The Plant Cell, Vol. 24: 2257–2261, June 2012, www.plantcell.org

COMMENTARY

Reading the Second Code: Mapping Epigenomes to Understand Plant Growth, Development, and Adaptation to the Environment [OA]

The EPIC Planning Committee[1,2]

We have entered a new era in agricultural and biomedical science made possible by remarkable advances in DNA sequencing technologies. The complete sequence of an individual's set of chromosomes (collectively, its genome) provides a primary genetic code for what makes that individual unique, just as the contents of every personal computer reflect the unique attributes of its owner. But a second code, composed of "epigenetic" layers of information, affects the accessibility of the stored information and the execution of specific tasks. Nature's second code is enigmatic and must be deciphered if we are to fully understand and optimize the genetic potential of crop plants. The goal of the Epigenomics of Plants International Consortium is to crack this second code, and ultimately master its control, to help catalyze a new green revolution.

INTRODUCTION

Global Needs and Challenges

Since the dawn of human existence, we have relied on plants as sources of food, medicine, building materials, fiber, and fuel. Remarkable increases in agricultural productivity have occurred during the past century, made possible by advances in engineering (equipment and mechanization), breeding (development of high-performing germplasm), and chemistry (e.g., fertilizers and pesticides) while drawing upon multiple disciplines of biology, including genetics, physiology, plant pathology, and biochemistry. To face the challenges inherent to an ever-increasing human population with increasing consumption, shrinking agricultural lands, and a changing, unpredictable environment, it is critical that we maximize the usefulness and productivity of traditional crop plants while exploiting the potential of new crops. Moreover, for agriculture to be environmentally responsible and sustainable, increased productivity must be attained with fewer inputs, including fertilizer, water, pesticides, and fossil fuels.

To meet the global challenges ahead, academic, industry, and government scientists, together with representatives from government funding agencies, growers, and commodity groups, have begun to identify and articulate priorities for the next decade of plant research (Ledford, 2011; Pennisi, 2011). Among these priorities are the following: (1) the need to understand how the genes that constitute a plant's genome specify that plant's form and function(s); (2) the need to better understand how genes are switched on, switched off, or tuned to different levels of expression; (3) the need to be able to predict a plant's phenotype (traits) from its genotype (sets of genes); and (4) the need to understand the molecular basis for the genotype–environment interactions that alter a plant's performance in different circumstances. These are challenging questions that require an understanding of genetic regulation, mediated through the action of transcription factors and other regulatory proteins, as well as epigenetic regulation, the elusive "second code" that can explain variable or alternative gene expression states in response to environmental or developmental cues. In recognition of this need, the Epigenomics of Plants International Consortium (EPIC) has been organized to develop enabling infrastructure and tools, identify research priorities, coordinate research efforts, and share data to accelerate the pace of discovery for the common good.

Interrelationships between Genetics, Epigenetics, and Epigenomics

High-throughput DNA sequencing has revolutionized genetics, allowing the rapid elucidation of genome sequences (the sum of all chromosomes) for a growing list of species. Computer-based bioinformatic analyses then allow prediction of the thousands of genes and genetic elements encoded by the genome. Together, sequencing and bioinformatics are core activities of the discipline of genomics. Knowing an organism's genome sequence provides insight into its genetic and biochemical potential. Likewise, genome comparisons can identify candidate genes that might explain trait differences among breeding lines or species, such as drought, salt, or cold tolerance, pathogen resistance, or biofuel potential.

Importantly, genome sequence information alone does not reveal how genes are regulated. Every cell of a species has the same genome, yet different sets of genes are expressed in specific cell types at different times in development or in response to the environment. Although the genome sequence provides the genetic blueprint, including the programmed expression of transcription factors that are critical for the specification of specific cell types, additional "epigenetic" layers of information are imposed upon this primary code to influence gene accessibility and readout. Epigenetic information, therefore, constitutes a second code that must be deciphered to understand both developmental gene regulation and genomic responses to environmental changes.

[1]A list of participants and their affiliations is provided at the end of this article.
[2]Address correspondence to wagnerdo@sas.upenn.edu.

www.plantcell.org/cgi/doi/10.1105/tpc.112.100636

COMMENTARY

To understand the epigenetic second code, it is essential to recognize that chromosomes are not simply DNA molecules but DNA–protein complexes known collectively as chromatin. Chromatin can be modified in many ways. For instance, the DNA can be methylated on its cytosines, one of the four nucleotides that form the genetic code; the histone proteins that wrap and organize the DNA can be chemically modified; the positions of histones and/or other chromatin proteins along the DNA can be altered; and distinct histone protein variants can be employed at specific sites along the DNA. These and other reversible modifications influence whether genes are turned on or off. The field of epigenomics is adapting the high-throughput methodologies of genomics to investigate the chemical nature of epigenetic information, to map the occurrence of chromatin modifications, and to understand the dynamics or stability of these modifications in the context of development, responses to the environment, and heritability in subsequent generations.

