



The Epigenetic Landscape of Plants

Xiaoyu Zhang, *et al.*
Science **320**, 489 (2008);
DOI: 10.1126/science.1153996

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 25, 2008):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/320/5875/489>

This article **cites 53 articles**, 22 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/320/5875/489#otherarticles>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

PERSPECTIVE

The Epigenetic Landscape of Plants

Xiaoyu Zhang

In plants, DNA methylation, histone modifications, and RNA interference play critically important roles in regulating chromatin structure, thereby profoundly affecting transcription and other molecular events. Recent advances in microarray and high-throughput sequencing technologies have enabled genome-wide studies of these pathways in great detail. The vast amounts of “epigenomic” data generated so far have provided new insights into the mechanisms and functions of these pathways and have broadened our understanding of the structure and organization of plant chromatin as a whole.

The genomes of several plants have been sequenced, and those of many others are under way. But genetic information alone cannot fully address the fundamental question of how genes are differentially expressed during cell differentiation and plant development, as the DNA sequences in all cells in a plant are essentially the same. Numerous studies in the past decade have unveiled the importance of several mechanisms in regulating transcription by affecting the structural properties of the chromatin. These mechanisms, including DNA cytosine methylation, covalent modifications of histones, and certain aspects of RNA interference (RNAi), are referred to as “epigenetic” because they direct “the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states” (1).

The components and the mechanistic aspects of plant epigenetic pathways as well as their functions in regulating plant development have been the subjects of several excellent recent reviews (2–6). My goal here is to summarize the results from genome-wide profiling studies of DNA methylation, histone modifications, and the aspects of RNAi that are relevant to chromatin modifications.

To date, most such epigenomic studies have been performed in *Arabidopsis thaliana*, largely as a result of the availability of powerful high-throughput tools (e.g., high-density whole-genome microarrays). The compact genome (~120 Mb) of *Arabidopsis* has been completely sequenced (7). Most of the repetitive sequences (~20 Mb; mostly transposons and their relics) cluster in the pericentromeric regions, whereas the majority of the ~27,000 protein-coding genes are distributed on the arms of the five chromosomes (Fig. 1).

DNA Methylation

Three distinct DNA methylation pathways with overlapping functions have been characterized in *Arabidopsis*. The mammalian DNMT1 homolog METHYLTRANSFERASE 1 (MET1) primarily maintains DNA methylation at CG sites (CG methylation) (8). The plant-specific CHROMO-METHYLASE3 (CMT3) interacts with the H3 Lys⁹ dimethylation (H3K9me2) pathway to main-

tain DNA methylation at CHG sites (CHG methylation, H = A, C, or T) (9, 10). The DNMT3a/3b homologs DOMAINS REARRANGED METHYLASE 1 and 2 (DRM1/2) maintain DNA methylation at CHH sites (CHH methylation), which requires the active targeting of small interfering RNAs (siRNAs) (11, 12).

The genome-wide pattern of DNA methylation has been the subject of several waves of microarray studies (13). Methylated and unmethylated DNA can be distinguished by three major types of experimental approaches: sodium bisulfite treatment that converts cytosine (but not methyl-cytosine) to uracil, enzymatic digestion (using methylation-specific endonucleases or methylation-sensitive isoschizomers), and affinity purification or immunoprecipitation (with methyl-cytosine binding proteins and antibodies to methyl-cytosine, respectively). The methylated fraction of the genome is then visualized by hybridizing treated DNA to microarrays (14–20). Results from these microarray studies were largely consistent: About ~20% of the *Arabidopsis* genome is methylated, with transposons and other repeats comprising the largest fraction, whereas the promoters of endogenous genes are rarely methylated. Surprisingly, methylation in the transcribed regions of endogenous genes is unexpectedly rampant. More than one-third of all genes contain methylation (called “body methylation”) that is enriched in the 3' half of the transcribed regions and primarily occurs at CG sites (Fig. 1 and Fig. 2A).

