Genomes and Epigenomes

Beyond the double helix...

Alain RIVAL

Palm Development Group
Cirad BioS
Montpellier
France.

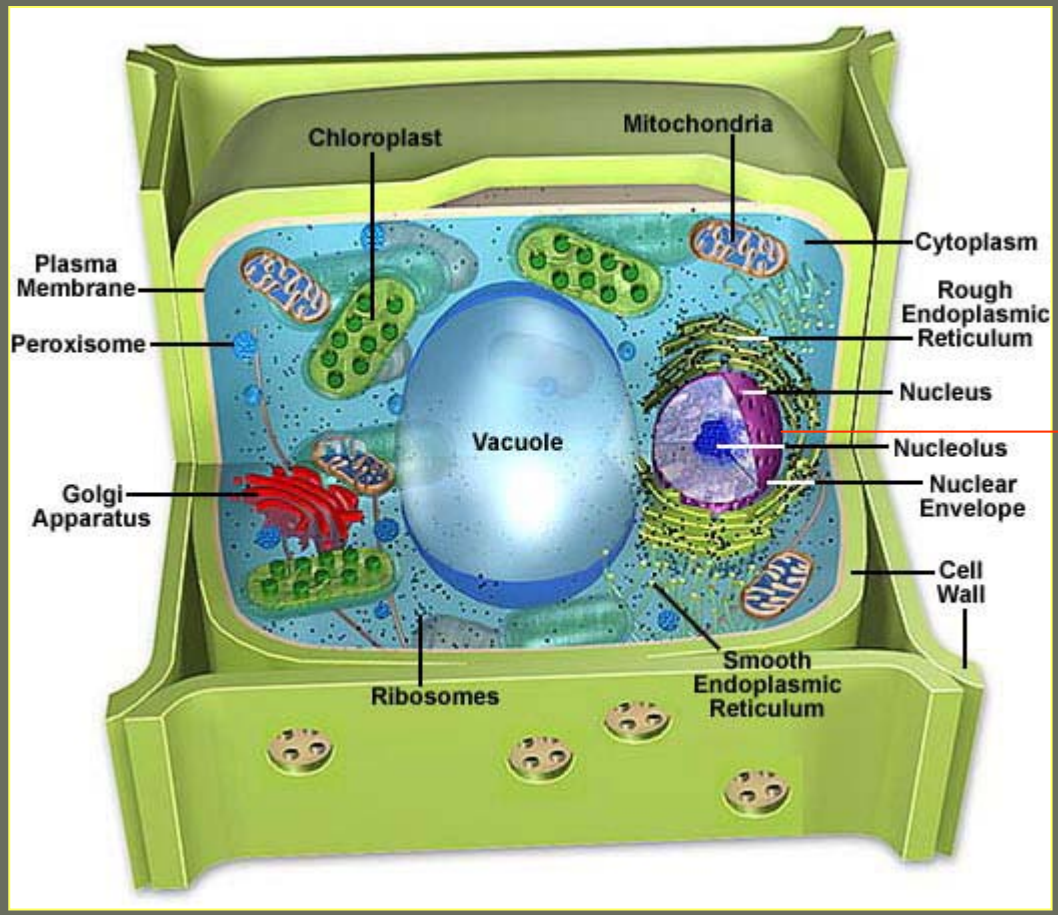
What is Epigenetics?

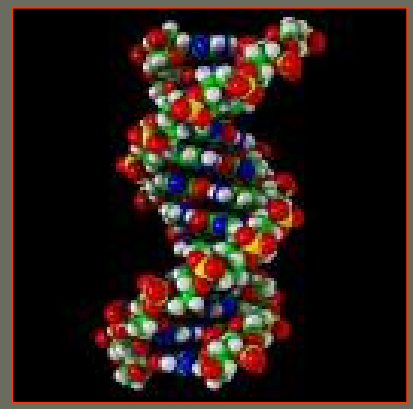
« The interactions of genes with their environment, which bring the phenotype into being ».

Waddington, C. (1942) The Epigenoptype. Endeavour, 1, 18–20

"Changes in gene expression not attributable to nucleotide sequence variation".

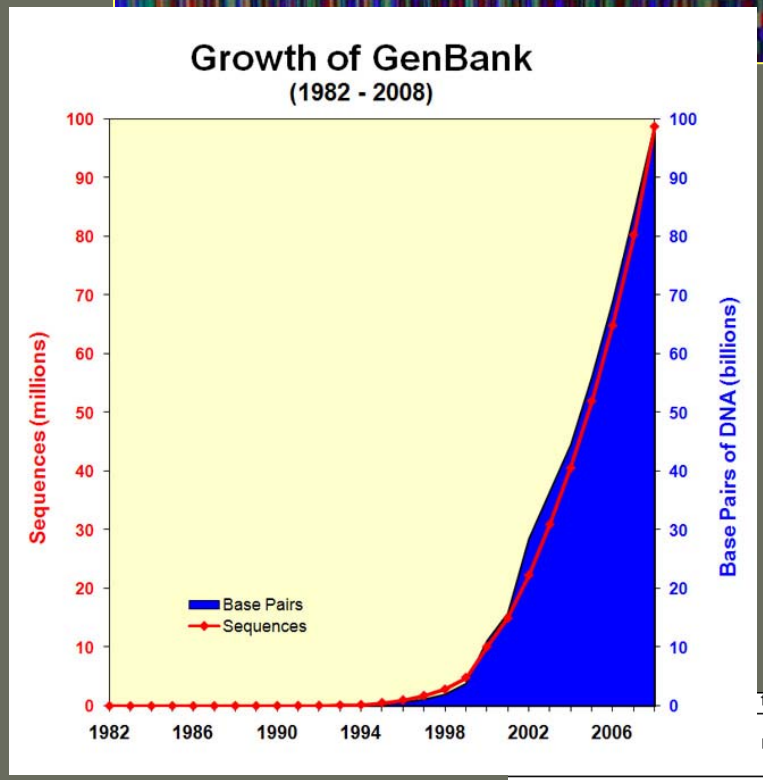
The Cell and the Genome





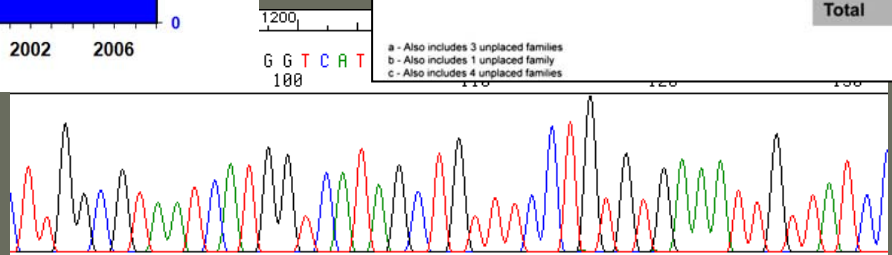
The cell nucleus contains the majority of the plant cell's genetic material

Genomics: sequencing plant genomes



	Order	Family	Species	Represented Families
Eurosid I (Fabids)	8	64	*Glycine max, Lotus japonicus, Medicago, Populus trichocarpa*	2
Eurosid II (Malvids)	3	34	*Arabidopsis lyrata, A. thaliana, Brassica oleracea, B. rapa, Capsella rubella, Carica papya*	2
Other Rosids	5	35	*Eucalyptus globulus. Vitis vinifera*	2
Euasterid I (Lamiids)	4	40	*Lycopersicon esculentum, Mimulus guttatus, Solanum tuberosum, Triphysaria versicolor*	3
Euasterid II (Campanulids)	4	35		0
Other Asterids	2	30		0
Other core eudicots	4	24		0
Early diverging dicots	2[a]	12		0
Magnoliids	4	18		0
Commelinids	3[b]	29	*Musa acuminata, Oryza sativa, Sorghum bicolor, Triticum aestivum, Zea mays*	2
Other monocots	8	53		0
Basal Angiosperms	1[c]	5		0
Gymnosperms	4	11		0
Ferns	16	55		0
Lycophytes	3	3	*Selaginella moellendorfii*	1
Bryophytes	13	158	*Physcomitrella patens*	1
Green Algae			*Chlamydomonas reinhardtii, Micromonas pusilla, Ostreococcus tauri*	
Total	84	606		13

a - Also includes 3 unplaced families
b - Also includes 1 unplaced family
c - Also includes 4 unplaced families

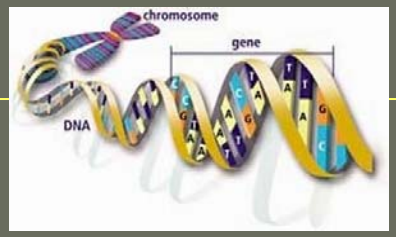


Beyond the sequencing of plant genomes

- Several new sequencing technologies are emerging that have the potential to provide increases in throughput and reductions in cost

