RESEARCH-PGR:Tools for epigenome engineering crops to uncover roles for epigenetic natural variation

PROJECT DESCRIPTION

SPECIFIC AIMS

Natural variation for gene expression levels plays a critical role in controlling plant traits important for agricultural productivity (1-3). Studies of many plant genomes have revealed that transposable elements and DNA methylation are frequently associated with natural variation for gene expression (4-8). Over 80% of the maize genome is derived from transposable elements (TEs) and the specific types and patterns of TEs vary substantially among haplotypes (9-11). There is also evidence for widespread variation in DNA methylation levels in low-copy regions of the maize genome that is often associated with changes in the presence of nearby TEs (5, 12-13). The proposed research project would utilize newly available genomic resources in maize to provide an unprecedented view of the connections between TEs, DNA methylation and gene expression in a crop plant. In addition, the proposed research would develop tools to enable the manipulation of these relationships with an eye towards developing new approaches for engineering crop epigenomes. The research would be fully integrated with activities aimed at training the next generation of scientists while emphasizing the ability to communicate with the general public.

Aim 1. Document natural variation in DNA methylation, transposons and gene expression [JRI, KLH, NMS, MV]. The availability of an improved maize B73 reference genome along with a novel structure-based annotation of TEs in this genome will allow for a substantially improved analysis of DNA methylation relative to transposons and gene expression levels in the reference genotype. In addition, *de novo* genome assemblies are now available for several additional maize genotypes, providing an opportunity to compare DNA methylation, transposon content and gene expression in multiple lines. The primary goal of this aim will be to document examples of variation in gene expression among diverse lines of maize that can be associated with transposon and/or DNA methylation variation. This aim will involve annotation of TE content in multiple maize genomes and analysis of DNA methylation patterns and gene expression. In addition, we will use resources available from existing mapping populations to quantify the role of transposons and DNA methylation in controlling gene expression and plant phenotype.

Aim 2. Determine the role of DNA methylation in controlling gene and TE expression in maize [CNS, RJS, NMS]. We would like to know which genes and TEs in the maize genome are controlled by DNA methylation to understand more about the potential roles for DNA methylation in controlling phenotypes and inheritance in maize. To date, forward- and reverse-genetics efforts have not been successful in recovering viable plants with strong reductions in DNA methylation, likely due to defects in proper development. We propose to develop a high-throughput protoplast transformation system to enable the creation of cell lines with mutations in key DNA methyltransferase genes in maize. These cell lines would be used for profiling of gene/TE expression and DNA methylation to document which genes/TEs are controlled by DNA methylation in a single tissue type of maize.

Aim 3. Develop tools for epigenome engineering in crop plants [RJS, CNS, NMS].
While aims 1 and 2 will develop detailed associations between DNA methylation, TEs and gene expression, they lack the ability to fully document cause and effect relationships between DNA methylation and expression of genes or transposons. We will develop tools such as CRISPR-dCAS9 fusions that would target methylation or demethylation machinery to specific loci to enable precise engineering of DNA methylation levels in maize. We will apply these tools to clarify the causal role of DNA methylation and transposons in influencing gene expression and phenotype.

Broader impacts activities: Developing science communication skills and data/software literacy [KH/KH]. We will have focused efforts to improve the science communication skills for all members of our

project. In addition, we will develop online resources and workshops to provide science communication skills to our scientific community. Our team will also partner with the Data Carpentry and Software Carpentry initiatives to develop tools for epigenomic analysis and to provide open access to these tools.

RELEVANCE AND JUSTIFICATION

Gene expression levels and patterns are critical for controlling phenotypes. The genomes of maize and many other crops consist of islands of genes interspersed with TEs. Research in model species has revealed distinct types of chromatin at genes and TEs. A key question is how crop species manage to maintain proper regulation of gene expression despite the close proximity to TEs with silenced chromatin. In particular, the diversity of TE content within species highlights the potential for TEs to generate expression diversity through genetic or epigenetic mechanisms. The proposed research would provide an unprecedented view of the relationship between TEs, DNA methylation and gene expression. In addition, this project would develop tools and resources to enable the targeted manipulation of DNA methylation at specific loci in crop plants.

The proposed research fits well with several of the PGRP areas for research and tool development. Our project fits within the "*Genotype to Phenotype*" research area as we would be documenting the role of genetic variation for TEs and epigenetic variation for DNA methylation in influencing gene expression and phenotype. We would also be addressing the "*Genomes by Epigenomes by Environment Interactions*" research theme in the proposed research. While our research would be more focused on the genomes by epigenomes interaction we would develop key data resources that could be extended to environmental interactions in future studies and our evaluation of the impact of TEs and methylation on phenotype will take advantage of mapping populations grown in multiple environments. The proposed research would also develop "*Breakthrough Technologies*" for high-throughput CRISPR application in cell cultures and for targeted epigenome engineering. These technologies could have widespread applicability to address a number of basic and applied research questions beyond the specific uses in our proposed research project. The successful completion of our project would also involve work in the areas of "*Data Challenges*" and "*Visualization*". In order to compare multiple de novo genome assemblies we will face challenges in how to anchor the comparisons between distinct genomes. This will provide significant challenges in terms of data analysis and visualization and the methods that are developed for this work will be made available through Cyverse or other sources.

RESULTS FROM PRIOR WORK

NSF-PGRP IOS-1237931 "Causes and consequences of epigenetic variation in maize." $2,724,265. Dates: Oct 1 2012 - Sept 30 2016, No-cost extension through Sept 30 2017. PI: **Springer**; CoPIs **Vaughn**, Lisch, Makarevitch. *Intellectual merit:* This focused on developing tools for characterizing genome-wide DNA methylation patterns in maize and applying these tools to study the factors that influence DNA methylation. We developed robust approaches for generating and analyzing whole-genome methylation data for maize, a crop species with a large, complex genome. We also developed tools for performing sequence-capture bisulfite sequencing to reduce costs and enable population-level scans of DNA methylation. These tools were successfully applied to document changes in methylation in populations of maize, plants derived from doubled haploids, teosinte and maize plants subjected to abiotic stress. To date **13** publications have resulted from this project (5, 13-25) and the data produced by this project has been made publicly available through NCBI SRA. *Broader impacts:* The project has supported the training of **9** undergraduates**, 4** graduate students and **6** postdoctoral researchers including 4 from underrepresented groups. In addition, we have developed tools and data resources that are being widely used by the research community. The results of this work have been important in developing

proper frameworks to consider how to include epigenetic information in public and private crop breeding programs. The project has also supported outreach efforts to K-12 students and the general public.

NSF-PGRP MCB-1339194 "Comparative-, Functional-, and Epi-Genomics of Legumes and Nodule Formation." $2,530,455. Dates: Nov 1 2013 - Oct 31 2017. PI: Jackson; Co-PIs **Schmitz**, Libault. *Intellectual merit:* This project develops genomic resources across three legume species (*Glycine max, Phaseolus vulgaris* and *Medicago truncatula*). This project will reveal natural variation in DNA methylation and identify regions targeted by the RNA-directed DNA methylation-silencing pathway. This proposal also is studying single-cell type epigenomes of root hairs and on cell types sorted from nodules. Over **11** publications have resulted from this support (14, 26-36). *Broader impacts:* Two graduate students and two hosted undergraduate interns from URMs have been trained to learn experimental/computational biology. PI Schmitz also co-organized and led a workshop/career development panel in New Mexico to URM undergraduates from around the state.

NSF-PGRP IOS-1238014 "Biology of Rare Alleles in Maize and Its Wild Relatives" $13,311,185 ($3.2M to **Ross-Ibarra**). Dates: May 15 2013 - April 15 2018. PI Edward Buckler, co- PIs J. Doebley, J. Holland, S. Flint-Garcia, Q. Sun, S. Mitchell, J. Ross-Ibarra. *Intellectual merit* This project has developed accurate imputation approaches, found evidence for the importance of deleterious variants and non-genic polymorphisms in heterosis and GWAS, documented differences in recombination among the parents of the NAM population, developed methods to annotate transposable elements in plant genomes, demonstrated the utility of *a priori* identification of deleterious alleles in genomic prediction, and found population genetic evidence suggesting the importance of demography and purifying selection across the genome. The grant has so far produced >**30** publications including references (19, 37-48) from CoPI Ross-Ibarra. *Broader impacts* The project has included **12** postdoctoral and **12** graduate trainees, including **4** former trainees of CoPI Ross-Ibarra all of whom have gone on to careers in science in industry, USDA, or as faculty. The grant also funded GBS workshops and a traveling maize exhibit as well as an on-line maize evolution resource for teachers.

NSF-PGRP IOS-1546708 "Meeting on Transformation-Enabled Genomic Research in Crop Plants" $68,583 PI Fredy Altpeter, coPIs **Stewart, Springer**. *Intellectual merit:* The crop genetics community has serious bottlenecks related to the production of transgenic plants. The costs are quite high and there is limited capacity in the community. This meeting brought together a group of plant transformation experts from academia and industry as well as plant genetics researchers that utilize transformed plants. A perspective paper summarizing the current state of plant transformation and needs for the community was published (49). *Broader impacts:* This effort was focused on providing information to help the funding agencies and research community identify key opportunities for improving plant transformation capacity and efficiency. A white paper was provided to federal agencies highlighting these opportunities.

coPIs **Happe** and **Hertweck** do not have prior NSF support.