Most recently, Cokus, Feng, and colleagues combined sodium bisulfite treatment of genomic DNA with ultrahigh-throughput sequencing (>20× genome coverage) to generate the first DNA methylation map for any organism at single-base resolution (21). Relative to microarray-based methods, this “BS-Seq” method offers several advantages. First, it can detect methylation in important genomic regions that are not covered by any microarray platform (such as telomeres, ribosomal DNA, etc.). Second, it reveals the sequence contexts of DNA methylation (i.e., CG, CHG, and CHH) and therefore provides important information regarding the epigenetic pathways that function at any given locus. For example, all three types of methylation colocalize to transposons, but gene body methylation occurs ex-

clusively at CG sites. Third, BS-Seq is more effective in detecting light methylation and subtle changes (e.g., in mutants). Fourth, the theoretically unlimited sequencing depth makes it possible to quantitatively measure the percentage of cells in which any particular cytosine is methylated, thereby offering important clues regarding potential cell-specific DNA methylation.

DNA methylation is critically important in silencing transposons and regulating plant development. Severe loss of methylation results in a genome-wide massive transcriptional reactivation of transposons (14, 17, 18), and quadruple mutations in *drm1 drm2 cmt3 met1* cause embryo lethality (22). Interestingly, the role of DNA methylation in regulating transcription appears to depend on the position of methylation relative to genes. Methylation in promoters appears to repress transcription (18). Paradoxically, however, body-methylated genes are usually transcribed at moderate to high levels and are transcribed less tissue-specifically relative to unmethylated genes (16–18). Loss of body methylation does not seem to trigger a systematic and drastic overexpression of body-methylated genes to the same extent as transposon reactivation (14, 17, 18). However, a moderate up-regulation of body-methylated genes was observed, suggesting that body methylation might be involved in fine-tuning transcription levels (17). In addition, differential body methylation patterns in different *Arabidopsis* ecotypes are not preferentially correlated with differential gene expression (16). Thus, the exact role of body methylation in regulating transcription remains unknown (2). However, erasure of body methylation in the *met1* mutant seems to trigger stochastic redistribution of histone modifications (e.g., H3K9me2) and hyper-CHG and CHH methylation within genes, at least some of which result in ectopic gene silencing (20, 23).

The ability to profile DNA methylation in a high-throughput manner also enables studies of DNA methylation from population genetics and evolutionary perspectives. Much like comparative genomics, a “comparative epigenomics” approach should allow the identification of evolutionarily conserved DNA methylation at specific loci with important functions, as well as differences in DNA methylation that contribute to phenotypic variations. The first such study identified numerous differentially methylated sites between two *Arabidopsis* ecotypes that diverged thousands of years ago (16). However, these two ecotypes share DNA methylation patterns at the majority of the sites assayed, indicating that DNA methylation could be stably maintained genome-wide for thousands of years.

Histone Modifications

The patterns of covalent histone modification are highly complex because of the large number of residues that can be modified as well as the multiple, combinatorial modifications (e.g., meth-

Department of Plant Biology, University of Georgia, Athens, GA 30602, USA. E-mail: xiaoyu@plantbio.uga.edu

