CellPress
PARTNER JOURNAL

Molecular Plant
Comment

Opportunities to Use DNA Methylation to Distil Functional Elements in Large Crop Genomes

DNA methylation is a chromatin modification that is often associated with the exciting and sometimes unpredictable patterns of inheritance that can unfold with epigenetic phenomena. The stability and heritability of DNA methylation patterns perhaps allow us to utilize DNA methylation profiles to distil the suite of potentially functional elements in large crop genomes. Here, we discuss the potential and possible ways to use the absence of DNA methylation to identify potential regulatory regions within intergenic sequences and the presence of DNA methylation to identify pseudogenes.

DNA methylation plays important roles in the regulation of transposable elements and may also contribute to the regulation of gene expression and control of recombination. A wealth of information from genetic and genomic analyses has revealed the molecular mechanisms that control DNA methylation in *Arabidopsis thaliana* (*Arabidopsis*) (Niederhuth and Schmitz, 2017). Methylation is found in three contexts depending on the nucleotides adjacent to the methylated cytosine: CG, CHG, and CHH (H representing A, T, or G). Broad surveys of context-specific DNA methylation levels in other plant species have revealed many similarities to the *Arabidopsis* methylome and several notable differences (Niederhuth et al., 2016). The genomic levels of DNA methylation vary among species but are often correlated with the proportion of the genome derived from transposable elements. Transposable elements are highly methylated, while genes tend to show reduced levels of methylation relative to flanking sequences (West et al., 2014; Niederhuth et al., 2016). The specific role of DNA methylation in plants remains the subject of research efforts in many species. Profiles of DNA methylation, particularly in the CHG context, can be useful as a tool to identify the small fraction of intergenic space that contains potential regulatory function and as a filter to identify annotated genes that may represent cryptic information.

What is it about DNA methylation, especially in the CHG context, that makes it potentially valuable to identify functional elements in plant genomes? One major advantage of using methylation data compared with other chromatin modifications is the stability of DNA methylation throughout the majority of vegetative development (Eichten et al., 2013; Schmitz et al., 2013; Kawakatsu et al., 2016). Yet, the different methylation contexts have differing values and predictability. While CHH methylation can exhibit dynamic patterns, the patterns of CG and CHG methylation are largely stable outside of some cell types involved in plant reproduction (Kawakatsu et al., 2016; Bouyer et al., 2017). CG and CHG methylation largely occurs in heterochromatic and other silent parts of the genome. A complication with CG methylation is that it also occurs in gene bodies, where it has an enigmatic function, somewhat correlated with increased levels of gene expression (Niederhuth and Schmitz, 2017). Thus, of the three sequence contexts, CHG likely holds the most valuable information, largely stable over development and predominantly associated with silent chromatin. In this opinion article, we are focused on using the absence of CHG methylation as an indicator that a region may have potential functional relevance.

The natural variation that can influence traits associated with domestication or agronomic improvement of plant species is often found to affect *cis*-regulatory elements that influence the developmental patterns or levels of gene expression. As we consider new uses for genome editing technologies in crop improvement, it is likely that changes in the regulation of gene expression will be a fruitful endeavor in the search for new traits or improved resilience of plants. However, we will be limited by our ability to identify the *cis*-regulatory regions that are critical for the proper regulation of gene expression among the vast intergenic space of crop genomes. Many crop species have relatively large genomes with complex organization of genes and repetitive elements. This leads to substantial increases in the amount of intergenic space relative to species such as *Arabidopsis*. For example, maize and barley contain ~2 and ~4 Gb, respectively, of intergenic space >5 kb from the nearest gene (Figure 1A). There is growing evidence for the presence of distal (up to 100 kb away) regulatory regions that can be critical for important plant traits (reviewed by Weber et al., 2016). Arguably, large crop genomes have more functional genome space and adapt primarily through mutations in regulatory regions (Mei et al., 2018). This has led to critical questions about how to identify the functionally relevant portions of the intergenic space of species such as maize to find candidate regulatory regions.

