UNIVERSITEIT VAN AMSTERDAM

*Identification and Validation of Maize Enhancers. A Cartography of the Maize Regulatory Genome*

B.B.M. Weber

# Summary


## Identification and Validation of Maize Enhancers

A cartography of the maize regulatory genome

The work described in this thesis aims at the genome-wide identification of novel enhancer sequences in the maize genome and their subsequent characterization and validation. Enhancers are non-coding DNA sequences that regulate transcription of their target genes and are necessary to establish the complex and diverse patterns of gene expression of multicellular organisms. Enhancers are therefore vital contributors to genic transcriptional regulation. However, despite their importance, the genome-wide identification and characterization of plant enhancers received until recently little attention. This thesis therefore aims at a better characterisation of plant enhancers through identification and validation of candidate sequences in the crop plant *Zea mays.*

In **chapter 1** of this thesis, we review knowledge on plant and animal enhancers and provide a list of previously identified enhancers in plants. In addition, we review the methods allowing medium to large-scale identification of enhancer sequences and validation of identified sequences, and discuss the advantages and drawbacks associated with each method.

In **chapter 2**, we used a combination of methods described in chapter 1 to identify novel enhancers in the genome of the B73 maize line. Candidate sequences were defined as regions of low DNA methylation displaying enriched levels of chromatin accessibility and H3K9ac. In total, using two tissues (V2-IST and husk), we identified 1702 candidate enhancers. In more detail, we identified 472 and 1500 enhancers in V2-IST and husk tissues, respectively, while 270 enhancer candidates were predicted to be active in both tissues. Also, when possible, each candidate enhancer was assigned one or two target genes, based on the expression levels of the directly upstream and downstream located coding sequences. Importantly, three out of five previously reported enhancers were present in our list of candidates: the putative DICE enhancer and the enhancers of gene *b1* and *tb1*. The profiles of candidate sequences appeared similar to those of promoters. Unlike seen in animals, both displayed asymmetric enrichment of H3K9ac and, the unidirectional production of transcripts. In animals, enhancers are indicated to show symmetric enrichment of histone acetylation and

produce transcripts in a bidirectional manner. In addition, 10% and 18% of the candidates in V2-IST and husk, respectively, overlapped with conserved non-coding sequences (rice to maize comparison), supporting a functional role for these candidate enhancers. Finally, about 30% of candidate enhancers overlapped with transposable elements and a large part of these candidates possibly emerged from transposable elements.

In **chapter 3**, attention was given to the *Vgt1-ZmRap2.7* locus. *Vgt1* was previously predicted as a regulatory sequence controlling expression of the downstream *ZmRap2.7* gene, which encodes a protein that represses the switch from vegetative growth to reproductive growth (i.e. flowering). In addition, we examined an H3K9ac enriched region just upstream of *Vgt1*, coined *uva1*. We showed that the combined region is characterized by enhancer-like characteristics (H3K9ac enrichment, high levels of accessibility and low levels of DNA methylation). Moreover, by using GUS reporter genes, we demonstrated that both *uva1* and *Vgt1* might function as tissue-specific regulatory sequences: transcriptional enhancers in some tissues and transcriptional silencers or insulators in others. We show that DNA methylation-mediated silencing of *Vgt1* was sufficient to induce early flowering, and also increased growth rate of the corresponding transgenic plants. The effect on flowering time is consistent with *ZmRap2.7* being the target gene of *Vgt1*. Surprisingly, no clear downregulation of *ZmRap2.7* expression was detected upon induced silencing of *Vgt1*. These results indicate that (*i*) *Vgt1* may regulate *ZmRap2.7* in only a few cells within the tissues examined, (*ii*) *ZmRap2.7* expression is also controlled by other regulatory sequences, (*iii*) the effect on *ZmRap2.7* expression can only be observed in other tissues than examined, or alternatively, (*iv*) that *Vgt1* affects flowering time by regulating other genes. Interestingly, using 4C-seq we observed physical interactions between the promoter of *ZmRap2.7* and a genomic region containing *Vgt1*. In addition, the promoter of *ZmRap2.7* was also found to interact with other protein-coding genes, lncRNA loci and candidate *cis*-regulatory regions. Altogether our results show that *Vgt1* is a regulatory sequence that can enhance expression of a reporter gene, and delay flowering time, the latter possibly via enhancing *ZmRap2.7* expression, potentially in collaboration with other *cis*-regulatory sequences.

In **chapter 4**, a subset of candidate enhancers and control regions was selected and cloned into reporter systems for validation purposes. Two candidate regions, H11/V426 and H112/V441, appeared able to function as a silencer and/or insulator in the tissues examined. These results indicate that regions identified based on enhancer-associated marks in chapter 2 are not necessarily all acting as transcriptional enhancers; part of them are likely to act as silencers or insulators or combine different functions. Hence, enhancer candidates should

therefore rather be defined as *cis*-regulatory regions. We also show that regions defined as LUMRs (low or un-methylated regions; ex-V29 and neutral region 4) are capable of acting as enhancers and/or silencer or insulator in other tissues than the ones used to predict enhancer candidates. This suggests that LUMRs represent the entire pool of *cis*-regulatory elements in B73. Finally, using 4C-seq, we show that candidate H112/V441 and the LUMR Neutral 4 display tissue-specific interactomes, and are interacting with other active regions of the genome. Interestingly, genic regions specifically interacting with H112/V441 in V2-IST tissue were in general expressed at higher levels in this tissue than in husk tissue, suggesting a potential role for the candidate region as a transcriptional enhancer.

In **chapter 5,** a protocol for Chromatin Conformation Capture (3C) is detailed. 3C offers the opportunity to resolve chromatin interactions between sequences of interest. We therefore provided a step-by-step 3C protocol particularly adapted to both maize and Arabidopsis, and included tips and details aimed at facilitating the successful application of the 3C method.

Finally, in the **General Discussion**, the main findings of this thesis are discussed and critically evaluated in the context of literature, providing a bird eye view on the maize regulatory genome.