Plant Genomes

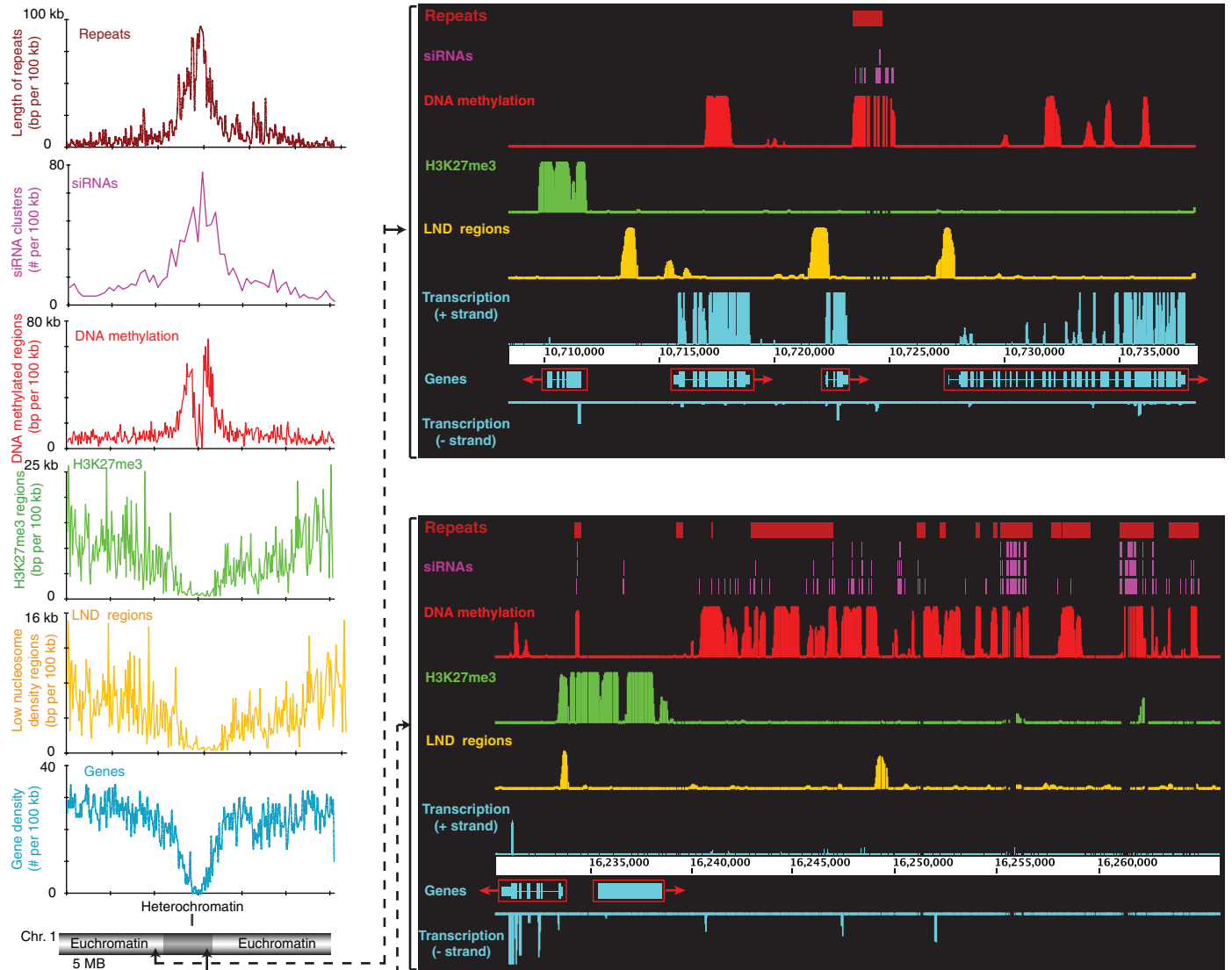


Fig. 1. Distribution of genes, repetitive sequences, DNA methylation, siRNAs, H3K27me3, and low nucleosome density (LND) regions in *Arabidopsis* (18, 25, 31). The chromosomal distributions are shown on the left, using chromosome 1 as an example. The x axis shows chromosomal position.

Right panels: detailed distribution patterns and transcription activity (vertical blue bars) in a gene-rich region (top) and a repeat-rich region (bottom). Red boxes indicate genes; arrows indicate the direction of transcription.

ylation, acetylation, phosphorylation, ubiquitination, etc.) (6). Some modifications directly alter chromatin structure, whereas others serve as binding platforms to recruit additional effectors.

To date, the genomic distribution patterns of histone H3 methylated at several lysine residues (H3K4me2, H3K9me2/3, and H3K27me3) have been determined in *Arabidopsis* by microarray analysis of samples from chromatin immunoprecipitation (ChIP-chip) (14, 24, 25). The results are generally consistent with the functions of these modifications inferred from locus-specific studies. H3K4me2 is involved in gene activation and is preferentially localized to endogenous genes but depleted in transposons (14). In contrast, H3K9me2 overlaps almost exclusively with transposons and other repeats, consistent with its primary function in transposon silencing (14, 24). A genome-wide

profiling of H3K27me3 revealed the presence of this modification at a large number of genes (~4400), most of which are highly tissue-specific and transcriptionally silent in the tissues assayed (young seedling) (Fig. 1 and Fig. 2A) (24, 25). It is possible that H3K27me3-mediated repression may be generally involved in the maintenance of tissue-specific gene expression patterns. The function of H3K9me3 is not yet characterized in plants. Unlike H3K9me2, H3K9me3 appears to be excluded from repetitive sequences and instead localizes to genes, but it does not seem to overlap substantially with either H3K4me2 or H3K27me3 (24).