Several studies have highlighted the potential for knowledge of chromatin accessibility to provide insights into potentially relevant regions of the maize genome. Regions with accessible chromatin exhibit 20-fold enrichment for functional variation (Rodgers-Melnick et al., 2016). Oka et al. (2017) utilized a combination of chromatin accessibility, histone acetylation, and DNA methylation data to identify a set of 1500 putative regulatory regions within the intergenic space of maize; although this is likely an underestimate, as it only surveyed chromatin of two tissues. Chromatin accessibility has proven useful for predicting regulatory interactions. However, ACRs and their regulatory interactions are highly dynamic across tissues and conditions (Sijacic et al., 2018). This means that a comprehensive identification of regulatory regions would require profiling of chromatin accessibility and/or histone modifications

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, SIBS, CAS.

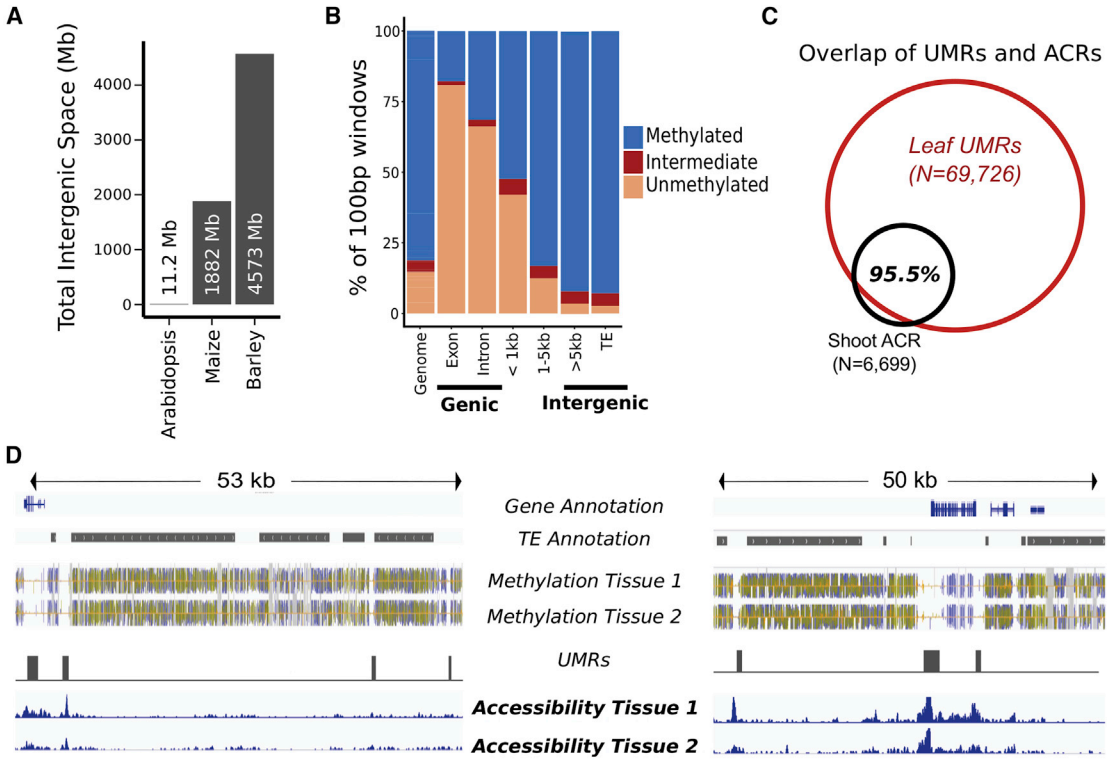

Figure 1. Uses of DNA Methylation to Distil Potential Regulatory Elements in Large Crop Genomes.

(A) The total amount of intergenic space (DNA >5 kb from nearest annotated gene) is shown for *Arabidopsis*, maize, and barley.

(B). The level of CHG methylation was used to classify 100 bp windows of the maize genome as methylated (>40% CHG methylation), intermediate (20–40% CHG methylation), or unmethylated (<20% CHG methylation). The distribution of these methylation levels was assessed for various portions of the maize genome.