- Companies such as 454 Life Sciences, Solexa, and Helicos Biosciences all have competing technologies, vying to be widely adopted for the next generation of sequencing machines

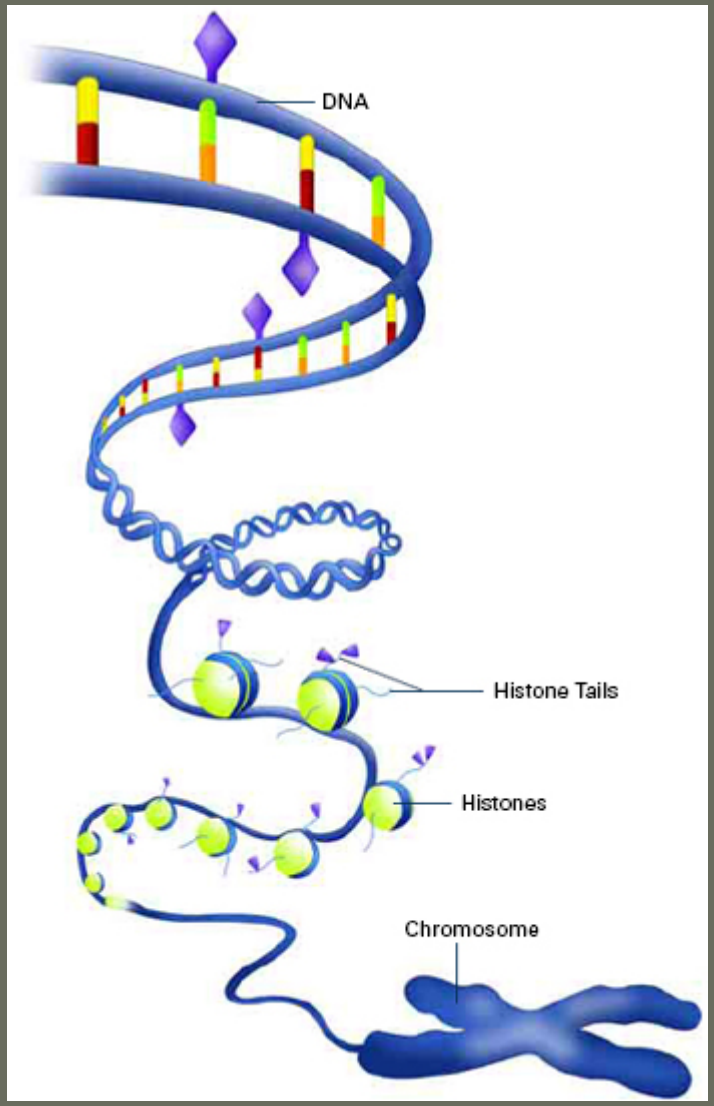
- The success of genome sequencing project has created wide-spread interest in exploring epigenomes in order to elucidate how the genome executes the information it holds.

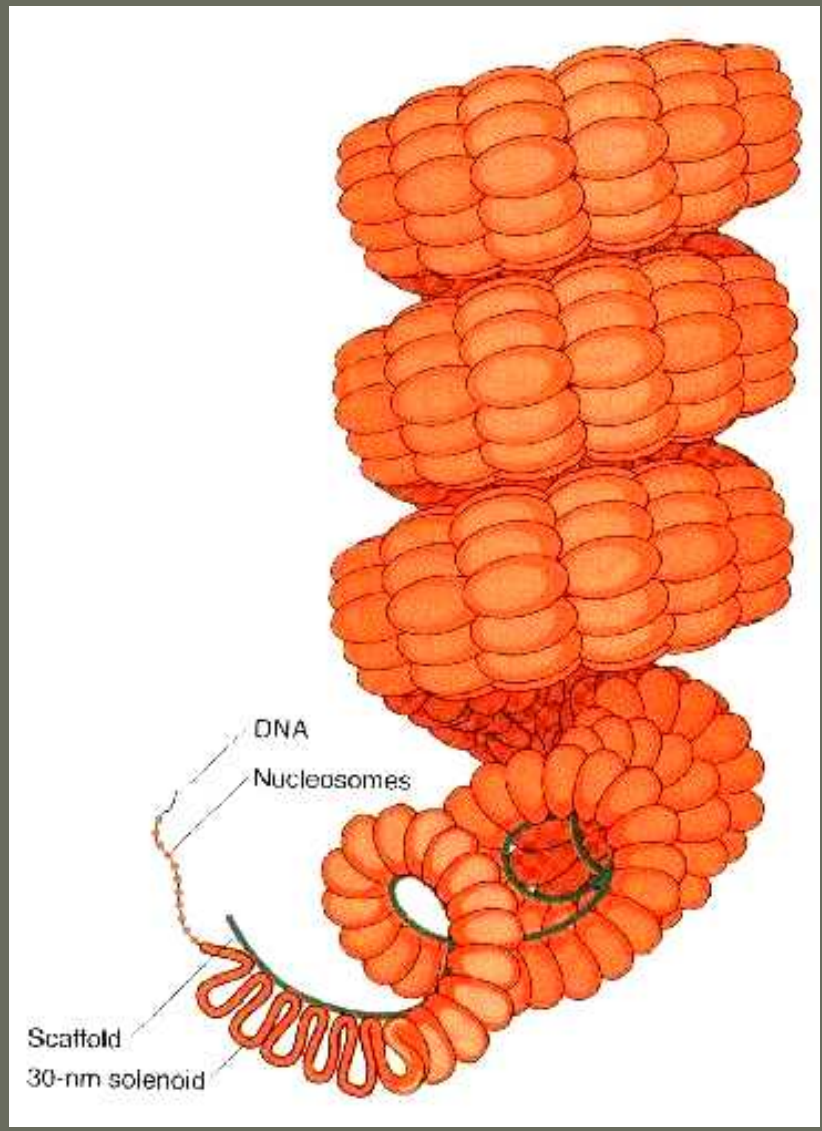


An organised DNA Packaging

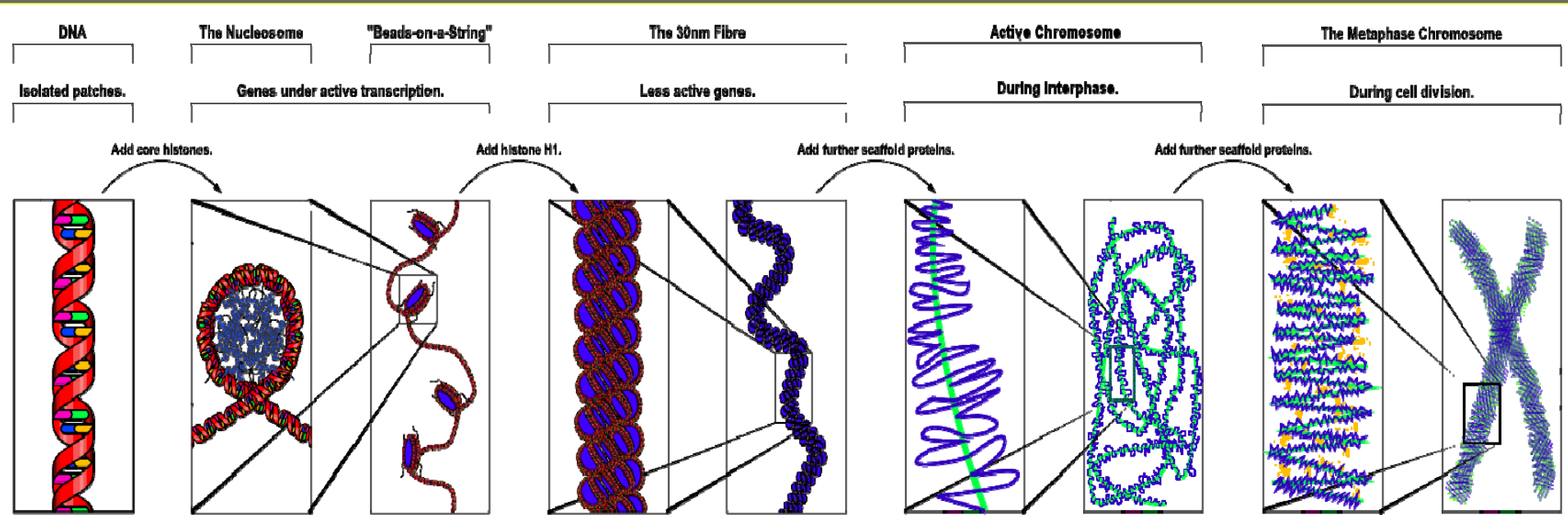
- A mammal diploid cell contains in average 6000 Mb of DNA, corresponding to 1.8 meters, which is packed within a nucleus of 6 microns diameter, following a strongly organized packaging
- Each chromosome yields one single molecule of linear DNA which is folded several times and distributed over the centromere
- Within the nucleus, DNA is never encountered under a free form but associated to other molecules, principally histones. Histones are small basic proteins which are present with the same proportion as DNA.
- Other proteins associated to DNA are non histones acidic proteins, which account for 10 to 30% of the whole structure.
- This DNA-Protein complex is called chromatin (as it is easily stained) and it constitutes the chromosomes