Small RNAs

Four major endogenous RNAi pathways have been described in *Arabidopsis*. The microRNA

(miRNA), transacting siRNA (ta-siRNA), and natural-antisense siRNA (nat-siRNA) pathways mainly function at the posttranscriptional level through mRNA degradation and/or translation inhibition (26, 27). In contrast, the siRNA pathway is involved in gene silencing both transcriptionally by directing DNA methylation and posttranscriptionally by guiding mRNA cleavage (12).

Millions of 21- to 24-nucleotide (nt) siRNAs have been cloned and sequenced from wild-type *Arabidopsis* plants and siRNA pathway mutants (28–37). Most of these studies generated not only sequence information necessary to map the siRNAs back to their originating genomic loci, but also the length information of siRNAs that is indicative of the processing enzymes involved (e.g., DICER-LIKE enzymes, DCLs). The majority of the siRNAs (>90%) are produced from double-

stranded RNA (dsRNA) precursors generated by RNA polymerase IV isoform a (Pol IVa) and RNA-dependent RNA polymerase 2 (RDR2). These dsRNA precursors are then processed by DCL3 to 24-nt siRNAs (with partially redundant contributions from DCL2 and DCL4) and become preferentially associated with ARGONAUTE4, which then interacts with Pol IVb to direct DRM1/2-mediated CHH methylation (28, 29, 31, 36–38). Most of these siRNAs are derived from genomic loci corresponding to transposons with high levels of CHH DNA methylation, and very few are found in protein-coding genes (Fig. 1 and Fig. 2A) (21, 29–31).

Independent siRNA profiles generated from the same tissue type (i.e., inflorescence) are remarkably similar (30–37). In addition, when key components of the siRNA pathway were eliminated and then reintroduced (e.g., Pol IV and RDR2), restoration of a siRNA population with the original composition immediately occurred (31, 39). These observations suggest that the targeting of Pol IVa to its “sites of action” is highly reproducible. On the other hand, important differences exist between the siRNAs that accumulate in seedlings and inflorescence, which suggests that the targeting of Pol IVa may be developmentally regulated (30). Specific sequence motifs near the boundaries of Pol IVa-transcribed regions (where transcription initiation might occur) has yet to be reported. It is also attractive to speculate that certain epigenetic marks colocalizing with siRNAs might be involved in Pol IVa targeting. However, mutations in other epigenetic pathways that eliminate siRNA production have not been identified.

The majority of the remaining siRNAs (<10%, mostly 21 nt long) are derived from genomic regions corresponding to inverted repeats, independently of Pol IVa and RDR2 (29, 31, 32, 37). These siRNAs are most likely processed by DCL1 from single-stranded hairpin RNA precursors generated by Pol II, a process resembling miRNA biogenesis. However, most of these siRNAs are present at relatively low levels, and it is not yet clear whether they target endogenous mRNAs to regulate plant development. Nevertheless, these siRNAs might represent “evolving miRNAs” that could eventually acquire endogenous gene targets (29, 31, 32, 35).

Out of *Arabidopsis*?

Most known genes involved in epigenetic pathways are shared between monocots and dicots (many are even present in moss), and loss-of-function mutations in corresponding genes lead to similar molecular defects (40). In addition, the chromosomal-level distribution of several histone modifications is similar between maize and *Arabidopsis* (41). However, recent data suggest that, unlike in *Arabidopsis*, genic DNA methylation in rice appears to be enriched in the promoter regions of endogenous genes and associated