BACKGROUND AND PRELIMINARY DATA

The proposed research would focus on characterizing the connections between DNA methylation, TEs and gene expression in maize. In contrast to *Arabidopsis*, maize has a large genome with abundant TEs interspersed with genes. While work in *Arabidopsis* has highlighted the potential for transposons and DNA methylation to contribute to epigenetic diversity, we might expect an even greater role in maize. The increase in heterochromatin, TE content (9) and differences in patterning of the methylome between maize and *Arabidopsis* (22, 24) highlight the importance of performing these studies in maize. Here, we highlight background information and preliminary data from our groups on DNA methylation and TEs and

describe preliminary work on developing high-throughput transformation systems and methods for epigenome engineering.

Mechanisms controlling DNA methylation (maize genes) and lessons from mutant alleles: DNA methylation refers to the methylation of the 5' carbon of cytosine bases in many eukaryotic genomes. There are several different classes of enzymes that utilize distinct pathways for adding DNA methylation and they tend to act at slightly different contexts. The main pathway for maintenance of CG methylation involves the MET1 gene in *Arabidopsis* (reviewed by 50-51). CG sites are symmetrical across the two strands and hemi-methylated sites resulting from DNA replication are the preferred targets for MET1, providing a mechanism for inheritance of CG methylation patterns following DNA replication. The CMT3 enzyme in *Arabidopsis* and relatives in other species are responsible for the bulk of CHG methylation. CHH methylation at heterochromatin/euchromatin boundaries is maintained by the complex RNA-directed DNA methylation (RdDM) pathway which involves the DRM methyltransferase (52). CHH methylation within large heterochromatic regions is maintained by CMT2 in *Arabidopsis* (53). The chromatin remodeling gene DDM1 also plays a critical role in maintaining DNA methylation in the *Arabidopsis* genome (50). Maize contains orthologs for each of these genes except CMT2 (16, 53). Loss-of-function mutations have been isolated for several maize methyltransferase genes and some of these exhibit partial reductions in DNA methylation in specific contexts (16). In addition, forward genetic screens for factors required for paramutation or transgene silencing have recovered mutations affecting the RdDM pathway and many of these mutations affect CHH methylation at specific loci. A number of genes in this pathway have retained paralogs from the maize whole-genome duplication event, and it seems likely that these paralogs have partially redundant function (16). Segregation analysis suggests that double mutant gametes are viable but there is evidence for early abortion of developing seeds, consistent with RNAi experiments that were able to reduce DNA methylation in cell cultures but could not be regenerated into plantlets (16). Similar results in rice (54) suggest that many crop species will not tolerate massive reductions in DNA methylation without severe phenotypic consequences.

Natural variation for DNA methylation: Rapid technology improvements in DNA sequencing have enabled the description of many plant methylomes. Recent work from the Schmitz group documented similarities and variation in the methylomes of 34 plants species (36). The 1001 epigenomes project in *Arabidopsis* documented natural variation for DNA methylation in a very large set of *Arabidopsis* ecotypes (55). While we continue to learn about natural variation in DNA methylation and its inheritance in a variety of plant species we will focus our discussion on recent results in maize. The methylome of maize is quite distinct from that of *Arabidopsis* (22, 24) and the massive increase in TE content (9) provides additional opportunities for TE induced natural variation (12). The first genome-wide comparison of DNA methylation between maize inbreds (using meDIP-microarray approaches) revealed many loci with quantitative differences in DNA methylation (56). Further analysis of additional inbreds revealed thousands of differentially methylated regions (DMRs) (5). Many of these DMRs appeared to be stably inherited in biparental populations and to be significantly associated with nearby SNPs (5, 15) or TE insertions (12), though at least some evidence suggested that DMRs can occur in the absence of any nearby genetic changes (56). Regulski et al. (57) provided the first context-specific analysis of variation in DNA methylation between B73 and Mo17 using whole-genome bisulfite sequencing (WGBS). A similar analysis provided context-specific information on DNA methylation in 5 diverse inbreds (17). In maize, DMRs frequently involve differences in both CG and CHG methylation, with fewer involving only CG or CHG sites. CHH methylation tends to be found at lower levels in the maize genome and is found in fewer DMRs. These studies have provided many examples of altered levels of DNA methylation, including some that could be associated with polymorphic TEs (12) or altered levels of gene expression (5,17). To date, however, generalizing these results to the entire genome and to other inbred lines has been limited

by the quality of the genome assembly, the quality of TE annotations and the lack of de novo assemblies for other inbreds.

TE: New annotation methods and evidence for polymorphisms: Although TEs comprise the majority of DNA in the maize genome (9, 58), previous understanding has focused on consensus copies of elements found at high copy numbers, and inference was extended to the whole genome by searching for similar sequences using RepeatMasker (59). While these homology-based approaches identify most repeated bases, they often fragment TE annotations and split intact TE insertions into multiple pieces, often identified as belonging to distinct TE families. To understand the impact TEs have on local patterns of gene expression and methylation, identification of full-length TE copies is essential to address their polymorphism among genotypes and identify their proximity to variant genes and DMRs.

The Ross-Ibarra group has developed an alternative, *de novo* approach that does not rely on prior information from known TE copies. Instead, the method takes advantage of the biology of transposition and knowledge of TE structure to identify both intact full-length TEs as well as fragments (e.g. solo LTRs) and nested insertions. Structural features such as the long terminal repeat of some retrotransposons, target site duplications (TSDs), and features necessary for the initiation and progression of reverse transcription like primer binding sites and polypurine tracts occur at recognizable distances from each other and can be identified in primary sequence data without pre-specification of TE homology. This structural approach comprehensively recovers intact TEs that have landed at a particular location in the genome, and as such is able to identify TE copies that can be used to answer questions about the functional relevance of individual insertions. Comparison of this novel automated approach with hand-annotated BACs suggests a high degree of accuracy (Figure 1), and this annotation has been incorporated into the publicly released v4 assembly of the maize genome (60). In the B73 genome, this annotation joins fragments previously attributed to heterogeneous TEs, resulting in ~180,000 copies which can be assayed for presence/absence in other genotypes.

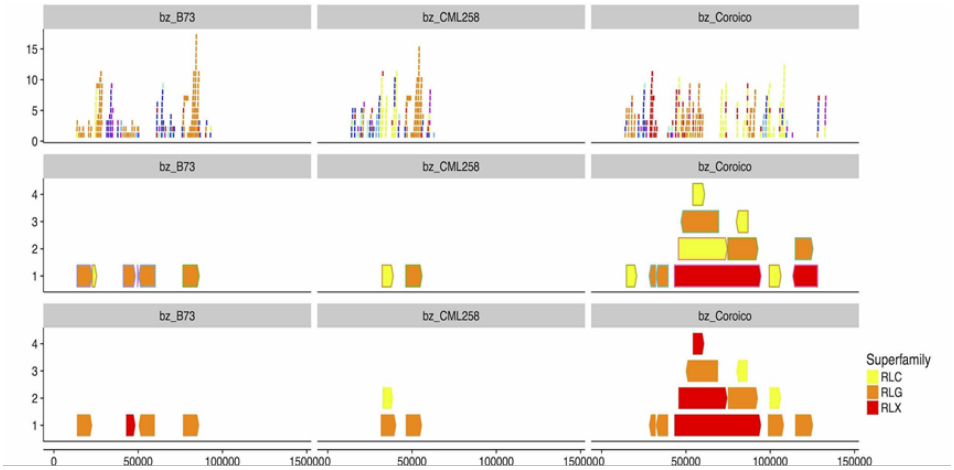

Figure 1: Comparison of TE annotation approaches of BACs surrounding the *bz1* locus of maize. Columns represent four maize genotypes, and rows represent automated repeat-masking (top), hand-annotation (Wang et al. 2006) (middle), and our automated structure-based annotation pipeline (bottom). TEs are colored by superfamily identification: Copia (RLC), Gypsy (RLG), or unknown (RLX).

Our current understanding of the contribution of intact TEs to variation among maize genotypes comes from haplotype differences in TE content in limited genomic regions (10-11, 61). Work from the Ross-Ibarra group has extended these inferences genome-wide using the structural annotation of the B73 v4 genome and applying approaches developed for identifying new TE insertions in non-reference genotypes (62-63). Preliminary analysis of short-read sequence from the inbred line Oh43, for example, has identified thousands of TE insertions not present in the reference B73 genome.