Plants as Model Systems for Epigenetics and Epigenomics

Within the fields of epigenetics and epigenomics, discoveries in plants have had broad relevance to human and animal biology. The elucidation of key components of DNA methylation, DNA demethylation, and small RNA–mediated gene silencing pathways are prime examples where plant research has laid the foundation for studies in mammals and other experimental model systems. In plants, as in humans, DNA methylation is required for development and for silencing potentially deleterious mobile genetic elements. By contrast, other model genetic systems, such as yeast, nematodes, and fruit flies, do not appreciably methylate their DNA. Plant research led to the discovery that small RNAs specify sites of DNA methylation and epigenetic modification. It is now recognized that similar mechanisms control DNA methylation in mammals. In plants, as in mammals, DNA demethylation is an essential aspect of reproductive development, and the enzymes responsible for demethylating DNA were first identified in plants. These examples illustrate the fact that plant studies have had a major impact on understanding the regulation of DNA methylation in normal development and its misregulation in disease states, including cancer and genetic disorders. Importantly, plants tolerate null mutations that eliminate whole pathways or chromatin regulators whose loss is lethal in animals, allowing the functions of the missing activities to be studied. For these reasons, plant epigenetic and epigenomic studies continually yield biomedically relevant discoveries, not possible in humans or other model systems, in the course of revealing regulatory mechanisms that are important to the unique biology of plants and their usefulness as sources of food, fiber, medicine, and fuel.

Data Sets That Collectively Define an Epigenome

Plants, like humans, are composed of multiple cell types, each with the same DNA, but with different sets of genes turned on or off and thereby having different epigenomes. The epigenome includes the cell's chromosomal DNA sequences (genome), the epigenetic marks in place on its genomic chromatin, the expression state of its genes, and the populations of small and noncoding RNAs that can trigger or reinforce epigenetic chromatin states. The technique of bisulfite-mediated DNA sequencing allows one to determine the frequency at which every cytosine in the genome is methylated. The technique of chromatin immunoprecipitation, combined with high-throughput DNA sequencing, is used to map the genomic positions of chromatin-associated proteins that can be targeted by specific antibodies, including transcription factors, repressors, or histones displaying chemical modifications indicative of gene "on" or "off" states. Deep sequencing of small RNAs allows one to determine their abundance, occurrence, and potential involvement in specifying epigenetic marks throughout the epigenome. Likewise, deep sequencing of protein-encoding mRNAs and long noncoding RNAs defines the genome's expression state. Comparison of such data sets allows correlations between different epigenetic modifications to be discerned. For instance, the correlation between 24-nucleotide small interfering RNA abundance, high cytosine methylation, and mRNA suppression is striking and indicative of an RNA-mediated gene-silencing mechanism operating genome-wide. Such epigenomic approaches and maps are now well developed in select plant models, such as *Arabidopsis thaliana* and rice (*Oryza sativa*).

Background

The EPIC initiative began with an international conference in Australia in 2008, with funding from Australia (Commonwealth Scientific and Industrial Research Organization [CSIRO]), the United States (National Science Foundation), and the United Kingdom (Biotechnology and Biological Science Research Council). The deliberations of this conference led to a Research Coordination Network proposal to launch the EPIC initiative, submitted in early 2009 to the U.S. National Science Foundation by coauthors Doris Wagner, Craig Pikaard, and Robert Martienssen. The Research Coordination Network grant was funded mid-year in 2010. Launch of the EPIC website (https://www.plant-epigenome.org/), a first meeting of an interim International Steering Committee, and a workshop to announce publicly the EPIC initiative occurred in January 2011, coinciding with the annual Plant and Animal Genome Conference in San Diego. Additional EPIC workshops were held at the International Conference on Arabidopsis Research in 2011 and at the 2012 Plant and Animal Genome Conference. During 2011, epigenomics experts from North America, Europe, Asia, Australia, and South America were recruited and merged with an interim Steering Committee to form an EPIC Planning Committee (see the end of the article for members). In November 2011, the Planning Committee convened a Banbury Conference at Cold Spring Harbor,

COMMENTARY

New York, to establish a set of recommendations and priorities for plant epigenomics research in the next decade. A resulting draft policy document was then distributed to the Planning Committee and discussed at an EPIC workshop in Beijing in April 2012. The current policy document, presented here, is the outcome of these meetings and deliberations.