An organised DNA Packing





DNA increasing orders of packing

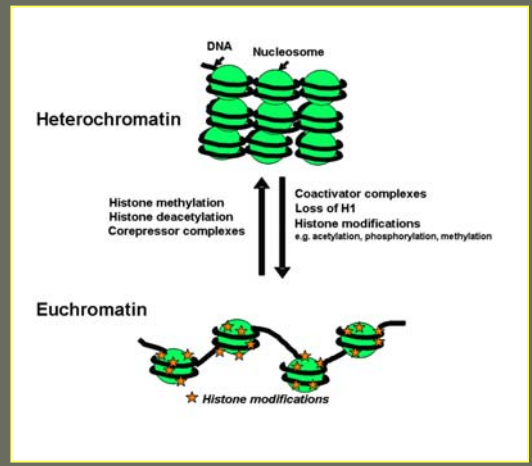


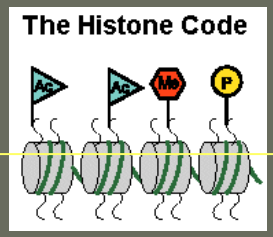
Chromatin structure and gene activity

- **Heterochromatin** refers to regions of the genome that have low gene density, contain satellite repeat elements and are late replicating.

- **Euchromatin** is a lightly packed form of chromatin that is rich in gene concentration, and is often under active transcription. Euchromatin comprises the most active portion of the genome within the cell nucleus.

- Heterochromatin and euchromatin are associated with distinct DNA methylation and histone modification patterns that correlate with particular states of gene activity

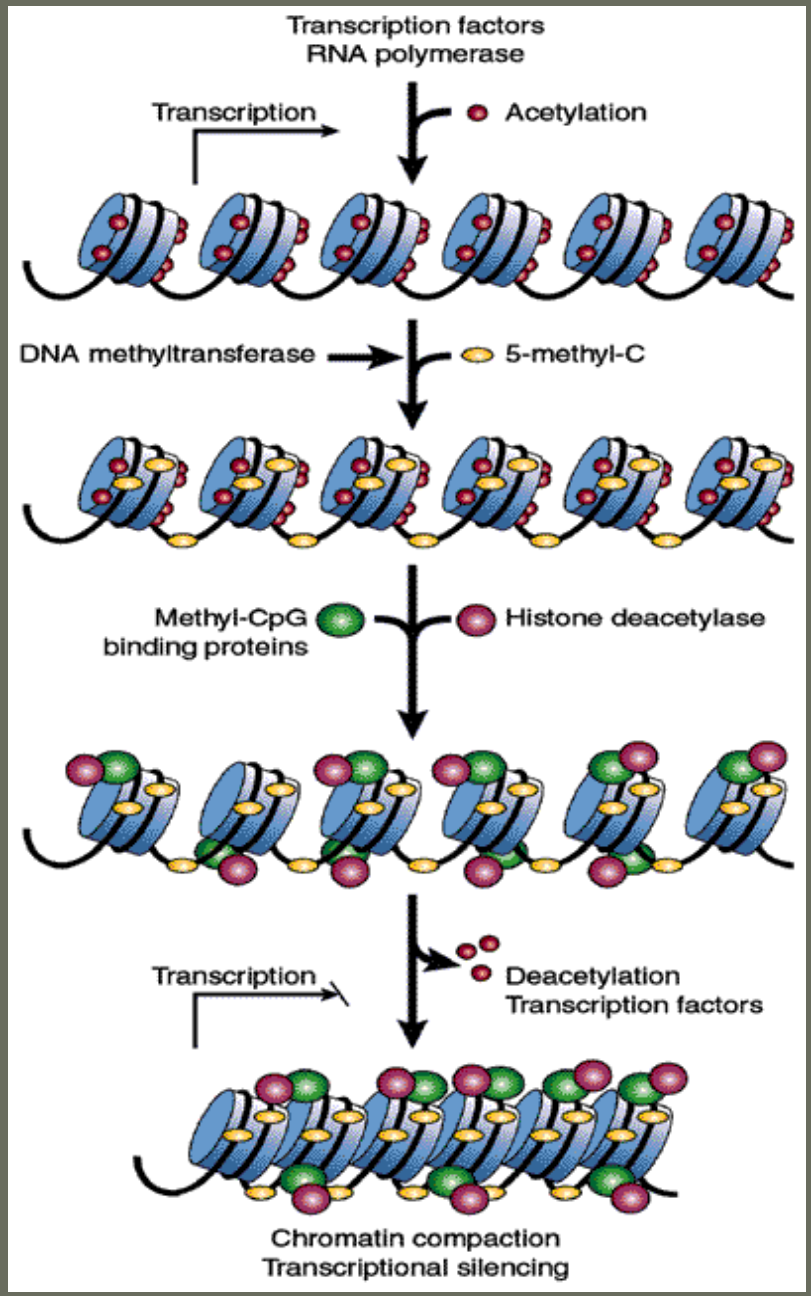




Chromatine structure and gene activity

- Acetylation of histones loosens the chromatin, facilitating replication and transcription.
- Methylated histones hold the DNA more tightly, thus restricting access and impeding transcription.
- Methylation of Lysine 4 and Lysine-27 on Histone3 may be involved in development.
- Specifically, there is fewer methylated Lysine 27 in the chromatin of differentiated cells.
- Lysine-4 methylation acts to promote transcription by recruiting nucleosome remodeling enzymes and histone acetylase.

Epigenetic prints on chromatin



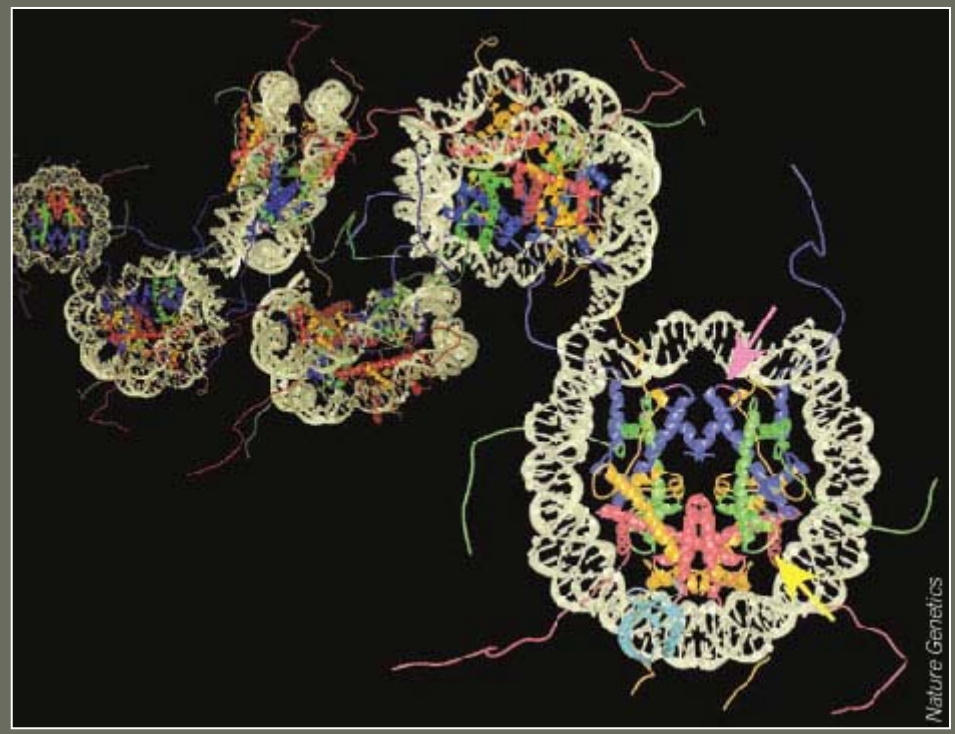
Transcriptionally active chromatin regions tend to be hyperacetylated and hypomethylated.

If a region of DNA or a gene is destined for silencing, chromatin remodeling enzymes such as histone deacetylases and ATP-dependent chromatin remodelers likely begin the gene silencing process.

One or more of these activities may recruit DNA methyltransferase resulting in DNA methylation, followed finally by recruitment of the methyl-CpG binding proteins.