Maize protoplast transformation system: description and evidence of feasibility: The proposed research would utilize transformed cell lines of maize to study the roles of DNA methylation and to develop epigenome editing tools. Recently, an automated plant protoplast extraction and transfection/transformation platform has been developed in CoPI Stewart's lab that uses a 96-well plate platform for robotic plant single cell assays (64). The system uses inexpensive enzymes to digest cell walls (65) and the robot distributes cells into plates, where the time to transfection is as little as 4 hours (64). A number of species (tobacco, switchgrass, maize, soybean) and tissues (suspension cell cultures, callus, intact leaf segments) have been successfully processed on the robot. Transformed cells can be analyzed using an integrated plate reader for fluorescence. The plate reader is tunable for custom excitation and emission, which enables multiple fluorescent proteins (FP) to be used and read simultaneously. For example, we have recently used a dual FP plasmid (containing constructs producing green and orange FPs) for screening dual promoters in a relatively recalcitrant species: *Arundo donax* (66). In this system, one FP is driven by a constitutive promoter to act as an internal standard for normalizing fluorescence while a second FP is used as a readout for promoter strength. A similar system (we now have 9 FPs that can be used) can be employed in the maize to standardize selection and transformation. After screening for transformation, cells can be sorted, proliferated, and further analyzed for the traits of interest.

Approaches for epigenome engineering and preliminary data from Arabidopsis: One thrust of the proposed research is to develop methods to site-specifically engineer DNA methylation states in plant genomes to identify causal relationships between differential DNA methylation and traits of interest. An ideal platform would allow easy targeting to diverse sites within plant genomes and effective addition or removal of DNA methylation at these sites. In addition, ideal systems might result in newly engineered methylation states that are inherited stably over generations, enabling the development of novel allelic variation. Directed DNA methylation has been achieved in plants using a zinc finger as a sequence-specific guide fused with SUVH2 (67) which recruits RdDM activity. However, the ability to quickly target zinc-finger systems to diverse sequences is limited. CRISPR/dCAS9 fusion systems could allow more efficient targeting of proteins or biochemical activities to genomic regions. This has recently been achieved in animals by fusing sequence-specific guides such as TAL/dCAS9 to enzymes such as DNMT3A/TET1 that catalyze the addition/removal of DNA methylation (68), but there are no published reports on the use of CRISPR for targeting DNA methylation or demethylation activities in plants.

The enzymes/mechanisms required for establishment and maintenance of DNA methylation share similarities between plants and mammals. However, mechanisms for DNA demethylation are more diverse. The major mode of DNA demethylation in plants utilizes enzymes involved in base-excision repair such as ROS1 and DEMETER. However, DNA demethylation in mammals utilizes the family of Tet methylcytosine dioxygenases, which hydroxylate 5-methylcytosine into 5-hydroxymethylcytosine and subsequently into 5-formylcytosine and 5-carboxylcytosine, which is then removed by base-excision repair. Tet enzymes are not found in plant genomes, but preliminary data from CoPI Schmitz demonstrates that expression of Tet1 from humans in *Arabidopsis thaliana* leads to widespread loss of DNA methylation (Figure 2). Therefore, the fusion of Tet1 to sequence-specific guides could provide a powerful tool to site-specifically engineer DNA methylation states in plant genomes. Based on these preliminary results we will plan to over-express Tet in maize cell lines to reduce DNA methylation and will also use fusions of Tet or maize ROS1 orthologs in an attempt to perform targeted demethylation.

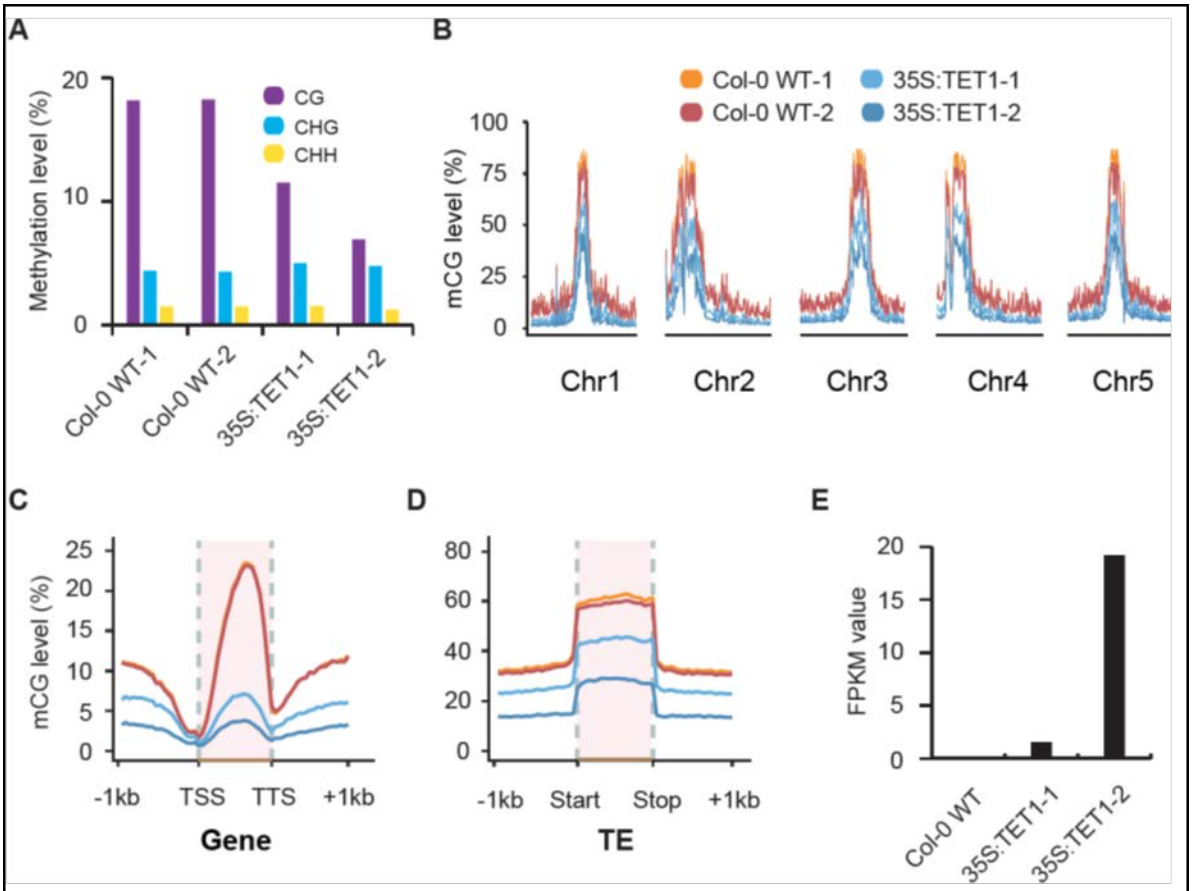

Figure 2. Overexpression of TET1 induced global CG demethylation in *A. thaliana*. (A) Bar graph of global methylation levels in two Col-0 WT replicates and two 35S:TET1 transgenic individuals. (B) Metaplot of CG methylation levels (100 kb windows) across five *A. thaliana* chromosomes. Metagene plots of CG methylation level across (C) gene bodies and (D) transposable elements. (E) Expression level (FPKM) of *TET1* in Col-0 WT and 35S:TET1 transgenic plants.

RESEARCH PLAN

The proposed research will involve collaborative efforts focused on several themes. The first aim will focus on developing a comprehensive understanding of the relationship between natural variation in DNA methylation, TE content, gene expression, and phenotype using de novo genome assemblies and existing mapping populations. The second aim will document the role of DNA methylation in controlling gene and TE expression in a cell culture system. These two aims will develop a number of hypotheses about the roles of DNA methylation in controlling the expression of specific alleles or specific TE insertions. The third aim will develop systems to test these hypotheses through targeted modification of the maize epigenome.

Aim 1. Document natural variation in DNA methylation, transposons and gene expression [JRI, KLH, NMS, MV]. As described in the preliminary data, the Ross-Ibarra group has developed approaches for performing structural annotations of TEs in the maize genome and has applied these methods to the recently improved B73 reference genome (60). The Ross-Ibarra and Springer groups are currently preparing a manuscript that will describe genomic distributions, evolutionary history and functional

characteristics such as expression and chromatin state for each family of TEs present in B73. The Springer and Vaughn groups have developed approaches for whole-genome methylation profiling and have generated profiles for over 20 inbred lines of maize. These approaches will be combined to provide a detailed analysis of variation in a set of maize inbreds with de novo genome assemblies. We will then utilize additional populations of maize with existing genomic resources to test the associations between DNA methylation, TE content, gene expression, and phenotype.

1.1 TE discovery and impacts across de novo assembled maize genomes. Several groups are preparing de novo genome assemblies (http://www.maizegdb.org/genome/assemblies_overview for the current list) for additional maize genotypes. Some of these, such as PH207 (69) and CML247 (70) have been published. Others (B104, W22, Mo17) are publicly available under limited use restrictions until publication or have been provided to our group (TIL1, TIL11 - Hufford letter), and several additional genomes still being assembled will be shared when available (PHJ89 - Kaeppler letter). In addition, it is widely anticipated that additional genotypes will have assemblies augmented with long-read technologies within the coming 1-2 years. The Ross-Ibarra group will perform structural annotation of TEs in each of these genomes using the pipelines described in Ware et al. (60). Variation in TE content at specific loci will be documented with particular attention to variation in which TEs are located closest to the same gene in different haplotypes.