(C) A comparison of the unmethylated regions (CHG methylation <10%) and accessible regions (Oka et al., 2017) within the 1882 Mb of intergenic space of the maize genome.

(D) Two regions of the maize genome are shown to illustrate the patterns of unmethylated regions relative to annotated features and accessible DNA. The methylation levels are shown with the blue, orange, and green bars showing levels of CG, CHG, and CHH methylation on both strands. The genes tend to have low methylation or only CG methylation. Several unmethylated regions (UMRs) are indicated. Some of these overlap with ACRs or gene regions, while others occur in intergenic space and are not accessible in the shoot tissue.

in many different tissues and environmental conditions. In contrast, the developmental stability of DNA methylation may provide a guide to the full set of intergenic regions with potential regulatory roles based on the profiling of a single tissue.

The vast majority of intergenic space (regions >5 kb from genes) is highly methylated in maize (Figure 1B). Although much of this can be attributed to the fact that a majority of this sequence is derived from transposable elements, we find that even the non-transposon sequences within these regions tend to be highly methylated (Figure 1B). However, there are many unmethylated regions that are found in this intergenic space. The majority of regions with accessible chromatin also have low levels of DNA methylation but there are many regions of low methylation without accessible chromatin (Figure 1C and 1D). One possible interpretation is that many of these lowly methylated regions may exhibit accessible chromatin in other tissues. Importantly, the intergenic regions that exhibit low levels of DNA methylation are reproducible among different tissues (Figure 1D). The identification of the unmethylated portion of the intergenic space of complex crop genomes could be used as a filter to focus on regions with potential roles in regulation.

A second use for DNA methylation data is the filtering of putative gene models. Annotation of genes is a complex problem with many factors that can lead to the identification of false positives or false negatives. In some cases, a true gene may lack expression in the set of tissues for which expression data has been generated and therefore will have limited experimental support. In other cases, putative genes may represent fragments of genes that have been captured by transposable elements. DNA methylation signatures may be helpful as a filter to identify annotations that reflect either cryptic information or non-functional gene-like elements as opposed to genes that have potential for expression in tissue types or environments not sampled. In other words, DNA methylation may reveal which genes have the potential for expression. This approach has been successfully used for the annotation of sorghum gene models (Olson et al., 2014). In most plants species, there is very little to no CHG methylation in the regions immediately surrounding the TSS of expressed genes (Niederhuth et al., 2016), and genes that are expressed in other tissues also have little or no CHG methylation in these regions (Li et al., 2015). In contrast, genes that are not detected as expressed in any of the sampled tissues include some with higher levels of CHG methylation (Li et al., 2015). The vast

majority of these genes with high levels of DNA methylation are located in non-syntenic genomic positions, and many of them are actually located within an annotated transposon.

The presence of CHG DNA methylation near the annotated TSS could be useful as a filter to flag gene models that are likely silenced in that inbred line but may not indicate that the gene is silenced in all genotypes. A subset of these genes show variable methylation and expression levels between different inbreds. These could reflect examples of cryptic information that is expressed in some lines and not others, potentially due to epigenetic variation. Likewise, genes that are highly methylated are unlikely to provide functional elements, although they may represent cryptic information that could be utilized if DNA methylation patterns are perturbed (Cortijo et al., 2014).

In short, we present our views on potential uses and advantages of using DNA methylation data to further the understanding of functional elements of plant genomes. The identification of regions lacking methylation provides the opportunity to rapidly distil the genome down to *cis*-regulatory regions and genes with expression potential.

FUNDING

P.A.C. and N.M.S. are supported by a grant from NSF (IOS-1802848), J.M.N. is supported by a Hatch grant from the Minnesota Agricultural Experiment Station (MIN 71-068) and S.N.A. and N.M.S. are supported by a grant from USDA-NIFA (2016-67013-24747).