The region of DNA will then be heritably maintained in an inactive state.

The Da-Histone Code ...



Covalent histone modifications, including acetylation, methylation, phosphorylation and ubiquitylation, represent a 'histone code' that controls the transcriptional status of genes.

From genome to epigenome

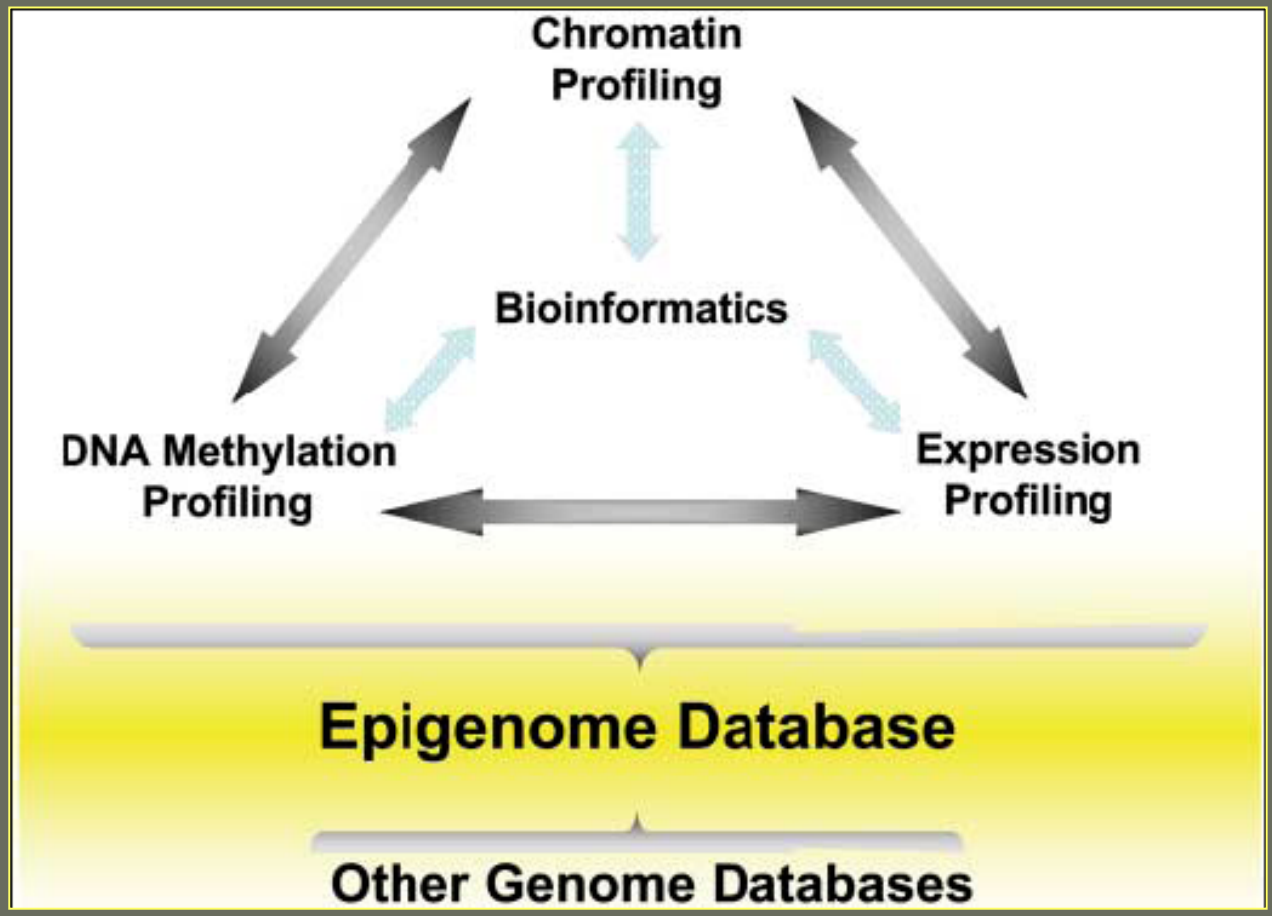
Epigenomics is the large scale study of epigenetic marks on the genome including:

- Covalent modifications of histone tails
  (acetylation, methylation, phosphorylation, ubiquitination)
- DNA methylation.
- Small RNAs

Epigenetic components are all amenable to genome-wide studies

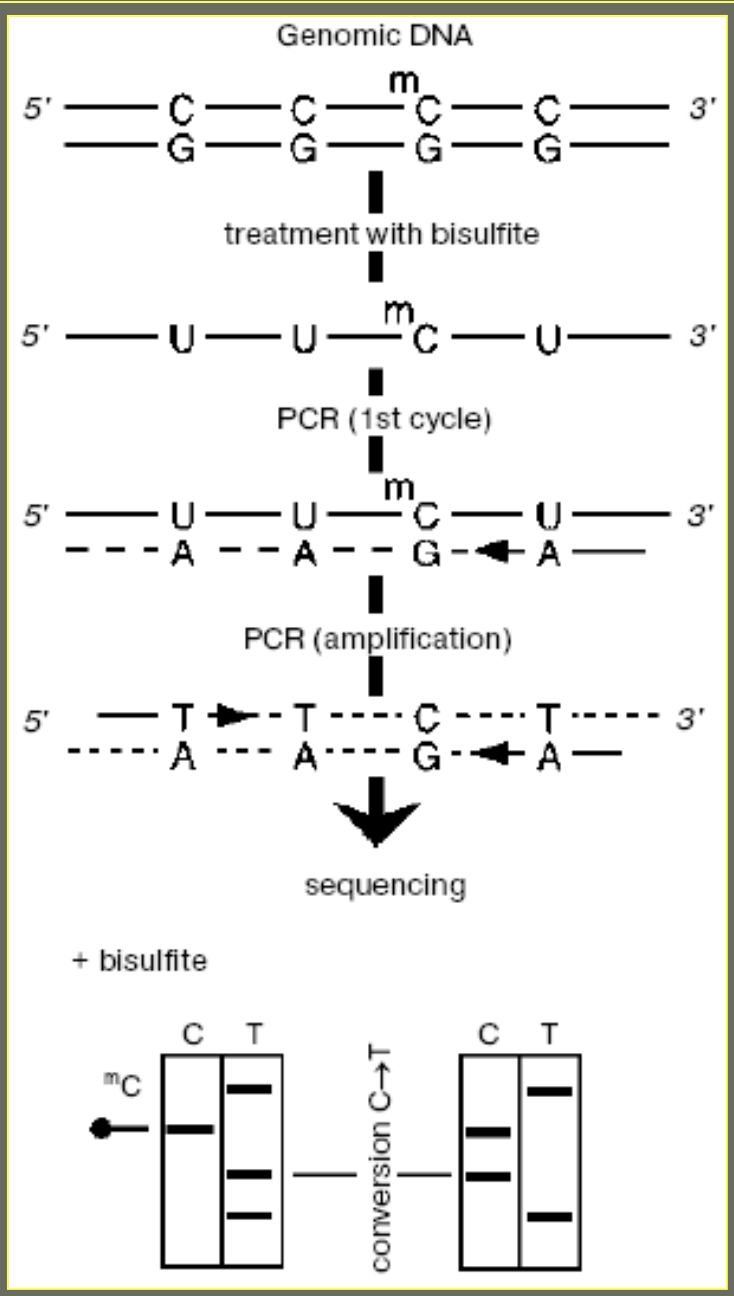
Integrated approaches that correlate gene expression with DNA methylation and chromatin profiles are being designed.

From genome to epigenome



Murell et al, Human Molecular Genetics 2005 14(Review Issue 1):

Analysing DNA Methylation: the bisulfite reaction



- Cytosine is converted stoichiometrically to uracil, but 5-methylcytosine remains nonreactive.

- PCR amplification of the region of interest in the bisulfite-reacted DNA yield a fragment in which all uracil, formerly cytosine, and thymine residues have been amplified as thymine and only 5-methylcytosines have been amplified as cytosine

- Sequences will provide methylation maps of single DNA strands from individual DNA molecules in the original genomic DNA sample.