GOALS AND RECOMMENDATIONS

Essential Data Sets

Critical to EPIC is the generation, analysis, and display of data sets that define epigenomes and allow correlations (positive or negative) among epigenetic marks and gene expression states to be evaluated. Essential genome-wide data sets include the following:

- Deep-sequence data identifying all RNA transcripts, including protein-coding mRNAs, long noncoding RNAs, and small noncoding RNAs (e.g., microRNAs and short-interfering RNAs). These data collectively reflect the expression state of the epigenome.
- Positions, and frequencies, of methylated cytosines, determined at single nucleotide resolution, genome-wide.
- Positions of nucleosomal histones bearing informative modifications, including, but not limited to, marks typical of silent chromatin (e.g., histone H3 that is monomethylated, dimethylated, or trimethylated on Lys-9 [abbreviated as H3K9me1, H3K9me2, or H3K9me3] or Lys-27 [H3K27me1, H3K27me2, or H3K27me3]) and marks typical of active chromatin (H3K4me1, H3K4me2, and H3K4me3, H3K36me, or acetylated H3 and H4).
- Positions of histone variants (e.g., H2AZ, centromere-specific H3).
- Positions of DNase-hypersensitive sites, corresponding to likely gene regulatory regions.
- Positions of nucleosomes, which can reveal nucleosome-free or nucleosome-depleted regions that are hotspots of gene regulation.

In addition to these core analyses, numerous additional data sets are highly desirable. These include positions of histones bearing other posttranslational modifications, positions of hydroxymethylated cytosines or other modified bases, positions of the DNA-dependent RNA polymerases that carry out transcription, positions of important DNA- and histone-modifying enzymes, and positions of transcriptional coactivator or corepressor complexes. Importantly, an advantage of plants for such studies is that cell-specific marker and sorting technologies allow epigenomic analyses to be conducted for nearly every cell type of the plant body, allowing studies of the highest possible resolution and promising important insights into the unique physiologies of different cell types important as sources of food, fiber, fuel, or medicine. We envision that the most in-depth analyses will be performed using a select subset of species to elucidate a set of reference epigenomes that can guide subsequent studies. These reference species will include *Arabidopsis*, rice, and maize (*Zea mays*). However, EPIC's goal is to facilitate the development of data deposition, data access, and data analysis tools that can be used for studies of species beyond the reference set.

Research Priorities

Through the acquisition of the data sets described above, EPIC's goal is to facilitate the epigenomics community's search for answers to important biological questions, including:

- What are the relative roles of transcription factors versus epigenetic regulators in the establishment or maintenance of transcriptional states?
- How similar are the epigenomes of different species, and can epigenomic states in model systems be extrapolated to other plants, including crops that may be difficult to study?
- Is hybrid vigor (which results from the mating of two dissimilar parents) an epigenetic phenomenon?
- How does the epigenome change in response to environmental stresses (e.g., drought, salt, cold, heat, pollution, nutrient limitation, etc.) or pathogen attack?
- What is the molecular basis for epigenetic inheritance and the perpetuation of acquired states?
- How much epigenetic variation exists within a population, such as a breeding population, and is this variation linked to genetic variation?
- Can epigenomes be engineered?
- How does the epigenome change during developmental transitions, such as when stem cells and their progeny differentiate into specific cell types?
- What epigenetic changes occur in shoot apical meristems as they transition from being vegetative meristems to floral meristems, and how do photoperiod, light quality, and temperature affect these transitions?
- How are important aspects of chromosome biology epigenetically regulated, including centromere function, telomere function, replication, recombination, and DNA repair?
- How extensive is the crosstalk between DNA methylation and histone posttranslational modification, and how does this crosstalk occur?
- How do different types of epigenetic chromatin modifications communicate with, and specify, one another?
- What are the epigenetic consequences of polyploidy, imprinting, paramutation, nucleolar dominance, apomixis, and other common features of plant genomes?