The Springer and Vaughn groups will use methods as described in Li et al. (17) to generate ~20X coverage methylomes for each of these genotypes using DNA collected from the third leaf (this has already been completed for B73, Mo17, PH207 and W22). Much of our analysis of DNA methylation focuses on comparing methylation in 100bp tiles across the genome. We will develop tools for cross-referencing 100bp tiles between different assemblies in order to allow for discovery of DMRs among each of these genotypes. RNAseq data (~15M reads for 3 biological replicates) will be collected for four different tissues (control seedling, heat-stressed seedling, embryo, endosperm) for each of eight genotypes (B73, PH207, W22, CML247, Mo17, B104, PHJ89, TIL1, or other genotypes with high-quality assemblies that are released in the coming year) to comprehensively document variation in gene expression levels in the inbred lines. We will also perform RNAseq to analyze allele-specific expression in the seven F1 genotypes generated from crossing each genotype to B73 to confirm that changes in gene expression are due to cis- rather than trans-acting differences (71). In addition, we will use approaches recently developed in PI Springer's group to quantify the expression of each TE family (using both uniquely and multiple-mapped reads) in all genotypes.

A coordinate analysis of the three datasets (TE content in each genome; expression levels of genes and DNA methylation levels) will be performed to assess the relative roles of TE and DNA methylation in gene expression variation. A primary goal is to identify (epi)haplotype variation in maize that drives changes in gene expression. We would like to identify and characterize examples for three different types of variation: (1) gene expression differences associated with nearby DNA methylation changes but no TE polymorphisms, (2) gene expression differences associated with nearby TE polymorphisms without changes in DNA methylation and (3) gene expression differences associated with both TE polymorphism and altered DNA methylation. Preliminary work (17) focusing on DNA methylation and gene expression in five maize inbreds identified numerous examples of genes that exhibit altered promoter methylation and altered gene expression (heavily methylated allele was silent or lowly expressed), but could not assess TE polymorphisms. Our targeted analysis of the eight maize genotypes with de novo genome assemblies will be used to document examples of gene expression variation that can be associated with either DNA methylation or TE variation.

1.2 Assess the connections between TEs or DNA methylation and gene expression in maize association panels or mapping populations. The availability of de novo genome assemblies will provide opportunities to document TE variation that can be difficult to detect solely through short-read resequencing. However, once TE polymorphisms have been documented among these eight genomes, existing short read data can be used to determine the presence or absence of a particular insertion in other genotypes. There are several populations of maize that have substantial resources for genotyping and expression. The NAM population (72) includes 26 diverse parents as well as 5,000 RILs. The 282 population (73) includes 282 diverse maize inbreds and the WiDiv population (74) includes over 500 maize inbreds that can grow to maturity in the upper midwest. For each of these populations the parents have existing whole-genome shotgun short read data that can be used to determine the presence or absence of TE insertions identified in the de novo assembled genomes. In addition, there is existing RNAseq data for each of these populations (75), which will allow for an evaluation of the association between the presence of a TE and expression levels of nearby genes. By independently evaluating the relationship between TE presence and gene expression in multiple panels we can identify robust associations that are replicated in multiple populations.

We also wish to assess the relationship between presence or absence of a TE and DNA methylation in nearby regions. We currently have WGBS data on 27 maize inbreds including 12 of the NAM parents and additional lines from the 282 panel. For this project we will perform WGBS on an additional 40 genotypes including the remaining NAM parents as well as a subset of lines that are represented in both the 282 and WiDiv populations. The selected lines will include genotypes that represent diverse germplasm or key founders in breeding programs. In total we will have WGBS for over 60 maize genotypes which will be used to identify DMRs. We will focus on comparing the relationship between DNA methylation levels in haplotypes containing a TE insertion relative to haplotypes without the insertion to assess the potential impact of TEs on methylation of nearby regions. We will also perform a comparison of the DNA methylation levels and gene expression levels in these genotypes. While our focused analysis of the 8 genotypes with de novo assemblies can provide important clues to the association between TEs or methylation and gene expression, the in-depth analysis of 60 diverse genotypes will provide a much more powerful analysis of these associations.

An additional experiment will be performed to assess whether the nearby TEs or DMRs are associated with expression of nearby genes. The NAM population (76) provides a very large population with multiple crossovers per gene. We will search for lines in this population that contain crossovers that occur between a gene of interest and a nearby DMR or polymorphic TE. Lines containing these crossovers will be used for targeted analysis of DNA methylation, TE content and DNA methylation, providing additional support for a role for variation in the region containing the TE / DMR in influencing gene expression.

1.3 TE impacts on phenotype. While activities 1 and 2 allow investigation of the impacts of TE activity on patterns of methylation and gene expression, ultimately we are interested in understanding the importance of each of these on phenotype. Maize is an excellent system for identifying phenotypic associations thanks to the existence of a number of extant mapping populations with both genotype (whole genome sequence of individuals or founders) and phenotype. We will take advantage of phenotypic data in the NAM (72), WiDiv (74-75), PVPdiallel (48), 282 (73), Magic (77), and an unpublished set of populations from CoPI Ross-Ibarra and colleagues that includes 200 sequenced landrace/teosinte parents and 20,000 genotyped and phenotyped progeny. In each population we will use two approaches to associate TEs to phenotype. First, we use short-read sequencing data from the parents to identify TE presence/absence polymorphism, then impute these polymorphisms to the progeny to perform a standard genome-wide association (GWA). This will allow rapid identification of TE

insertions with large effect on phenotype. Second, we will use simple mixed models to partition phenotypic variance to different subsets of the genome (78). While this latter approach does not allow identification of individual TEs or DMRs that are responsible for phenotypic change, it provides a comprehensive assessment of the explanatory power of particular types of polymorphism. Recent applications of this approach to maize have demonstrated that open chromatin -- as measured by MNase hypersensitive sites -- is profoundly enriched for functional variation (79) and that maize subgenomes differ in their functional relevance across dozens of phenotypes (80). For some of these populations (NAM, 282), additional expression and methylation data will allow us to investigate the impacts of individual insertions associated to phenotype as well as refine genomic subsets to be tested. A set of the most promising associations from these analyses will be validated experimentally by growing RILs from these populations with and without the insertions of interest in multiple different environments and document the impact of the insertion on expression, methylation, and phenotype.

Pitfalls and alternative approaches: The majority of work in this aim has been successfully performed on B73 or using other traits. The TE annotation pipeline can easily be applied to other genotypes. While the quality of the annotation will depend on the quality of these additional *de novo* assemblies, TE insertions in genic regions of the genome are both the easiest to assemble and those of greatest interest, suggesting that genome quality will not unduly limit our investigation. We have substantial experience in generating and analyzing WGBS and RNAseq data and do not anticipate difficulties in generating these datasets. We recognize that the repetitive nature of many maize TEs can result in difficulties in achieving unique alignments. For DNA methylation we are primarily interested in how the TEs affect methylation of flanking sequences which are usually lower copy and therefore can be assessed. For expression, we have developed approaches to estimate expression per TE family rather than per locus, which permits the use of multiple-mapping reads. There will be some complications in developing direct comparisons of expression or methylation between genotypes due to differences in gene content among maize genotypes. However, we have experience working on these comparisons with B73 and PH207 and expect to have tools that will allow for robust comparisons for much of the genome along with targeted analysis of genotype-specific segments in each line. Finally, the mixed model approaches for quantifying phenotypic impact are likely underpowered, providing only a conservative estimate of associations between our molecular traits and phenotypic traits. Given our preliminary data and expertise we do not anticipate major issues in this aim and expect to provide the most comprehensive analysis of variation in DNA methylation, TEs, gene expression and traits in a crop species to date.

Aim 2. Develop cell culture lines with reduced methylation to assess the role of DNA methylation in gene and TE regulation in maize [CNS/ RJS/ NMS]. Previous research has suggested that severe perturbation of the maize methylome has major fitness consequences for maize. We have not been able to recover double mutant lines for duplicate copies of the maize CMT3 or DDM1 orthologs (17). In maize many attempts were made to reduce DNA methylation through RNAi targeted silencing of MET1, but while cell cultures with reduced methylation could be recovered, transgenic plantlets expressing these RNAi constructs were never recovered (17), suggesting that major reductions in DNA methylation might be permissible in cell lines but not in developing plants. We propose to study the role of DNA methylation by developing cell lines with compromised DNA methylation machinery. One major concern is that cell cultures in rice and maize exhibit altered methylomes relative to other tissues (81-82). It is worth noting that these studies found that the majority of the methylome is unchanged but there are hundreds of regions with reproducible changes in DNA methylation. Therefore, it will be critical to include proper controls to document the background effects of tissue culture relative to the specific effects of particular mutations. Our hope is that by studying cell cultures with major perturbations to DNA methylation levels we can identify specific genes and TEs that require DNA methylation for silencing in maize. By focusing

on genes or TEs that are normally methylated in B104 but become demethylated and expressed following a perturbation of the DNA methylation machinery we hope to identify the direct targets of DNA methylation in the complex maize genome.