- The position of each 5-methylcytosine will be given by a positive band on a sequencing gel

DNA Methylation profiling

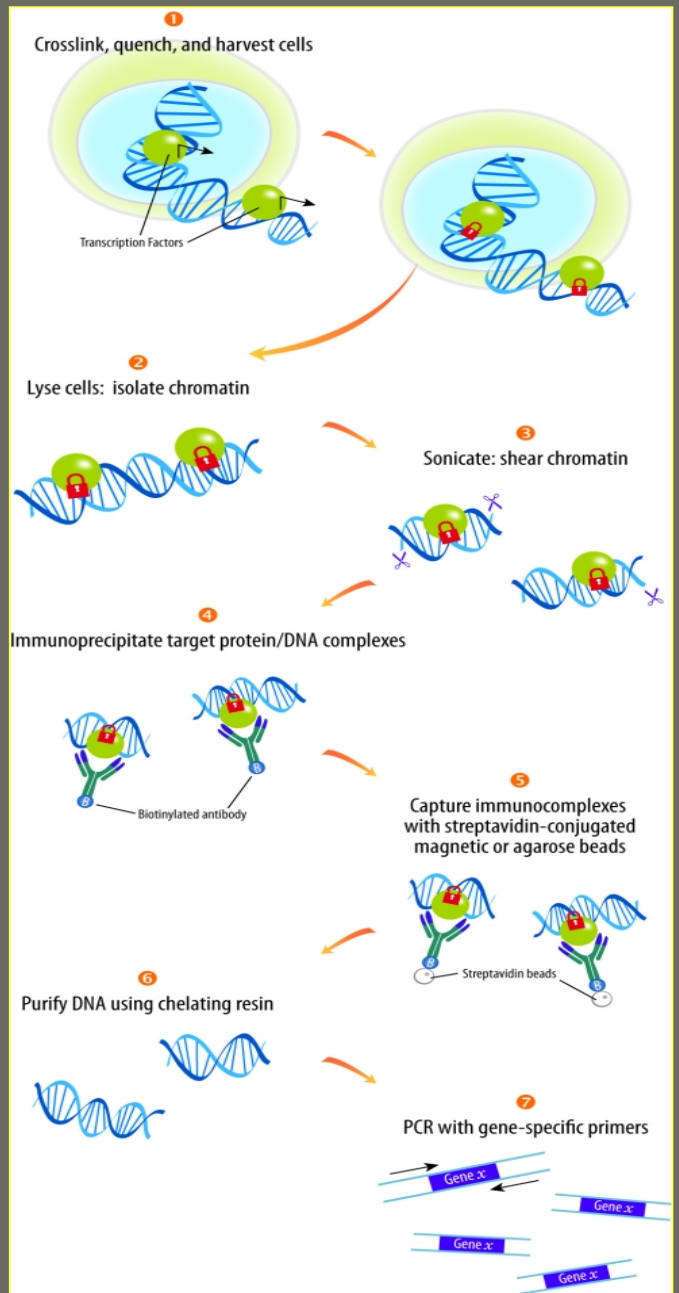
➢ Array-based methylation profiling

Conventional sodium bisulphite treatment of genomic DNA, followed by PCR amplification of regions of interest (300–400 bp in size).
These products are hybridized to custom microarrays that contained probes to discriminate converted versus unconverted cytosines at the CpG site of interest.

➢ Sequencing-based methylation profiling

Bisulphite sequencing has been adapted for pyrosequencing, which is based on quantitative, real-time sequencing-by-synthesis. In this approach, genetic polymorphisms such as single nucleotide polymorphisms (SNPs) and epigenetic polymorphisms such as methylation variable positions (MVPs) can be analysed in a single assay.

Chromatin ImmunoPrecipitation (ChIP)

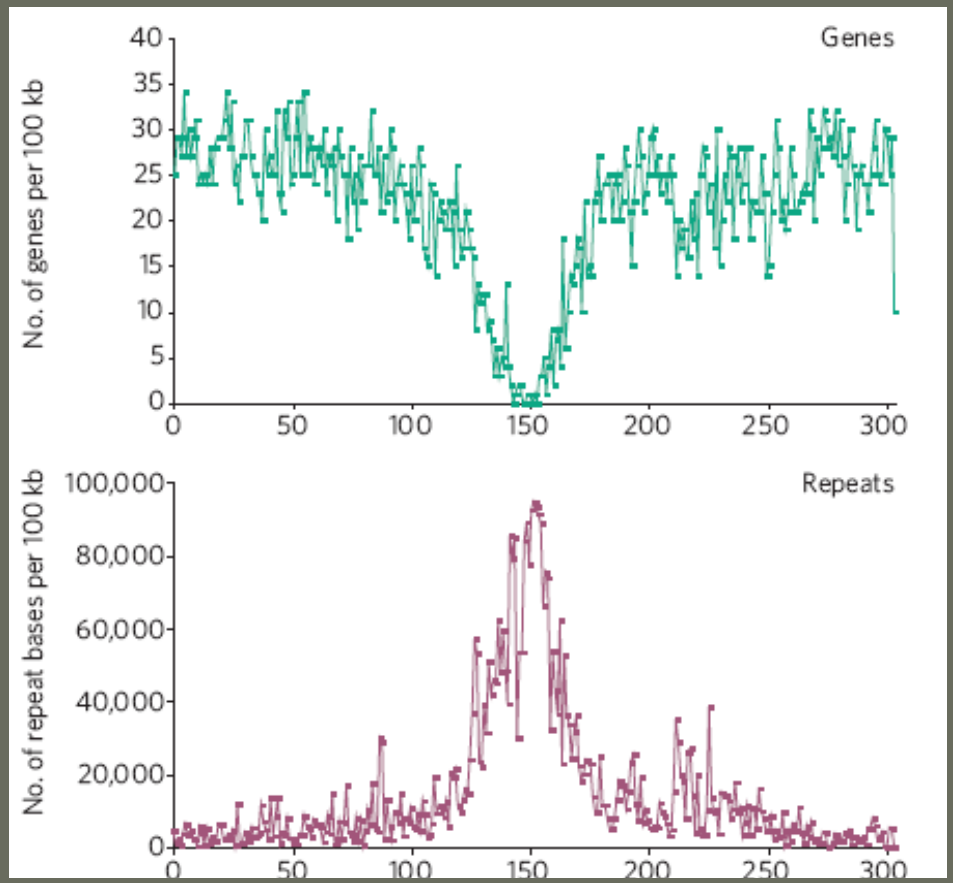


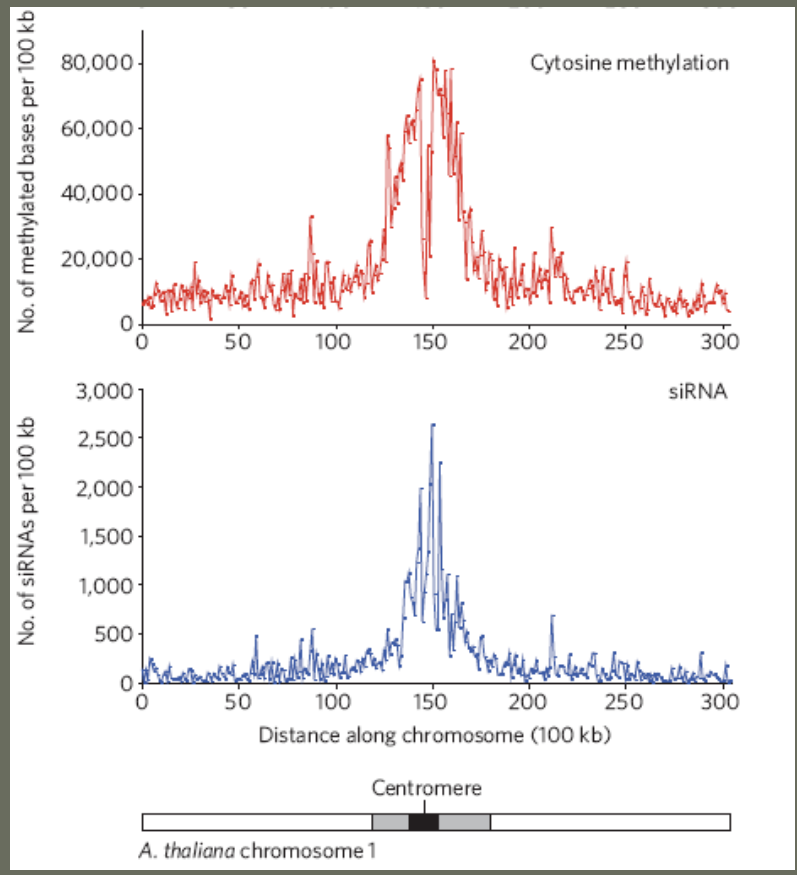
- The chromatin state of a given DNA sequence influences transcriptional activity and is regulated by potentially reversible covalent modifications of DNA and histones.

- Histone modifications at conserved lysine and arginine residues within the flexible N-terminal tails, such as phosphorylation, acetylation and methylation, specify a code which serves as an interaction platform with specific domains of chromatin-associated proteins.