We anticipate that experiments to address these questions will be conducted in a variety of plant species by individual investigators or self-organized consortia that possess the expertise to tackle the questions in the appropriate species. EPIC believes strongly in the value of investigator-initiated proposals that address priorities to be enunciated in calls for proposals by funding agencies within the nations represented by its members. EPIC's role is not to directly fund research but to facilitate research through the development of standards for data collection and analysis,

COMMENTARY

computational tools, data sharing, and coordination of efforts.

Vision for an EPIC Informatics Hub

To date, the largest number of plant epigenomics studies have been conducted using the dicotyledonous plant *Arabidopsis*. These studies have taken place in different laboratories, under different conditions, and sometimes using different methodologies, resulting in data that are in multiple formats and not available from a single database. Therefore, to facilitate analyses of existing and future data sets, the EPIC Planning Committee identified as an immediate priority the need for the creation of an EPIC informatics hub to serve as a central community resource. This vision includes collaborating with iPlant (funded by the U.S. National Science Foundation), the nascent Arabidopsis Information Portal, and other computational and informatics resource centers to:

- Establish a central online resource, providing access to data sets and experimental information.
- Develop data-formatting standards to facilitate easy access and comparison of new or existing data sets and require that information concerning the developmental stages, growth conditions, or tissue types examined be provided. EPIC researchers will be asked to deposit their data in the Gene Expression Omnibus (GEO) and then upload the GEO accession information to the EPIC website, generating a link to the data. Periodic automated searches of GEO to retrieve relevant data sets are also envisioned.
- Develop a user-friendly epigenomic data browser and sets of computational tools for epigenomic data comparisons and analyses. The browser should allow for easy upload and coordinated graphical display of new, user-generated data sets and existing data sets as customizable tracks. Among the features to display are positions of annotated genes and their transcripts; positions of methylated cytosines; positions of nucleosomes, histone variants, and histones bearing specific chemical modifications; positions of long and small noncoding RNAs; and positions of important chromatin-associated proteins, including transcription factors, corepressors, and chromatin-modifying activities.
- Develop computational tools for querying whether two or more epigenomic features overlap, or are independent of one another, and provide coordinates for overlapping features.
- Develop tools for searching DNA regulatory elements, Gene Ontology terms, metadata, or author information.
- Develop tools for the comparison of independent data sets that examined the same epigenetic mark to determine the degree of concordance or dissidence among the data sets.
- Develop tools for comparisons of orthologous genes in two or more species.
- Develop tools and pipelines for the generation of publishable graphical displays of features being compared, including statistical analyses where appropriate.

The overall goal for the informatics hub is to develop software and user-friendly computer interfaces that allow users to deposit, upload, and analyze data without the need for computer programming skills. Tools and resources generated by epigenetics initiatives focused on organisms other than plants will be utilized where possible, so as to avoid unnecessary duplications of effort (e.g., the Human Epigenome Browser [http://epigenomegateway.wustl.edu/]).

Training Needs

A limiting factor for epigenomics research is the lack of sufficient numbers of informaticians able to collaborate in the design and interpretation of epigenomics data sets. To meet this need, EPIC will advocate for increased funding of fellowships, training grants, and other initiatives that can spur the acquisition of computer language and bioinformatics skills among life science students, or of molecular biology knowledge and skills among computer science and informatics students.

We also plan to develop a more unified outreach from the scientific community to seed companies, growers, environmentalists, professional societies, government and nongovernment organizations, other stakeholders, and the general public. A broader awareness of the emerging challenges of epigenetics research in plants will be critical in garnering support for a major public initiative in this arena.

Structure of the Consortium

Community Members

Researchers, students, and other interested individuals become members of the EPIC community by self-registering at the EPIC website. By doing so, members are added to a listserve that allows for information dissemination, discussions, and voting.

Advisory Board

During the current planning phase of EPIC, prominent scientists recognizing the need for an international plant epigenomics initiative were invited to serve on the EPIC Planning Committee by the Steering Committee that administers the Research Coordination Network grant. As EPIC moves from the planning phase to a funded implementation phase, the planning committee will be replaced by a 12-member Scientific Advisory Board that will be elected by the community. The Scientific Advisory Board will be responsible for overseeing the activities of EPIC, with authority to appoint working committees and oversight committees, including representatives of funding entities.

Funding Entities

Nations, groups of nations (e.g., the European Union), corporations, foundations, and other entities providing funding for EPIC will not contribute funds to a pool but will make their own funding decisions according to the priorities of their citizens or stakeholders.