2.1 Optimize a high-throughput protoplast isolation and transformation system for maize and characterize methylome and transcriptomes of control cell lines. The Stewart group has successfully developed a robust protoplast isolation system for switchgrass and has preliminary success implementing a similar system in maize. We plan to focus the efforts in this aim on genotype B104, as this line is known to be culturable and can be used to regenerate plantlets and has a de novo genome assembly (currently available for download at http://www.maizegdb.org/genome/assemblies_overview under Toronto agreement) that will be improved in the near future with PacBio reads. In addition, >90% of the B104 genome is identical-by-descent to the very well-assembled B73 reference (83). This means that for the majority of the genome we will have a very high-quality assembly and annotation. Maize mesophyll protoplasts will be isolated and transformed according to previously established protocols with slight modifications (84-85). Embryogenic callus will be initiated from surface-sterilized immature embryos and subsequent cell suspension cultures will be established as previously described (86). Protoplasts will be isolated using food-grade enzymes (65) and intact protoplasts will be purified from cellular debris using a 23% sucrose gradient. For protoplast transformation, approximately 1x105 protoplasts will be added to a round bottom tube along with 20-30 µg of plasmid DNA and an equal volume of 40% PEG solution. After transformation, the protoplasts will be washed to remove the PEG and will be resuspended in incubation solution. Fluorescent protein tagging will be used in a plate-based detection assay. All steps from protoplast extraction to fluorescent protein analysis will be performed using the transformation robot (64). Transgenic protoplasts that are destined for plant regeneration will be collected by centrifugation and will be regenerated following the methods of Rhodes et al (87). Transgenic cultures, as determined based on a screenable FP marker, will be shipped to team members and used to perform targeted methylation profiling using a sequence-capture bisulfite approach developed by the Springer group (13) to reduce cost ~20-fold relative to WGBS. The current implementation of this approach allows for the profiling of 15Mb of the maize genome with ~10M 100bp PE reads. The 15Mb includes several thousand sites that survey regions of high, low and variable methylation and can be used to document changes in DNA methylation. We will perform targeted profiling of 12 independent cell lines that have been transformed with a selectable marker to assess the stability of DNA methylation at these regions in tissue culture. RNAseq will be conducted on the cell lines to document the robustness of gene and TE expression among independent cell cultures.

2.2 Generate and characterize cell lines with compromised DNA methylation machinery. CRISPR-Cas9 constructs will be developed to target conserved regions of the maize *CMT* orthologs (*Zmet2/5*), the maize *DDM1* orthologs *(Chr101/106*), the maize *DRM* orthologs (*Zmet3/7)* and the maize *MET1* orthologs (tandem duplicates of *Zmet1*). In each case it is possible to develop multiple gRNAs that would have 100% sequence identity to both copies of these genes. We will perform stable transformation with a vector containing the gRNA and Cas9 along with a selectable marker. We recognize that gene editing technologies are rapidly advancing and we will make use of new developments as appropriate for ideal vector construction. For example, new systems enable the use of multiple gRNAs in a single vector (88) and these could improve our ability to recover targeted mutations. 96 independent transgenic cell cultures will be developed on plates from single cell founders for each construct and a small portion of callus will be used to perform FASTmC (89) to identify cultures with altered methylomes. FASTmC is a highly scalable low cost method to screen for genome-wide changes in methylation, as it can be done with ~10,000 sequencing reads, allowing for screening of many lines at once with a quick read-out for functional changes. Cell lines with altered methylation levels will be identified and grown into larger

cultures. We will collect DNA for sequence capture profiling of DNA methylation and RNA for transcriptome profiling from at least six independent events for each of four constructs and from a defective CAS9 (dCAS9) control. We will also perform WGBS for two of the events for each construct to provide a high-resolution view of DNA methylation changes in these loss-of-function lines. The results will be analyzed to document changes in gene or TE expression in lines with strongly reduced DNA methylation and we will make use of the controls from aim 2.1 to make sure that observed changes are specific to loss-of-function lines.

2.3 Attempt to recover plants with reduced DNA methylation levels. While the use of cell culture systems provides an avenue to study the role of DNA methylation in maize it would be ideal to have viable maize plants with severely perturbed DNA methylation as well as having mutant alleles for individual genes in order to study the role of DNA methylation in plant traits. We will select several cell culture lines with moderate to severe reductions in DNA methylation and attempt to regenerate plantlets. It is possible that plants with partial loss-of-function alleles could be viable and provide reductions of 30-70% of genomic methylation levels. If we can successfully recover these plants, we will perform methylation profiling and expression analyses on vegetative tissues to assess whether similar genes are regulated by DNA methylation in callus and other tissues. If this is not successful we will isolate cell cultures that are heterozygous for loss-of-function alleles and regenerate these plants in attempts to generate homozygous materials (embryos or plants) through self-pollination.

Pitfalls and alternative approaches: Aim 2.1 is feasible. The Stewart group has had success in applying their high-throughput protoplast isolation and transformation system to a number of monocots, including maize. We anticipate the need to do some work to optimize the system for the genotypes we will need to use in this project. The molecular characterization of these lines will use techniques that are frequently used in the Springer and Schmitz groups. The activities in aim 2.2 have some potential risk. It will be important to sub-culture some of the transformed events to get genetically homogenous populations of cells that have undergone deletions at all four targeted loci/alleles. This will allow us to perform a robust analysis of alterations in DNA methylation and expression in null lines. There is a potential that cell lines with null alleles will be inviable or have slower growth. We intentionally included the DRM genes (Zmet3/7) as targets given the fact that null mutants in other RdDM components are viable in maize. If we are successful with this construct but not MET, CMT or DDM it will suggest that strong perturbations of DNA methylation result in reduced viability even in cell cultures. This would necessitate the creation of inducible alleles and we could work towards developing this type of allele. We may find that cell cultures with strongly reduced methylation are viable but that they cannot be regenerated into plants in aim 2.3. This would necessitate regenerating heterozygous events or partial loss-of-function alleles.

Aim 3. Develop tools for epigenome engineering [RJS/CNS/NMS]

Although aims 1 and 2 will develop very strong information on potential connections between DNA methylation, transposons and gene expression, they lack the ability to fully document cause and effect relationships between DNA methylation and expression of genes or transposons. An EAGER has been awarded to the Schmitz group to develop the use of DNMT3/Tet1 to add/remove DNA methylation in a targeted manner using yeast and *Arabidopsis thaliana* as model systems. In this aim we will build on the lessons learned from these model systems as we expand the repertoire of enzymes used to develop targeted engineering of DNA methylation levels in maize to further enable our studies. These tools would also be useful for epigenetic recovery of silenced transgenes or other applications in a variety of crop species.

RESEARCH-PGR:Tools for epigenome engineering crops to uncover roles for epigenetic natural variation

3.1 Develop tools for targeted methylation. A CRISPR-dCas9 will be fused to SUVH2 or other RdDM components in an attempt to target RdDM activity in maize. This activity will be focused on cell culture systems and stable transgenic cell lines will be recovered. The maize genes that are homologous to *SUVH2* (GRMZM2G025924) or *DMS3* (GRMZM2G309152) will be fused to dCAS9. The gRNAs will be selected to target promoter regions of genes that are unmethylated in B104 but have high levels of methylation in other haplotypes. We will include examples with and without 24nt sRNAs. Targeted methylation analysis will be used to document DNA methylation at these loci in at least 8 events per construct. For cell lines with novel methylation at the target locus we will also perform qRT-PCR to assess any potential changes in gene expression. We will also perform WGBS in several lines to assess off-target effects. Several lines with successful inductions of locus-specific methylation will be regenerated into plantlets and the inheritance of the acquired methylation will be assessed for one or two generations following segregation of the transgene.

3.2 Develop tools for targeted demethylation. It may be more useful for many epigenome engineering applications to target demethylation activities in an attempt to activate gene expression. Monocots contain orthologs of the DNA glycosylases ROS1 and DML3 that are present in *Arabidopsis* but lack orthologs of DME (90). We plan to develop constructs that target the expressed ROS1 ortholog (GRMZM2G422464) or DML3 ortholog (AC234169.1_FGP003) as well as the mammalian TET1 fused to a dCas9. The gRNA targets will include 5 loci that have high levels of DNA methylation for the B104 allele but low levels of methylation in other haplotypes. Some of the target loci will include nearby TE polymorphisms whereas others will exhibit a DNA methylation change with no nearby sequence change. A similar experimental design will be employed as in activity 1 of this aim. Multiple events will be assessed for each construct with targeted methylation and expression analysis. A small set of successful events will be used for WGBS to evaluate off-target effects and some events will be regenerated to document inheritance of the demethylated allele in the absence of the transgene.