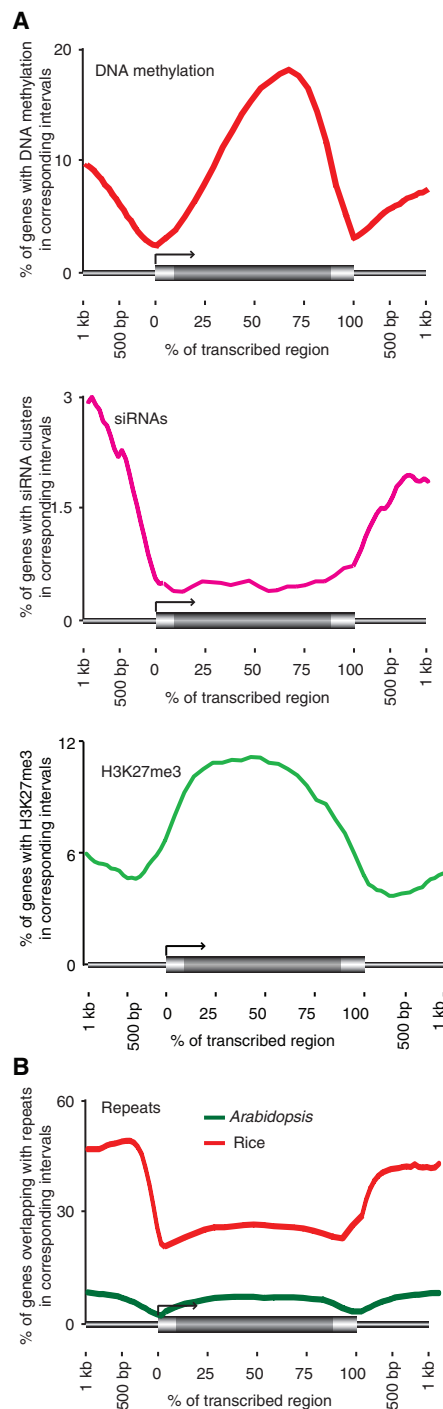


Fig. 2. (A) Distribution of DNA methylation, siRNAs, and H3K27me3 relative to *Arabidopsis* genes (18, 25, 31). One-kilobase regions upstream and downstream of each gene are divided into 50-base pair intervals, each gene is divided into 20 intervals (5% each interval), and the percentage of genes overlapping with each epigenetic mark in the corresponding regions is graphed. Thick and thin horizontal bars represent genes and intergenic regions, respectively. (B) Distribution of repetitive sequences relative to (genes in *Arabidopsis* (green) and rice (red).

with transcriptional repression (42). It is therefore possible that the mechanisms and functions of some epigenetic pathways might have diverged and acquired distinct functions during plant evolution.

The extent to which epigenetic pathways regulate gene expression in a particular species may also be affected by the genetic architecture of its genome. For example, results from *Arabidopsis* suggest that DNA methylation, siRNA, or H3K9me2 primarily regulate genes through nearby repetitive sequences (14, 18, 43–46). The close proximity of repeats to genes, although relatively rare in *Arabidopsis*, may be the norm rather than the exception in larger genomes where genes are commonly embedded in an ocean of repetitive sequences (47). This is already apparent in rice, which still has one of the “leanest” genomes in the grasses. As shown in Fig. 2B, relative to *Arabidopsis*, a significantly higher fraction of rice genes are closely associated with repetitive sequences. Consistent with this notion, although the loss of RDR2 (which primarily functions in transposon silencing) does not affect normal development in *Arabidopsis*, mutation of its homolog in maize (*Mediator of Paramutation1; MOP1*) leads to a number of development abnormalities (48–50). Finally, endogenous gene families could be subject to similar epigenetic controls as transposons because of their repetitive nature (51, 52). Such gene families may be present at higher copy numbers in larger plant genomes, particularly in relatively recent polyploids (e.g., *Brassica* or wheat). It therefore might be reasonable to expect that epigenetic silencing pathways play much broader roles in regulating gene expression in plants with larger and more complex genomes.

Conclusions

Recent high-throughput profiling studies in *Arabidopsis* have painted a picture of epigenetic compartmentation, where the two major fractions of the genome are associated with and regulated by different epigenetic mechanisms. That is, genes are regulated by pathways such as H3K27me3, H3K4me2, and miRNAs/ta-siRNAs/nat-siRNAs, whereas transposons and other repeats are silenced by DNA methylation, H3K9me2, and siRNAs. Such a functional distinction, however, is blurred when the two genetic fractions overlap, which occurs much more frequently in larger and more complex genomes. The function(s) of DNA methylation that are enriched in different fractions of the gene space in *Arabidopsis* (3' half of transcribed regions) and rice (promoter regions), as well as DNA methylation by the DEMETER (DME) family of DNA glycosylases (53), are not yet understood and warrant further functional studies.

Although increasingly comprehensive, such an epigenomic picture remains static. Relatively little is known about how the plant epigenome changes in response to developmental or envi-

ronmental cues. A particularly interesting question may be how mechanisms that evolved to stably silence transposons could offer the flexibility required for the developmental regulation of endogenous genes. In addition, we do not yet have a clear understanding of the nature and the maintenance of the boundaries separating epigenetically distinct chromatin compartments. In some cases, genetic landmarks (such as the transcription unit) may serve as borders; in other cases, the balancing acts of opposing epigenetic mechanisms may help to stably maintain the epigenetic landscape of plant genomes.