*Peter A. Crisp, Jaclyn M. Noshay, Sarah N. Anderson and Nathan M. Springer**

Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, MN 55108, USA

***Correspondence: Nathan M. Springer (springer@umn.edu)**

https://doi.org/10.1016/j.molp.2019.02.006

REFERENCES

Bouyer, D., Kramdi, A., Kassam, M., Heese, M., Schnittger, A., Roudier, F., and Colot, V. (2017). DNA methylation dynamics during early plant life. Genome Biol. **18**:179.

Cortijo, S., Wardenaar, R., Colome-Tatche, M., Gilly, A., Etcheverry, M., Labadie, K., Caillieux, E., Hospital, F., Aury, J.M., Wincker, P., et al. (2014). Mapping the epigenetic basis of complex traits. Science **343**:1145–1148.

Eichten, S.R., Vaughn, M.W., Hermanson, P.J., and Springer, N.M. (2013). Variation in DNA methylation patterns is more common among maize inbreds than among tissues. Plant Genome **6**. https://doi.org/10.3835/plantgenome2012.06.0009.

Kawakatsu, T., Stuart, T., Valdes, M., Breakfield, N., Schmitz, R.J., Nery, J.R., Urich, M.A., Han, X., Lister, R., Benfey, P.N., et al. (2016). Unique cell-type-specific patterns of DNA methylation in the root meristem. Nat. Plants **2**:16058.

Li, Q., Song, J., West, P.T., Zynda, G., Eichten, S.R., Vaughn, M.W., and Springer, N.M. (2015). Examining the causes and consequences of context-specific differential DNA methylation in maize. Plant Physiol. **168**:1262–1274.

Mei, W., Stetter, M.G., Gates, D.J., Stitzer, M.C., and Ross-Ibarra, J. (2018). Adaptation in plant genomes: bigger is different. Am. J. Bot. **105**:16–19.

Niederhuth, C.E., and Schmitz, R.J. (2017). Putting DNA methylation in context: from genomes to gene expression in plants. Biochim. Biophys. Acta **1860**:149–156.

Niederhuth, C.E., Bewick, A.J., Ji, L., Alabady, M.S., Kim, K.D., Page, J.T., Li, Q., Rohr, N.A., Rambani, A., Burke, J.M., et al. (2016). Widespread natural variation of DNA methylation within angiosperms. Genome Biol. **17**:194.

Oka, R., Zicola, J., Weber, B., Anderson, S.N., Hodgman, C., Gent, J.I., Wesselink, J.-J., Springer, N.M., Hoefsloot, H.C.J., Turck, F., et al. (2017). Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. Genome Biol. **18**:137.

Olson, A., Klein, R.R., Dugas, D.V., Lu, Z., Regulski, M., Klein, P.E., and Ware, D. (2014). Expanding and vetting sorghum bicolor gene annotations through transcriptome and methylome sequencing. Plant Genome **7**. https://doi.org/10.3835/plantgenome2013.08.0025.

Rodgers-Melnick, E., Vera, D.L., Bass, H.W., and Buckler, E.S. (2016). Open chromatin reveals the functional maize genome. Proc. Natl. Acad. Sci. U S A **113**:E3177–E3184.

Schmitz, R.J., Schultz, M.D., Urich, M.A., Nery, J.R., Pelizzola, M., Libiger, O., Alix, A., McCosh, R.B., Chen, H., Schork, N.J., et al. (2013). Patterns of population epigenomic diversity. Nature **495**:193–198.

Sijacic, P., Bajic, M., McKinney, E.C., Meagher, R.B., and Deal, R.B. (2018). Changes in chromatin accessibility between *Arabidopsis* stem cells and mesophyll cells illuminate cell type-specific transcription factor networks. Plant J. **94**:215–231.

Weber, B., Zicola, J., Oka, R., and Stam, M. (2016). Plant enhancers: a call for discovery. Trends Plant Sci. **21**:974–987.

West, P.T., Li, Q., Ji, L., Eichten, S.R., Song, J., Vaughn, M.W., Schmitz, R.J., and Springer, N.M. (2014). Genomic distribution of H3K9me2 and DNA methylation in a maize genome. PLoS One **9**:e105267.