3.3 Apply these tools to perform coordinate demethylation of TE families. One particularly attractive option would be to target demethylation activities to the TIRs or LTRs for a family of transposons. We will identify six families of transposons and develop gRNAs to target the LTRs or TIRs of these families. The families will be selected to include families that are reactivated in cell lines with reduced methylation, exhibit evidence for impacts on expression or phenotype in aim 1 or exhibit evidence for recent movement in maize populations. We anticipate using a combination of multiple gRNAs in a single vector (88) in an attempt to achieve strong recruitment of the demethylase activities to the LTR. This analysis will focus on families with 3-20 elements to allow specific testing of the methylation of multiple edges and potential influences on nearby genes. Multiple independent cell cultures expressing these 6 constructs will be isolated and used to perform Chop-PCR (using restriction enzymes affected by methylation) assays to assess whether the LTRs exhibit reduced methylation. Cell lines with reduced methylation will be selected for characterization and regeneration. We will characterize expression of the TE and nearby genes in cell cultures. In addition, regenerated plants will be used to study inheritance of reduced methylation, TE movement, expression of the TE and nearby genes.

Pitfalls and alternative approaches: This is clearly the highest risk aim in the proposal. We fully recognize that there are a number of technical and biological hurdles to the successful completion of this aim. While prior work in *Arabidopsis* has demonstrated the ability to recruit RdDM activities to a single locus in *Arabidopsis* using a Zinc-finger fusion (67) it could be difficult to adapt this system to CRISPR/dCAS9 targeting and other loci may be recalcitrant. In addition, while targeted demethylation has been described in mammalian systems it has not been described in plants to date. Despite these risks, we argue that it is critical to begin work to develop these tools to enable epigenome engineering approaches in plants. We

feel that our approach (focusing on a high-throughput cell culture system, using rapid molecular assays for screening events and targeting naturally variable loci) provides ideal opportunities for developing this system. We anticipate that during the coming year there will continue to be rapid developments in the basic CRISPR/genome editing technologies and we will take advantage of these. In addition, we anticipate that work in mammalian systems or *Arabidopsis* will highlight important features in the design of constructs or selection of targets and we will use this information to guide or efforts.

BROADER IMPACTS OF THE PROPOSED WORK

The proposed work will consider a range of broader impacts. The development of resources for engineering crop epigenomes will address basic research questions as well as open new avenues for crop improvement. Improving our ability to engineer crop epigenomes could help improve transgene expression or alter traits in a non-genetically modified fashion. The proposed work will include effective training of undergraduates, graduate students and postdoctoral researchers. A complete description of the mentoring plan for these individuals is included in the additional documents (A-5 and PDM). In this section we will highlight several novel activities that will occur in the proposed work with the aim of improving science communication skills and training in computational approaches. These activities will be spearheaded by coPIs Happe and Hertweck but members from all groups will be involved in each of these activities.

Science communication development workshops: The ASPB Decadal Vision ("Unleashing a Decade of Innovation in Plant Science") calls for increased focus on effective science communication. CoPI Happe will develop the science communication skills for individuals within this project to identify appropriate resources to aid in their communication with other scientists, policy-makers, and the general public. To better link research outcomes and broader impacts, the grantees will engage in training and outreach informed by the following principles:

- Science communicators must be able to identify and tailor messages about both the technical and social aspects of plant science to multiple stakeholder audiences, including scientists from related and allied disciplines, policy-makers, academics, and community groups.
- Science communication with the general public is most effective if it employs a two-way, deliberative communication model that allows stakeholders to voice hopes, concerns, and alternative/complementary perspectives on food security. For example, a successful message will be one that articulates the interests of the scientific community with those of the public—interests that are documented by both research about public attitudes as well as direct engagement with community groups.
- Adoption of risk communication practices should emphasize transparency and uncertainty with the aim of gaining the public's trust. Risk communication research shows that when researchers, regulators, corporate actors, and policy-makers are honest and transparent about both known and unknown risks, public trust is enhanced.

The aforementioned principles will ground the following activities:

1. Regular video calls (approx. every 6-8 weeks) between co-PI Dr. Kelly Happe and personnel of grant-funded labs to discuss project research, outreach activities, and science communication best practices. Dr. Happe will evaluate the science communication needs of project personnel and engage in direct science communication training as needed.
2. Organization of an annual science communication workshop at a major plant sciences conference such as PAG. This workshop will train members of the plant sciences community in science communication best-practices including but not limited to audience analysis, analogy, framing,

consensus-building, visual imagery, emotional appeals, and understanding the complexities of scientific controversy such as that surrounding the topic of GMOs.
3. Development of communication materials to support on-going efforts of grantees to communicate the technical aspects of their research to various audiences. These materials will be made available through the project website.
4. Publications on the topic of effective science communication about plant science, including articles on successful science communication pedagogy.
5. Development of a "Plants and Society" course that introduces undergraduate and graduate students to the field of science studies, offers opportunities to discuss the plant sciences from an interdisciplinary perspective, and trains students in effective science communication.

Training in epigenetic and genomic computation: Skills in data management and computational analysis are increasingly important in scientific professions, especially in relation to genomics (91). We will collaborate with Software Carpentry (SWC) and Data Carpentry (DC), two closely related non-profit organizations, to train a broad range of learners in skills related to data-driven research (92). We propose joint organizational membership with SWC/DC for the duration of the grant (silver level in years 1 and 3, bronze in years 2 and 4), which will provide sustained support for the following activities:
1. Hosting SWC and DC workshops. Both organizations have well-established, two-day active-learning workshops taught by trained volunteers. DC workshops focus on data management and basic scripting for learners with little to no prior computational experience, while SWC emphasizes automating repetitive tasks and source control through programming. We will host a total of 12 SWC or DC workshops over the course of the grant at UT Tyler, UMN, and UGA (other grant-affiliated institutions have existing training programs). Given ca. 40 learners per workshop, we will train ca. 480 students, postdocs, and researchers, and plan to advertise to nearby minority-serving institutions and community colleges to broaden workshop participation.
2. Training grant-affiliated participants as certified SWC and DC instructors. Instructors for SWC and DC are trained in evidence-based teaching practices that provide an effective framework for helping people learn better and faster (93). The two years of proposed silver membership includes training for twelve grant-affiliated participants as instructors. Following completion of the training course and checkout procedure administered by SWC, these instructors will be approved to serve as instructors for an SWC/DC workshop at any of the participating institutions, including those described above.
3. Epigenetics workshop development in collaboration with DC. In addition to the joint organizational membership, co-PI Hertweck will partner with DC to develop a two-day epigenetics (DNA methylation) workshop. A multitude of computational tools currently exist for DNA methylation analysis (25). Learning objectives for this workshop will include transferrable data management and reproducible scripting skills that will enable researchers to both use these tools and collaborate on open-source software development projects. Following DC's roadmap for lesson development (94), we will spend year 1 designing the workshop and drafting lesson content. We will host a three-day lesson development hackathon at UT Tyler in year 2. During this hackathon, ca. 17 participants from both the grant, DC, and the broader epigenetics community will test and refine the course materials. Pilot workshops in years 3 and 4 will allow the opportunity for assessment by DC staff and further refinement. After completion of the lesson material, it will be included in DC's list of freely available (CC-BY) lessons (95) and supported for long-term use by other workshops.
4. Curriculum development for undergraduate bioinformatics course. Following design of the DC epigenetics workshop, which teaches general skills for computational analysis, co-PI Hertweck will develop an instructional epigenetics module for her undergraduate bioinformatics class at UT Tyler. This module will emphasize content knowledge about epigenetics by integrating data collected as a part of Aim 1 in an undergraduate course.

REFERENCES

1. Studer A, Zhao Q, Ross-Ibarra J, Doebley J. 2011. Identification of a functional transposon insertion in the maize domestication gene tb1. Nat. Genet. 43: 1160-1163.

2. Yang Q, Li Z, Li W, Ku L, Wang C, Ye J, Li K, Yang N, Li Y, Zhong T, et al. 2013. CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. Proc. Natl. Acad. Sci. U. S. A. 110: 16969-16974.

3. Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashev S, Bruggemann E, et al. 2007. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc. Natl. Acad. Sci. U. S. A. 104: 11376-11381.

4. Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, Libiger O, Alix A, McCosh RB, Chen H, Schork NJ, et al. 2013. Patterns of population epigenomic diversity. Nature 495: 193-198.

5*. Eichten SR, Briskine R, Song J, Li Q, Swanson-Wagner R, Hermanson PJ, Waters AJ, Starr E, West PT, Tiffin P, et al. 2013. Epigenetic and genetic influences on DNA methylation variation in maize populations. Plant Cell 25: 2783-2797.

6. Kawakatsu T, Huang SS, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y, et al. 2016. Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. Cell 166: 492-505.

7. Stuart T, Eichten SR, Cahn J, Karpievitch YV, Borevitz JO, Lister R. 2016. Population scale mapping of transposable elements reveals links to gene regulation and epigenomic variation. Elife 5. pii: e20777.

8. Quadrana L, Bortolini Silveira A, Mayhew GF, LeBlanc C, Martienssen RA, Jeddeloh JA, Colot V. 2016. The Arabidopsis thaliana mobilome and its impact at the species level. Elife.5. pii: e15716.

9. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112-1115.

10. Wang Q,Dooner HK. 2006. Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. Proc. Natl. Acad. Sci. U. S. A. 103: 17644-17649.

11. Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A. 2005. Evolution of DNA Sequence Nonhomologies among Maize Inbreds. Plant Cell 17: 343-360.

12. Eichten SR, Ellis NA, Makarevitch I, Yeh CT, Gent JI, Guo L, McGinnis KM, Zhang X, Schnable PS, Vaughn MW, et al. 2012. Spreading of heterochromatin is limited to specific families of maize retrotransposons. PLoS Genet. 8: e1003127.

13*. Li Q, Suzuki M, Wendt J, Patterson N, Eichten SR, Hermanson PJ, Green D, Jeddeloh J, Richmond T, Rosenbaum H, et al. 2015. Post-conversion targeted capture of modified cytosines in mammalian and plant genomes. Nucleic Acids Res. 43(12):e81.

14*. Eichten SR, Schmitz RJ, Springer NM. 2014. Epigenetics: Beyond Chromatin Modifications and Complex Genetic Regulation. Plant Physiol. 165: 933-947.

15*. Li Q, Eichten SR, Hermanson PJ, Springer NM. 2014. Inheritance patterns and stability of DNA methylation variation in maize near-isogenic lines. Genetics 196: 667-676.

16*. Li Q, Eichten SR, Hermanson PJ, Zaunbrecher VM, Song J, Wendt J, Rosenbaum H, Madzima TF, Sloan AE, Huang J, et al. 2014. Genetic perturbation of the maize methylome. Plant Cell 26: 4602-4616.

17*. Li Q, Song J, West PT, Zynda G, Eichten SR, Vaughn MW, Springer NM. 2015. Examining the causes and consequences of context-specific differential DNA methylation in maize. Plant Physiol 168(4):1262-74.

18*. Makarevitch I, Eichten SR, Briskine R, Waters AJ, Danilevskaya ON, Meeley RB, Myers CL, Vaughn MW, Springer NM. 2013. Genomic distribution of maize facultative heterochromatin marked by trimethylation of H3K27. Plant Cell 25: 780-793.

19*. Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM. 2015. Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS Genet. 11: e1004915.

20*. Springer NM 2013. Epigenetics and crop improvement. Trends Genet. 29: 241-247.

21*. Springer NM,McGinnis KM. 2015. Paramutation in evolution, population genetics and breeding. Semin. Cell Dev. Biol 44:33-38.

22*. Springer NM, Lisch D, Li Q. 2016. Creating Order from Chaos: Epigenome Dynamics in Plants with Complex Genomes. Plant Cell, 2016, 28, 2, 314-325

23*. Waters AJ, Bilinski P, Eichten SR, Vaughn MW, Ross-Ibarra J, Gehring M, Springer NM. 2013. Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. Proc. Natl. Acad. Sci. U. S. A. 110: 19639-19644.

24*. West PT, Li Q, Ji L, Eichten SR, Song J, Vaughn MW, Schmitz RJ, Springer NM. 2014. Genomic distribution of H3K9me2 and DNA methylation in a maize genome. PLoS One 9: e105267.

25*. Song, J., Zynda, G., Beck, S., Springer, N.M., and Vaughn, M.W. 2016. Bisulfite sequence analyses using CyVerse discovery environment: From mapping to DMRs. Curr. Protoc. Plant Biol. 1:510-529. doi: 10.1002/cppb.20034

26*. Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EK, Liu X, Gao D, Clevenger J, Dash S, et al. 2016. The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. Nat. Genet. 48: 438-446.

27*. Bewick AJ, Ji L, Niederhuth CE, Willing EM, Hofmeister BT, Shi X, Wang L, Lu Z, Rohr NA, Hartwig B, et al. 2016. On the origin and evolutionary consequences of gene body DNA methylation. Proc. Natl. Acad. Sci. U. S. A. 113: 9111-9116.

28*. El Baidouri M, Kim KD, Abernathy B, Arikit S, Maumus F, Panaud O, Meyers BC, Jackson SA. 2015. A new approach for annotation of transposable elements using small RNA mapping. Nucleic Acids Res. 43: e84.

29*. Jacobs TB, LaFayette PR, Schmitz RJ, Parrott WA. 2015. Targeted genome modifications in soybean with CRISPR/Cas9. BMC Biotechnol. 15: 16-015-0131-2.

30*. Ji L, Neumann DA, Schmitz RJ. 2015. Crop Epigenomics: Identifying, Unlocking, and Harnessing Cryptic Variation in Crop Genomes. Mol. Plant. 8: 860-870.

31*. Ji L, Sasaki T, Sun X, Ma P, Lewis ZA, Schmitz RJ. 2014. Methylated DNA is over-represented in whole-genome bisulfite sequencing data. Front. Genet. 5: 341.

32*. Kim KD, El Baidouri M, Abernathy B, Iwata-Otsubo A, Chavarro C, Gonzales M, Libault M, Grimwood J, Jackson SA. 2015. A Comparative Epigenomic Analysis of Polyploidy-Derived Genes in Soybean and Common Bean. Plant Physiol. 168: 1433-1447.

33*. Lane AK, Niederhuth CE, Ji L, Schmitz RJ. 2014. pENCODE: a plant encyclopedia of DNA elements. Annu. Rev. Genet. 48: 49-70.

34*. Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR. 2015. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. Nat. Protoc. 10: 475-483.

35*. van der Graaf A, Wardenaar R, Neumann DA, Taudt A, Shaw RG, Jansen RC, Schmitz RJ, Colome-Tatche M, Johannes F. 2015. Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. Proc. Natl. Acad. Sci. U. S. A. 112: 6676-6681.

36*. Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, Rohr NA, Rambani A, Burke JM, Udall JA, Egesi C, Schmutz J, Grimwood J, Jackson SA, Springer NM, Schmitz RJ. 2016. Widespread natural variation of DNA methylation within angiosperms. Genome Biol. 2016 Sep 27;17(1):194

37*. Wills DM, Whipple CJ, Takuno S, Kursel LE, Shannon LM, Ross-Ibarra J, Doebley JF. 2013. From many, one: genetic control of prolificacy during maize domestication. PLoS Genet. 9: e1003604.

38*. Mezmouk S,Ross-Ibarra J. 2014. The pattern and distribution of deleterious mutations in maize. G3 (Bethesda) 4: 163-171.

39*. Tiffin P,Ross-Ibarra J. 2014. Advances and limits of using population genetics to understand local adaptation. Trends Ecol. Evol. 29: 673-680.

40*. da Fonseca RR, Smith BD, Wales N, Cappellini E, Skoglund P, Fumagalli M, Samaniego JA, Caroe C, Avila-Arcos MC, Hufnagel DE, et al. 2015. The origin and evolution of maize in the Southwestern United States. Nat. Plants 1: 14003.

41*. Hake S,Ross-Ibarra J. 2015. Genetic, evolutionary and plant breeding insights from the domestication of maize. Elife 4: 10.7554/eLife.05861.

42*. Takuno S, Ralph P, Swarts K, Elshire RJ, Glaubitz JC, Buckler ES, Hufford MB, Ross-Ibarra J. 2015. Independent Molecular Basis of Convergent Highland Adaptation in Maize. Genetics 200: 1297-1312.

43*. Beissinger TM, Wang L, Crosby K, Durvasula A, Hufford MB, Ross-Ibarra J. 2016. Recent demography drives changes in linked selection across the maize genome. Nat. Plants 2: 16084.

44*. Orozco-Ramirez Q, Ross-Ibarra J, Santacruz-Varela A, Brush S. 2016. Maize diversity associated with social origin and environmental variation in Southern Mexico. Heredity (Edinb) 116: 477-484.

45*. Durvasula A, Hoffman PJ, Kent TV, Liu C, Kono TJ, Morrell PL, Ross-Ibarra J. 2016. Angsd-Wrapper: Utilities for Analysing Next-Generation Sequencing Data. Mol. Ecol. Resour. 16: 1449-1454.

46*. Sosso D, Luo D, Li QB, Sasse J, Yang J, Gendrot G, Suzuki M, Koch KE, McCarty DR, Chourey PS, et al. 2015. Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport. Nat. Genet. 47: 1489-1493.

47*. Bilinski P, Distor K, Gutierrez-Lopez J, Mendoza GM, Shi J, Dawe RK, Ross-Ibarra J. 2015. Diversity and evolution of centromere repeats in the maize genome. Chromosoma 124: 57-65.

48*. Yang, J., Mezmouk, S., Baumgarten, A., Buckler, E.S., Guill, K.E., McMullen, M.D., Mumm, R.H. and Ross-Ibarra, J., 2016. Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize. bioRxiv, p.086132.

49*. Altpeter F, Springer NM, Bartley LE, Blechl AE, Brutnell TP, Citovsky V, Conrad LJ, Gelvin SB, Jackson DP, Kausch AP, Lemaux PG, Medford JI, Orozco-Cárdenas ML, Tricoli DM, Van Eck J, Voytas DF, Walbot V, Wang K, Zhang ZJ, Stewart CN Jr. 2016. Advancing Crop Transformation in the Era of Genome Editing. Plant Cell. 2016 Jul;28(7):1510-20

50. Law JA,Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. 11: 204-220.

51. Du J, Johnson LM, Jacobsen SE, Patel DJ. 2015. DNA methylation pathways and their crosstalk with histone methylation. Nat. Rev. Mol. Cell Biol. 16: 519-532.