- The immunoprecipitation (IP) of crosslinked chromatin with antibodies specific for certain histone modifications (chromatin immunoprecipitation; ChIP), followed by PCR to detect a potential enrichment or depletion of a DNA sequence of interest within IP fractions, constitutes an elegant and direct method to query specific chromatin states of individual genes

The epigenetic landscape of *A. thaliana*

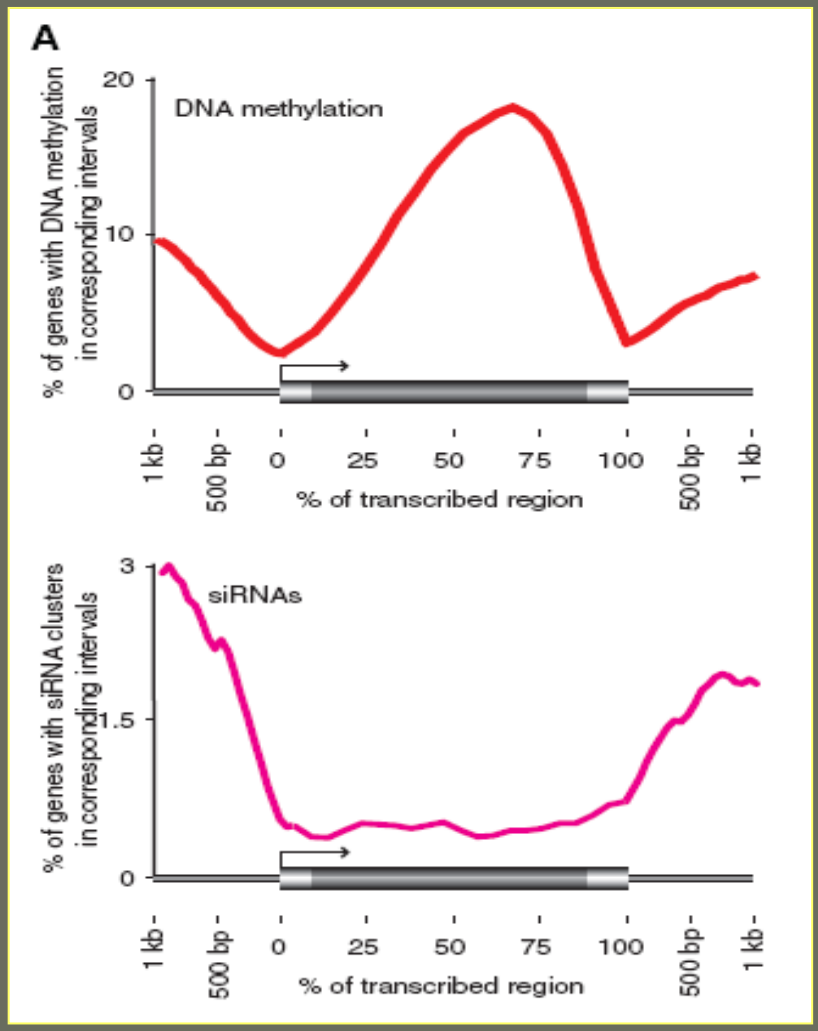




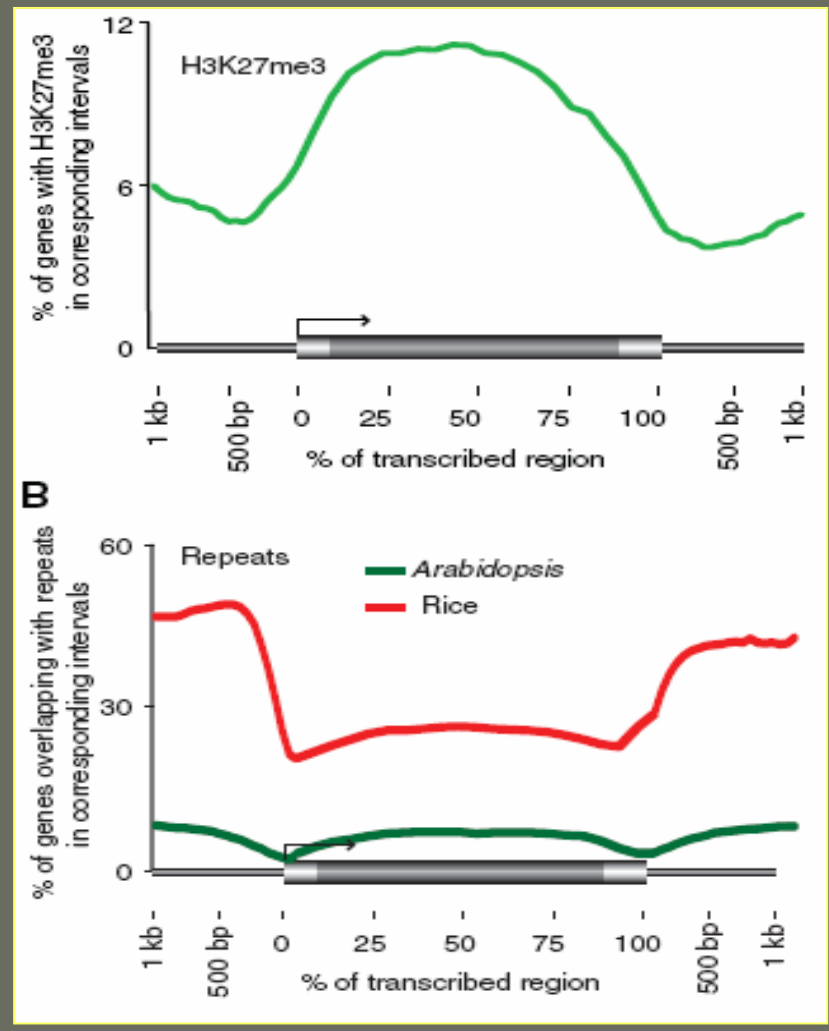
The relative abundance of genes, repeats, cytosine methylation and siRNAs is shown for the length of *A. thaliana* chromosome 1, which is ~30 Mb long.

Henderson & Jacobson, Nature 447, 418 (2007)

Positioning relative to *Arabidopsis* genes



(A) Distribution of DNA methylation, siRNAs, and H3K27me3 relative to *Arabidopsis* genes.



(B) Distribution of repetitive sequences relative to genes in Arabidopsis (green) and rice (red).

Thick and thin horizontal bars represent genes and intergenic regions, respectively.

Zhang, Science 320, 489 (2008)

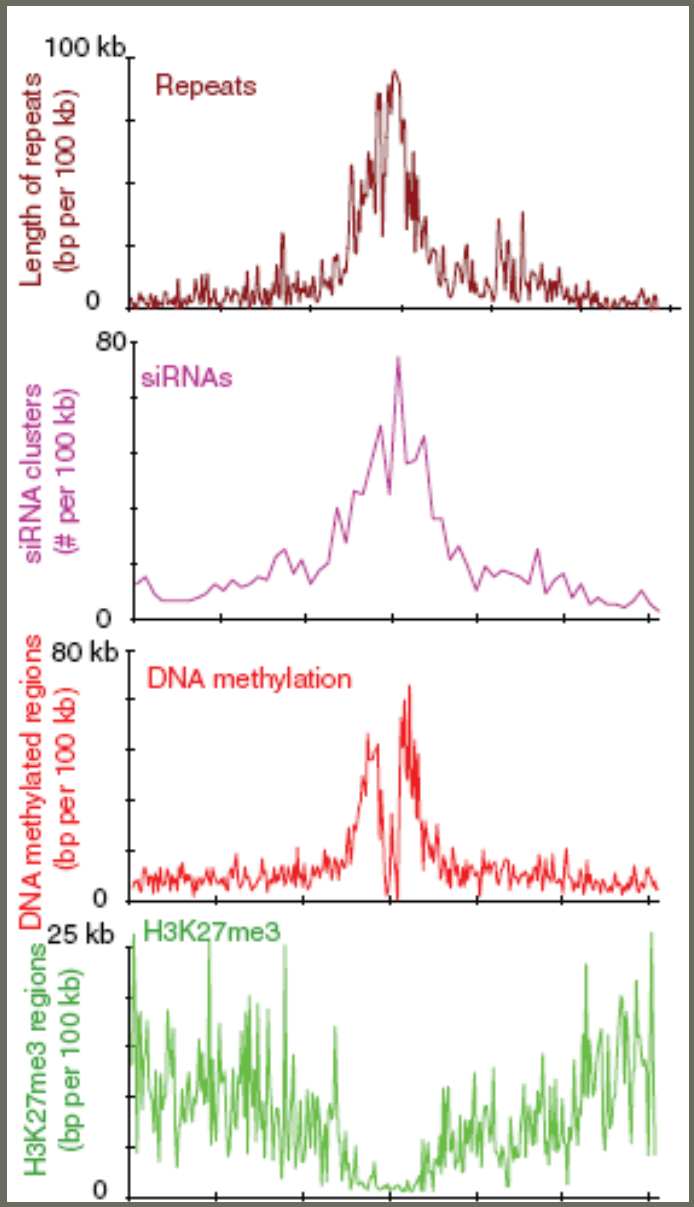
Distribution patterns and transcription activity

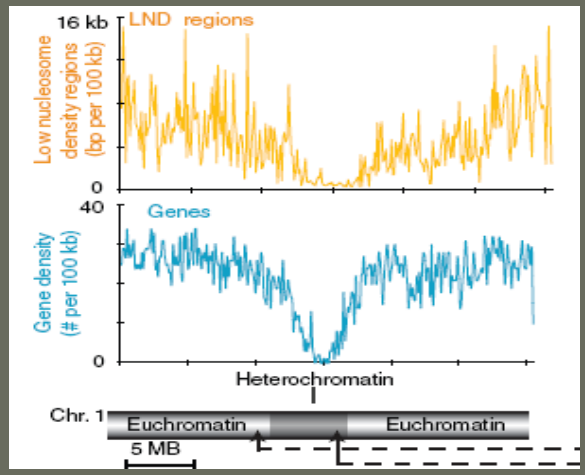
Detailed distribution patterns and transcription activity (vertical blue bars) in a gene-rich region (A) and a repeat-rich region (B).

