoride
QTL	Quantitative Trait Locus
RdDM	RNA-dependent DNA Methylation
Rice	<i>Oryza sativa</i>
RNA	Ribonucleic acid
RNase	Ribonuclease
RPM	Read Per Million mapped reads
RT	Room Temperature
SEM	Standard Error of the Mean
smRNA	small RNA
Sorghum	<i>Sorghum bicolor</i>
ssRNA	single-stranded RNA
SV40	Simian Virus 40
T	Thymine

T1, T2, T3, etc.	first, second, third, etc. filial generation after transformation
T-DNA	Transferred DNA
TE	Transposable Element
TF	Transcription Factor
TFBS	Transcription Factor Binding Site
TGS	Transcriptional Gene Silencing
TSS	Transcription Start Site
UTR	Untranslated Region
UV	Ultraviolet
UvA	University of Amsterdam
Vgt1	<i>Vegetative to generative transition1</i>
WT	Wild Type
<i>Z. mays</i>	<i>Zea mays</i>
ZT	Zeitgeber Time

Abbreviations of gene and protein names

AGO4	ARGONAUTE 4
AGO6	ARGONAUTE 6
AP1	APETALA 1
b1	booster1
CHS	CHALCONE SYNTHASE
CMT2	CHROMOMETHYLASE 2
CMT3	CHROMOMETHYLASE 3
CO	CONSTANS
DCL2	DICER-LIKE 2
DCL3	DICER-LIKE 3
DCL4	DICER-LIKE 4
DMR2	DOMAIN REARRANGED METHYLTRANSFERASE 2
EF1 α -A3	ELONGATION FACTOR 1 α -A3
FLC	FLOWERING LOCUS C
FT	FLOWERING LOCUS T
FWA	FLOWERING WAGENINGEN
GFP	GREEN FLUORESCENT PROTEIN
GUS	β -glucuronidase
HEN1	HUA ENHANCER 1
MET1	METHYLTRANSFERASE 1
NF-Y	NUCLEAR FACTOR-Y
p1	pericarp color1
PIF4	PHYTOCHROME-INTERACTING FACTOR 4
Pol II	RNA polymerase II
Pol IV	RNA polymerase IV
Pol V	RNA polymerase V
RDR2	RNA-dependent RNA polymerase 2
RIN	RIPENING INHIBITOR

Shh	Sonic hedgehog
SOC1	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1
SVP	SHORT VEGETATIVE PHASE
tb1	teosinte branched1
TMM	TOO MANY MOUTHS

17 Acknowledgments

I would first like to thank my supervisor **Dr. Franziska Turck** for having given me the chance to work on such an interesting research project with the great facilities provided by the Max Planck Institute for Plant Breeding Research. She was always going to the point during our discussions and trusted me for the details. I felt I had a total freedom in my research and I hope to continue to work this way in the future.

In addition, my PhD took place within the **Marie Skłodowska-Curie ITN EpiTRAITS (GA-316965)**. This network provided me exceptional conditions to perform my research with many workshops, conferences, collaborations, and opportunities to start building a strong professional network. Within this network, I would especially like to thank **Dr. Maïke Stam** and her two PhD students **Blaise Weber** and **Rurika Oka**, both fellows from the EpiTRAITS ITN. We collaborated on the maize part of this thesis. Dr. Maïke Stam supported me a lot during my thesis and was like a second supervisor to me.

I would like to thank **all the members of the Turck's lab**, especially **Emmanuel Tergemina** for the nice discussions on the Arabidopsis project, **Dr. Liangyu Liu** who helped me to get into the project at the beginning of my PhD, **Dr. Yue Zhou** for his technical help on basically everything happening in the lab, and **Petra Tänzler** for doing all the experiments during the writing of my PhD thesis and for her continuous good mood.

I would also like to thank:

Prof. Dr. Coupland who gave me the chance to work in his department, extended my contract so I could finish peacefully my thesis, and for being my first reviewer.

Prof. Dr. Ute Höcker for being my second reviewer.

Prof. Dr. Juliette De Meaux for being the head of my thesis committee.

At last, I would like to warmly thank **Olga Tur** for her support and her love.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit -einschließlich Tabellen, Karten und Abbildungen, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie abgesehen von unten angegebenen Teilpublikationen noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. George Coupland betreut worden.

Köln, November 2017

Johan Zicola

18 Lebenslauf

Persönliche Angaben

Name: Johan Zicola

Geburtsdatum: 02. Dezember 1987

Geburtsort: Nîmes, Frankreich

Nationalität: französisch

Familienstand: ledig

E-mail: johan.zicola@gmail.com

Ausbildung

Seit 03/2013: Promotions-Studium an der Universität zu Köln, Köln, Deutschland Max-Planck-Institut für Züchtungsforschung, unter der Leitung von Prof. Dr. G. Coupland Stipendium der Marie Skłodowska-Curie actions (EPITRAITS, project 316965, FP7-PEOPLE).

2009-2012: Master in Agrarwissenschaft, AGROCAMPUS OUEST, Rennes, Frankreich

2007-2009: Diplom in Labortechnik, IUT Paul Sabatier, Auch, Frankreich

2005-2007: Abitur, EPLEFPA Nîmes Rodhilan, Frankreich