52. Matzke MA,Mosher RA. 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. Genet. 15: 394-408.

53. Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D. 2013. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153: 193-205.

54. Hu L, Li N, Xu C, Zhong S, Lin X, Yang J, Zhou T, Yuliang A, Wu Y, Chen YR, et al. 2014. Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. Proc. Natl. Acad. Sci. U. S. A. 111: 10642-10647.

55. Kawakatsu T, Huang SS, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y, Chen H, Dubin M, Lee CR, Wang C, Bemm F, Becker C, O'Neil R, O'Malley RC, Quarless DX; 1001 Genomes Consortium., Schork NJ, Weigel D, Nordborg M, Ecker JR. 2016. Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. Cell. 166(2):492-505

56. Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, Liu S, Yeh CT, Jia Y, Gendler K, Freeling M, et al. 2011. Heritable epigenetic variation among maize inbreds. PLoS Genet. 7: e1002372.

57. Regulski M, Lu Z, Kendall J, Donoghue MT, Reinders J, Llaca V, Deschamps S, Smith A, Levy D, McCombie WR, 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

58. Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, Deragon JM, Westerman RP, Sanmiguel PJ, Bennetzen JL. 2009. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genet. 5: e1000732.

59. Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0. 2015 <http://www.repeatmasker.org>.

60. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin C, Guill K, Regulski M, Kumari S, Olson A, Gent J, Schneider KL, Wolfgruber TK, May MR, Springer NM, Antoniou E, McCombie R, Presting GG,McMullen M, Ross-Ibarra J, Dawe K, Hastie A, Rank DR, Ware D. (2016) The complex sequence landscape of maize revealed by single molecule technologies. BioRxiv http://dx.doi.org/10.1101/079004

61. Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A. 2005. Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat. Genet. 37: 997-1002.

62. Cridland JM, Macdonald SJ, Long AD, Thornton KR. 2013. Abundance and distribution of transposable elements in two Drosophila QTL mapping resources. Mol. Biol. Evol. 30: 2311-2327.

63. Robb SM, Lu L, Valencia E, Burnette JM,3rd, Okumoto Y, Wessler SR, Stajich JE. 2013. The use of RelocaTE and unassembled short reads to produce high-resolution snapshots of transposable element generated diversity in rice. G3 (Bethesda) 3: 949-957.

64. Dlugosz EM, Lenaghan SC, Stewart CN,Jr. 2016. A Robotic Platform for High-throughput Protoplast Isolation and Transformation. J. Vis. Exp. (115). doi: 10.3791/54300.

65. Burris KP, Dlugosz EM, Collins AG, Stewart CN,Jr, Lenaghan SC. 2016. Development of a rapid, low-cost protoplast transfection system for switchgrass (Panicum virgatum L.). Plant Cell Rep. 35: 693-704.

66. Pigna G, Dhillon T, Dlugosz EM, Yuan JS, Gorman C, Morandini P, Lenaghan SC, Stewart CN,Jr. 2016. Methods for suspension culture, protoplast extraction, and transformation of high-biomass yielding perennial grass Arundo donax. Biotechnol. J. 11: 1657-1666.

67. Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, Zhong X, Marson G, Pellegrini M, Segal DJ, et al. 2014. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature 507: 124-128.

68. Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, Shu J, Dadon D, Young RA, Jaenisch R. 2016. Editing DNA Methylation in the Mammalian Genome. Cell 167: 233-247.e17.

69. Hirsch CN, Hirsch CD, Brohammer AB, Bowman MJ, Soifer I, Barad O, Shem-Tov D, Baruch K, Lu F, Hernandez AG, Fields CJ, Wright CL, Koehler K, Springer NM, Buckler E, Buell CR, de Leon N, Kaeppler SM, Childs KL, Mikel MA. 2016. Draft Assembly of Elite Inbred Line PH207 Provides Insights into Genomic and Transcriptome Diversity in Maize. Plant Cell. 28(11):2700-2714.

70. Lu F, Romay MC, Glaubitz JC, Bradbury PJ, Elshire RJ, Wang T, Li Y, Li Y, Semagn K, Zhang X, et al. 2015. High-resolution genetic mapping of maize pan-genome sequence anchors. Nat. Commun. 6: 6914.

71. Waters AJ, Makarevitch I, Noshay J, Burghardt LT, Hirsch CN, Hirsch CD, Springer NM. 2017. 71 variation for gene expression responses to abiotic stress in maize. Plant J. Doi:10.1111/tpj.13414

72. McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, et al. 2009. Genetic properties of the maize nested association mapping population. Science 325: 737-740.

73. Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES. 2005. Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. 44: 1054-1064.

74. Hansey CN, Johnson, JM, Sekhon RS, Kaeppler SM, de Leon N. 2010. Genetic diversity of a maize association population with restricted phenology. Crop Science 51 (2), 704-715

75. Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Penagaricano F, Lindquist E, Pedraza MA, Barry K, et al. 2014. Insights into the maize pan-genome and pan-transcriptome. Plant Cell 26: 121-135.

76. Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, et al. 2009. The genetic architecture of maize flowering time. Science 325: 714-718.

77. Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, Hlaing AL, Aung HH, Nelissen H, _. Baute J, et al. 2015. Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. Genome Biol. 16: 167-015-0716-z.

78. Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88: 76-82.

79. Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES. 2016. Open chromatin reveals the functional maize genome. Proc. Natl. Acad. Sci. U. S. A. 113: E3177-84.

80. Renny-Byfield S, Rodgers-Melnick E, Ross-Ibarra J. 2016. Gene fractionation and function in the ancient subgenomes of maize. bioRxiv https://doi.org/10.1101/095547

81. Stelpflug SC, Eichten SR, Hermanson PJ, Springer NM, Kaeppler SM. 2014. Consistent and heritable alterations of DNA methylation are induced by tissue culture in maize. Genetics 198: 209-218.

82. Stroud H, Ding B, Simon SA, Feng S, Bellizzi M, Pellegrini M, Wang GL, Meyers BC, Jacobsen SE. 2013. Plants regenerated from tissue culture contain stable epigenome changes in rice. Elife 2: e00354.

83. Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, Elshire RJ, Acharya CB, Mitchell SE, Flint-Garcia SA, et al. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 14: R55-2013-14-6-r55.

84. Sheen J 1991. Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes. Plant Cell 3: 225-245.

85. Yoo SD, Cho YH, Sheen J. 2007. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat. Protoc. 2: 1565-1572.

86. Gordon-Kamm WJ, Spencer TM, Mangano ML, Adams TR, Daines RJ, Start WG, O'Brien JV, Chambers SA, Adams WR,Jr, Willetts NG, et al. 1990. Transformation of Maize Cells and Regeneration of Fertile Transgenic Plants. Plant Cell 2: 603-618.

87. Rhodes CA, Lowe KS, Ruby KL. 1988. Plant regeneration from protoplasts isolated from embryogenic maize cell cultures. Nature Biotechnology 6, 56 – 60.

88. Zetsche B, Heidenreich M, Mohanraju P, Fedorova I, Kneppers J, DeGennaro EM, Winblad N, Choudhury SR, Abudayyeh OO, Gootenberg JS, Wu WY, Scott DA, Severinov K, van der Oost J, Zhang F. 2016. multiplex gene editing by CRISPR-Cpf1 through autonomous processing of a single crRNA array. BioRxiv. https://doi.org/10.1101/049122

89. Bewick AJ, Hofmeister BT, Lee K, Zhang X, Hall DW, Schmitz RJ. 2015. FASTmC: A Suite of Predictive Models for Nonreference-Based Estimations of DNA Methylation. G3 (Bethesda) 6: 447-452.

90. Zemach A, Kim MY, Silva P, Rodrigues JA, Dotson B, Brooks MD, Zilberman D. 2010. Local DNA hypomethylation activates genes in rice endosperm. Proc. Natl. Acad. Sci. U. S. A. 107: 18729-18734.

91. Via A, Blicher T, Bongcam-Rudloff E, Brazas MD, Brooksbank C, Budd A, De Las Rivas J, Dreyer J, Fernandes PL, van Gelder C, et al. 2013. Best practices in bioinformatics training for life scientists. Brief Bioinform 14: 528-537.

92. Teal, T.K., Cranston, K.A., Lapp, H., White, E., Wilson, G., Ram, K., and Pawlik, A. 2015. Data Carpentry: Workshops to Increase Data Literacy for Researchers. International Journal of Digital Curation 10(1):135-143.

93. Koch, C and Wilson, G (eds.). 2016. Software Carpentry: Instructor Training. Zenodo. http://doi.org/10.5281/zenodo.57571, http://swcarpentry.github.io/instructor-training/

94. Data Carpentry. Accessed 2016. Lessons, http://www.datacarpentry.org/lessons/

95. Data Carpentry. Accessed 2016. Roadmap for Lesson Development, http://www.datacarpentry.org/lessons-incubation/