Core Guidelines

Researchers wishing to upload data to the EPIC browser are expected to abide by

COMMENTARY

data generation, deposition, and formatting standards detailed at the browser interface. These guidelines will include information concerning (1) minimal data sets for optimal comparability; (2) parameters for growth, harvest, treatment, etc.; and (3) time frames for the public release of the data, which may vary depending on the mandates of the funding agencies involved.

To allow EPIC data producers to publish the initial analyses of their own data, if data are released prior to publication, resource users who were not involved in generating the data sets are expected to observe a moratorium before any manuscripts containing analyses of the data are submitted. The moratorium period will be 9 months from the date of data submission, indicated by a time stamp generated by the data repository.

Funding entities should encourage their grantees to follow the data-release guidelines.

A core philosophy of EPIC is that all information should be freely accessible and unencumbered by legal considerations.

Deliverables

Short-Term Goals

These include the establishment of the EPIC informatics hub, including the creation of a central data access site, epigenome browser, and computational tools for the analysis of epigenomic data sets. With appropriate funding (estimated at several million dollars), the needed computer infrastructure and tools could be fully implemented within 2 years.

Mid-Term Goals

Within 5 years, we can expect to have complete data sets for essential epigenomic marks at specific developmental time points, and in single organ, tissue, or cell types, for *Arabidopsis*, maize, and rice and to have partial data sets for other crops or plants of interest. We can expect to have information as to the stability or instability of the epigenome and how it changes during development or in response to environmental variables. We expect to have a better understanding of how different epigenetic modifications specify, reinforce, or antagonize one another and how epigenome profiles are altered upon the loss of key epigenetic modifiers due to mutation or natural variation.

Long-Term Goals

Within 10 years, we can expect to understand the nature of epigenetic inheritance and the perpetuation of acquired traits as well as the roles of epigenetic modifications in the essential functions of chromosomes. We can expect to better understand the molecular basis for genotype–environment interactions and whether the epigenome can be engineered to alter gene expression networks, allowing desired traits to be predicted and realized.

PRIMARY CONTRIBUTORS (WORKSHOP ORGANIZERS AND SIGNIFICANT CONTRIBUTORS TO THE FINAL REPORT)

Fred Berger, Temasek Life Sciences Laboratory, Singapore; XiaoFeng Cao, Chinese Academy of Sciences, China; Vicki Chandler, The Gordon and Betty Moore Foundation, United States; Liz Dennis, CSIRO, Australia; Rob Martienssen, Cold Spring Harbor Laboratory, United States; Blake Meyers, University of Delaware, United States; Craig Pikaard, Indiana University, United States; Jim Peacock, CSIRO, Australia; Eric Richards, Boyce Thompson Institute, Cornell University, United States; Doris Wagner, University of Pennsylvania, United States; Detlef Weigel, Max Planck Institute, Germany.

OTHER SIGNIFICANT CONTRIBUTORS (WORKSHOP PARTICIPANTS AND REPORT SECTION CONTRIBUTORS)

Vincent Colot, Ecole Normale Superieur, France; Roger Deal, Emory University, United States; Caroline Dean, John Innes Centre, United Kingdom; Joe Ecker, SALK Institute, United States; Mary Gehring, Massachusetts Institute of Technology, United States; Zhizhong Gong, China Agricultural University, China; Brian Gregory, University of Pennsylvania, United States; Gutierrez Rodrigo, Pontifical Catholic University, Chile; Jose Gutierrez-Marcos, University of Warwick, United Kingdom; Mitsuyasu Hasebe, National Institute for Basic Biology, Japan; Il-Doo Hwang, Pohang University of Science and Technology, South Korea; Steve Jacobsen, University of California Los Angeles, United States; Tetsuji Kakutani, National Institute of Genetics, Japan; Jiayang Li, Chinese Academy of Sciences, China; Scott Michaels, Indiana University, United States; Yoo-Sun Noh, Seoul National University, South Korea; Nick Provart, University of Toronto, Canada; Matt Vaughn, University of Texas, United States.

AUTHOR CONTRIBUTIONS

Craig Pikaard wrote the article with help and input from the primary contributors; other contributors provided input at the workshop and in section reports.

Received May 16, 2012; revised June 8, 2012; accepted June 14, 2012; published June 29, 2012.

REFERENCES

Ledford, H. (2011). US plant scientists seek united front. Nature **477:** 259.

Pennisi, E. (September 26, 2011). Decadal plan for plant science begins to take shape. *Science*Insider September 2011. http://news.sciencemag.org/scienceinsider/2011/09/decadal-plan-for-plant-science.html. (Accessed May 1, 2012.)