References and Notes.

1. A. Bird, *Nature* **447**, 396 (2007).
2. M. Gehring, S. Henikoff, *Biochim. Biophys. Acta* **1769**, 276 (2007).
3. I. R. Henderson, S. E. Jacobsen, *Nature* **447**, 418 (2007).
4. M. Matzke, T. Kanno, B. Huettel, L. Daxinger, A. J. Matzke, *Curr. Opin. Plant Biol.* **10**, 512 (2007).
5. I. Vaillant, J. Paszkowski, *Curr. Opin. Plant Biol.* **10**, 528 (2007).
6. J. Pfluger, D. Wagner, *Curr. Opin. Plant Biol.* **10**, 645 (2007).
7. Arabidopsis Genome Initiative, *Nature* **408**, 796 (2000).
8. E. J. Finnegan, W. J. Peacock, E. S. Dennis, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8449 (1996).
9. A. M. Lindroth *et al.*, *Science* **292**, 2077 (2001).
10. L. Bartee, F. Malagnac, J. Bender, *Genes Dev.* **15**, 1753 (2001).
11. X. Cao, S. E. Jacobsen, *Curr. Biol.* **12**, 1138 (2002).
12. S. W. Chan *et al.*, *Science* **303**, 1336 (2004).
13. D. Zilberman, S. Henikoff, *Development* **134**, 3959 (2007).
14. Z. Lippman *et al.*, *Nature* **430**, 471 (2004).
15. R. K. Tran *et al.*, *Curr. Biol.* **15**, 154 (2005).
16. M. W. Vaughn *et al.*, *PLoS Biol.* **5**, e174 (2007).
17. D. Zilberman, M. Gehring, R. K. Tran, T. Ballinger, S. Henikoff, *Nat. Genet.* **39**, 61 (2007).
18. X. Zhang *et al.*, *Cell* **126**, 1189 (2006).
19. J. Reinders *et al.*, *Genome Res.* **18**, 469 (2008).
20. O. Mathieu, J. Reinders, M. Caiikovski, C. Smathajitt, J. Paszkowski, *Cell* **130**, 851 (2007).
21. S. J. Cokus *et al.*, *Nature* **452**, 215 (2008).
22. X. Zhang, S. E. Jacobsen, *Cold Spring Harb. Symp. Quant. Biol.* **71**, 439 (2006).
23. S. E. Jacobsen, E. M. Meyerowitz, *Science* **277**, 1100 (1997).
24. F. Turck *et al.*, *PLoS Genet* **3**, e86 (2007).
25. X. Zhang *et al.*, *PLoS Biol.* **5**, e129 (2007).
26. F. Vazquez, *Trends Plant Sci.* **11**, 460 (2006).
27. E. J. Chapman, J. C. Carrington, *Nat. Rev. Genet.* **8**, 884 (2007).
28. Y. Qi *et al.*, *Nature* **443**, 1008 (2006).
29. C. Lu *et al.*, *Genome Res.* **16**, 1276 (2006).
30. C. Lu *et al.*, *Science* **309**, 1567 (2005).
31. X. Zhang, I. R. Henderson, C. Lu, P. J. Green, S. E. Jacobsen, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4536 (2007).
32. R. Rajagopalan, H. Vaucheret, J. Trejo, D. P. Bartel, *Genes Dev.* **20**, 3407 (2006).
33. M. D. Howell *et al.*, *Plant Cell* **19**, 926 (2007).
34. K. D. Kasschau *et al.*, *PLoS Biol.* **5**, e57 (2007).
35. N. Fahlgren *et al.*, *PLoS ONE* **2**, e219 (2007).
36. I. R. Henderson *et al.*, *Nat. Genet.* **38**, 721 (2006).
37. R. A. Moshier, F. Schwach, D. Studholme, D. C. Baulcombe, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3145 (2008).
38. M. El-Shami *et al.*, *Genes Dev.* **21**, 2539 (2007).
39. S. W. Chan, X. Zhang, Y. V. Bernatavichute, S. E. Jacobsen, *PLoS Biol.* **4**, e363 (2006).
40. Y. Ding *et al.*, *Plant Cell* **19**, 9 (2007).
41. J. Shi, R. K. Dawe, *Genetics* **173**, 1571 (2006).
42. X. Li *et al.*, *Plant Cell* **20**, 259 (2008).
43. H. Saze, T. Kakutani, *EMBO J.* **26**, 3641 (2007).
44. W. J. Soppe *et al.*, *Mol. Cell* **6**, 791 (2000).
45. R. Martienssen, A. Baron, *Genetics* **136**, 1157 (1994).
46. R. K. Slotkin, R. Martienssen, *Nat. Rev. Genet.* **8**, 272 (2007).
47. K. M. Devos, J. Ma, A. C. Pontaroli, L. H. Pratt, J. L. Bennetzen, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 19243 (2005).
48. J. E. Dorweiler *et al.*, *Plant Cell* **12**, 2101 (2000).
49. M. R. Woodhouse, M. Freeling, D. Lisch, *PLoS Biol.* **4**, e339 (2006).
50. M. Alleman *et al.*, *Nature* **442**, 295 (2006).
51. T. L. Stokes, B. N. Kunkel, E. J. Richards, *Genes Dev.* **16**, 171 (2002).
52. J. Bender, G. R. Fink, *Cell* **83**, 725 (1995).
53. J. Penterman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6752 (2007).
54. I sincerely apologize to authors whose work is not cited here because of space limitations. I thank S. Wessler, K. Dawe, and S. Jacobsen for their critical reading of the manuscript and helpful suggestions, and N. Jiang for sharing unpublished data on rice repeats. Supported by a grant from the University of Georgia.