Red boxes: genes; Arrows indicate the direction of transcription.

Zhang, Science 320, 489 (2008)

Motiv density along chromosome





Distribution of genes, repetitive sequences, DNA methylation, siRNAs, H3K27me3, and low nucleosome density (LND) regions in Arabidopsis.

The chromosomal distributions use chromosome 1 as an example. The x axis shows chromosomal position.

Zhang, Science 320, 489 (2008)

The small RNAs world

Breakthrough Online
For an expanded version of this section, with references and links, see www.sciencemag.org/content/vol298/issue5602/#special

Breakthrough

#1 The Winner

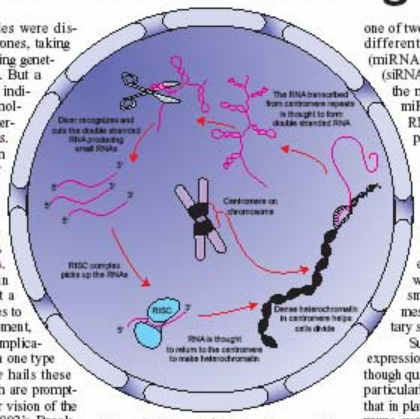
Just when scientists thought they had deciphered the roles played by the cell's leading actors, a familiar performer has turned up in a stunning variety of guises. RNA, long upstaged by its more glamorous sibling, DNA, is turning out to have star qualities of its own.

Small RNAs Make Big Splash

For decades, RNA molecules were dismissed as little more than drones, taking orders from DNA and converting genetic information into proteins. But a string of recent discoveries indicates that a class of RNA molecules called small RNAs operate many of the cell's controls. They can turn the tables on DNA, shutting down genes or altering their levels of expression. Remarkably, in some species, truncated RNA molecules literally shape genomes, carving out chunks to keep and discarding others. There are even hints that certain small RNAs might help chart a cell's destiny by directing genes to turn on or off during development, which could have profound implications for coaxing cells to form one type of tissue or another. *Science* hails these electrifying discoveries, which are prompting biologists to overhaul their vision of the cell and its evolution, as 2002's Breakthrough of the Year.

These astonishing feats are performed by short stretches of RNA ranging in length from 21 to 28 nucleotides. Their role had gone unnoticed until recently, in part because researchers, focused on the familiar larger RNA molecules, tossed out the crucial small ones during experiments. As a result, RNA has long been viewed primarily as an essential but rather dull molecule that ferries the genetic code from the nucleus to the ribosomes, the cell's protein factories, and helps assemble amino acids in the correct order during protein synthesis.

Signs that RNA might be more versatile came in the early 1990s, when biologists determined that some small RNAs could quash the expression of various genes in plant and, later, animal cells. But they didn't appreciate the molecules' true powers until 1998. That's when Andrew Fire of the Carnegie Institution of Washington in Baltimore, Maryland, Craig Mello of the University of Massachusetts Medical School in Worcester, and



Life cycle. With a helping hand from proteins RISC and Dicer, small RNAs are born. We now know that these molecules keep DNA in line and ensure a cell's good health.

their colleagues injected stretches of double-stranded RNA into worms. Double-stranded RNA forms when a familiar single strand kinks back in a hairpin bend, putting two complementary sequences alongside each other. To the researchers' surprise, double-stranded RNA dramatically inhibited genes that had helped generate the RNA in the first place. This inhibition, which was later seen in flies and other organisms, came to be known as RNA interference (RNAi). It helped prove that RNA molecules were behind some gene silencing.

Another crucial step came last year, when Gregory Hannon of Cold Spring Harbor Laboratory in New York and his colleagues identified an enzyme, appropriately dubbed Dicer, that generates the small RNA molecules by chopping double-stranded RNA into little pieces. These bits belong to one of two small RNA classes produced by different types of genes: microRNAs (miRNAs) and small interfering RNAs (siRNAs). SiRNAs are considered to be the main players in RNAi, although miRNAs, which inhibit translation of RNA into protein, were recently implicated in this machinery as well.

To bring about RNAi, small RNAs degrade the messenger RNA that transports a DNA sequence to the ribosome. Exactly how this degradation occurs isn't known, but scientists believe that Dicer delivers small RNAs to an enzyme complex called RISC, which uses the sequence in the small RNAs to identify and degrade messenger RNAs with a complementary sequence.

Such degradation ratchets down the expression of the gene into a protein. Although quashing expression might not sound particularly useful, biologists now believe that in plants, RNAi acts like a genome "immune system," protecting against harmful DNA or viruses that could disrupt the genome. Similar hints were unearthed in animals this year. In labs studying gene function, RNAi is now commonly used in place of gene "knockouts": Rather than delete a gene, a laborious process, double-stranded RNA is applied to ramp down its expression.

The year's most stunning revelations emerged in the fall, in four papers examining how RNA interference helps pilot a peculiar—and pervasive—genetic phenomenon known as epigenetics. Epigenetics refers to changes in gene expression that persist across at least one generation but are not caused by changes in the DNA code.

In recent years, researchers have found that one type of epigenetic regulation is caused by adjustments in the shape of complexes known as chromatin, the bundles of DNA and certain fundamental proteins that make up the chromosomes. By changing shape—becoming either more or less compact—chromatin can alter which genes are expressed. But what prompts this shape-

ILLUSTRATION: C. ZASTROW/SCIENCE

- These tiny RNAs (~21-26 nt) induce silencing through homologous sequence interactions
- They can control mRNA stability or translation, or target epigenetic modifications to specific regions of the genome.
- Small RNAs and evolutionarily conserved RNA-mediated silencing pathways have established a new paradigm for understanding eukaryotic gene regulation and revealed novel host defenses to viruses and transposons.