10.1126/science.1153996

PERSPECTIVE

Extending Genomics to Natural Communities and Ecosystems

Thomas G. Whitham,^{1,2} Stephen P. DiFazio,³ Jennifer A. Schweitzer,⁴ Stephen M. Shuster,^{1,2} Gery J. Allan,^{1,2} Joseph K. Bailey,⁴ Scott A. Woolbright^{1,2}

An important step in the integration of ecology and genomics is the progression from molecular studies of relatively simple model systems to complex field systems. The recent availability of sequenced genomes from key plants is leading to a new understanding of the molecular drivers of community composition and ecosystem processes. As genome sequences accumulate for species that form intimate associations in nature, a detailed view may emerge as to how these associations cause changes among species at the nucleotide level. This advance could dramatically alter views about the structure and evolution of communities and ecosystems.

The emerging field of community and ecosystem genetics has shown that genetic variation in one species can influence the composition of associated communities (the interacting species that live in a particular area) and

the functioning of ecosystems (the biotic community and its abiotic environment) (1, 2). To date, this field has focused mainly on the genetics of foundation plant species—those that structure their ecosystems by creating locally stable conditions and providing specific resources for diverse organisms (3). When a foundation species' genotype influences the relative fitness of other species, it constitutes an indirect genetic interaction (4), and when these interactions change species composition and abundance among individual tree genotypes, they cause community and ecosystem phenotypes to become statistically distinct from one another (2). Figure 1

shows some of the community and ecosystem phenotypes that have been quantified for the foundation tree species, *Populus angustifolia*. These include canopy arthropod communities, soil microbial communities, soil nutrient pools, and interactions between an insect herbivore and its avian predators. Common garden studies show that these phenotypes are all heritable (4–6) in the broad sense (i.e., additive, dominant, and epistatic genetic effects influence these phenotypes). For example, genes control the susceptibility to the leaf-galling aphid, *Pemphigus betae*, in *P. angustifolia*. The presence or absence of this insect is determined by susceptible or resistant tree genotypes. This interaction predictably affects other trophic levels and alters the composition of a diverse community of fungi, insects, spiders, and avian predators (2, 5). This work suggests that indirect genetic interactions among relatively few species (e.g., a tree, herbivores, mutualists, and/or pathogens) and their environment may structure the composition and abundance of a much larger community of organisms and define ecosystem processes such as nutrient cycling. However, much work needs to be undertaken to understand the genomic components underlying these phenotypes, such as additive genes and epigenetic effects. Here, we examine how a genomics perspective may aid our understanding of communities and ecosystems by focusing on the indirect genetic interactions that arise among a few foundation species rather than attempting to

¹Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA. ²Merriam-Powell Center for Environmental Research, Northern Arizona University, Flagstaff, AZ 86011, USA. ³Department of Biology, West Virginia University, Morgantown, WV 26506–6057, USA. ⁴Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA.

*To whom correspondence should be addressed. E-mail: Thomas.Whitham@nau.edu