Function of small RNAs

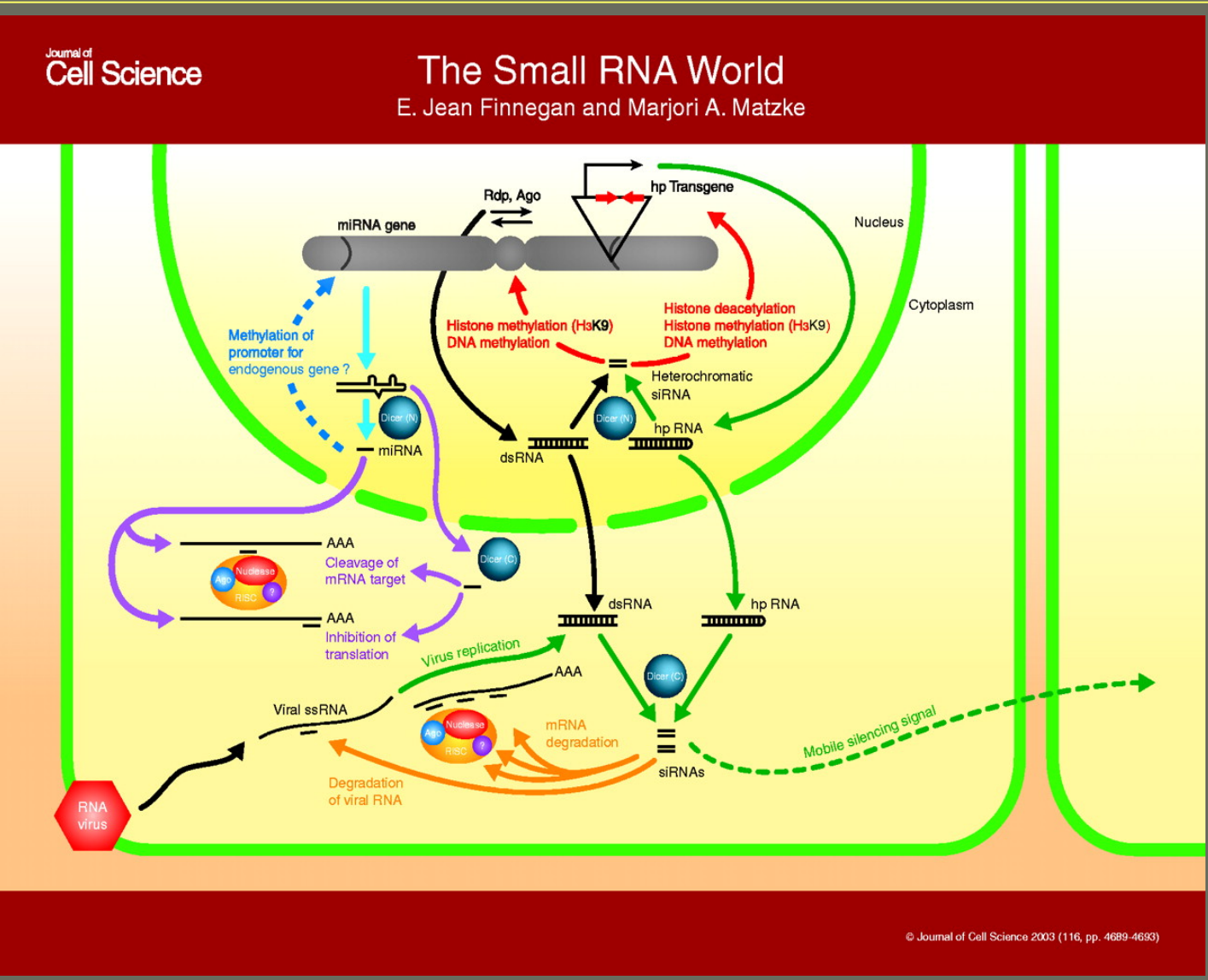
- MicroRNAs (**miRNAs**) and transacting siRNAs (**tasiRNAs**) are primarily involved in regulating **gene expression** and **plant development**,

- siRNAs play a major role in defending the genome against the **proliferation** of invading **viruses** and **endogenous transposable elements**.

- The function of the fourth type of sRNAs, natural-antisense siRNAs (nat-siRNAs), is not entirely clear but is likely related to **plant stress responses**

Zhang et al., PNAS 104, 4536 (2007)

Deciphering the small RNAs machinery



RNA-directed DNA methylation (RdDM), first discovered in plants requires a dsRNA that is processed to 21-24 nt small RNAs.

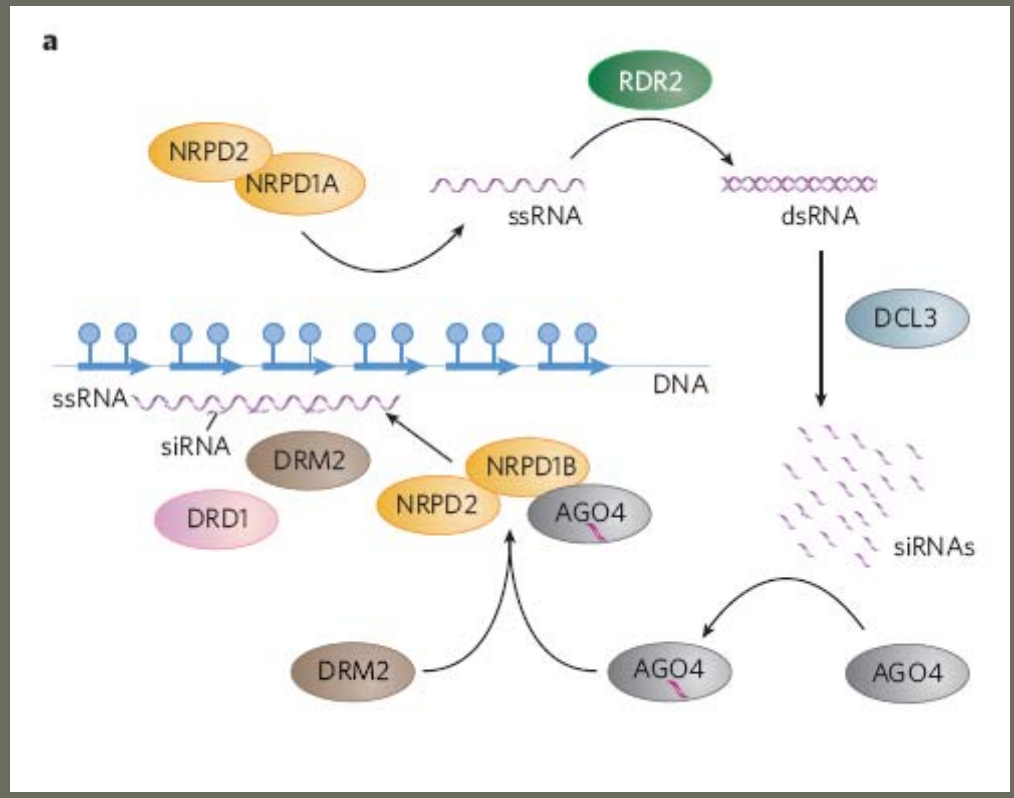
In *Arabidopsis*, links between locus specific small RNAs, DNA methyltransferases, and histone modifications, including deacetylation and histone H3 lysine 9 (H3K9) methylation have been identified (see red pathway).

Some small RNAs might target native promoters of endogenous genes (see dashed blue pathway).

RNAi and epigenetic alterations of the genome, such as DNA methylation and histone modifications

RNA-directed DNA methylation

Putative pathway for RNA directed DNA methylation in *A. thaliana*. Target loci (in this case tandemly repeated sequences; coloured arrows) recruit an RNA polymerase IV complex consisting of NRPD1A and NRPD2 through an unknown mechanism, and this results in the generation of a single-stranded RNA (ssRNA) species. This ssRNA is converted to double-stranded RNA (dsRNA) by the RNA-dependent RNA polymerase RDR2.



The **dsRNA** is then processed into 24-nucleotide **siRNAs** by **DCL3**. The siRNAs are subsequently loaded into the protein **AGO4**, which associates with another form of the RNA polymerase IV complex, **NRPD1B–NRPD2**. AGO4 that is 'programmed' with siRNAs can then locate homologous genomic sequences and guide the protein **DRM2**, which has *de novo* cytosine methyltransferase activity. Targeting of DRM2 to DNA sequences also involves the chromatin remodelling protein **DRD1**.

Henderson & Jacobson, Nature 447, 418 (2007)

Conclusions

✓ The 'epigenetic code' considerably extends the information potential of the genetic code.

✓ Thus, one genome can generate many 'epigenomes' as the fertilised egg progresses through development and translates its information into a multitude of cell fates.

✓ The transcriptomes of an organism are continually changing in response to developmental and environmental cues.

✓ The epigenome is not static and can be molded by developmental signals, environmental perturbations, and disease states.

✓ Therefore, many epigenomes will need to be sequenced for a single organism, making epigenome sequencing perhaps even more challenging than genome sequencing.

The search for epigenetic factors of flower development in oil palm

Alain Rival (Cirad)

Sylvie Doulbeau (IRD), Frédérique Aberlenc (IRD), James Tregear (IRD), Estelle Jaligot (Cirad), Pascal Ilbert (Cirad), Thierry Beulé (Cirad)

IRD Institut de recherche pour le développement

CIRAD

Rationale

- The commercial development of large scale propagation of oil palm is hampered by the generation of somaclonal variants affecting the floral architecture: the « mantled » genotype.

- Field observation of clonal plantings on the long term and results from DNA markers analysis (Rival et al, 1998) are consistent with an epigenetic origin for somaclonal variation.

- The “mantled” phenotype is correlated with a global hypomethylation of genomic DNA, which was revealed through a whole genome approach (Jaligot et al, 2000) and the use of Methylation-Sensitive DNA markers, such as MS-RFLPS and MSAPs (Jaligot et al, 2002; 2004).

Journal of Experimental Botany, Vol. 59, No. 12, pp. 3271–3281, 2008
doi:10.1093/jxb/ern178 Advance Access publication 17 July, 2008

Journal of Experimental Botany
www.jxb.oxfordjournals.org

RESEARCH PAPER

Isolation and expression analysis of genes encoding MET, CMT, and DRM methyltransferases in oil palm (*Elaeis guineensis* Jacq.) in relation to the 'mantled' somaclonal variation

Alain Rival[1,*,†], Estelle Jaligot[1,*], Thierry Beulé[1] and E. Jean Finnegan[2]

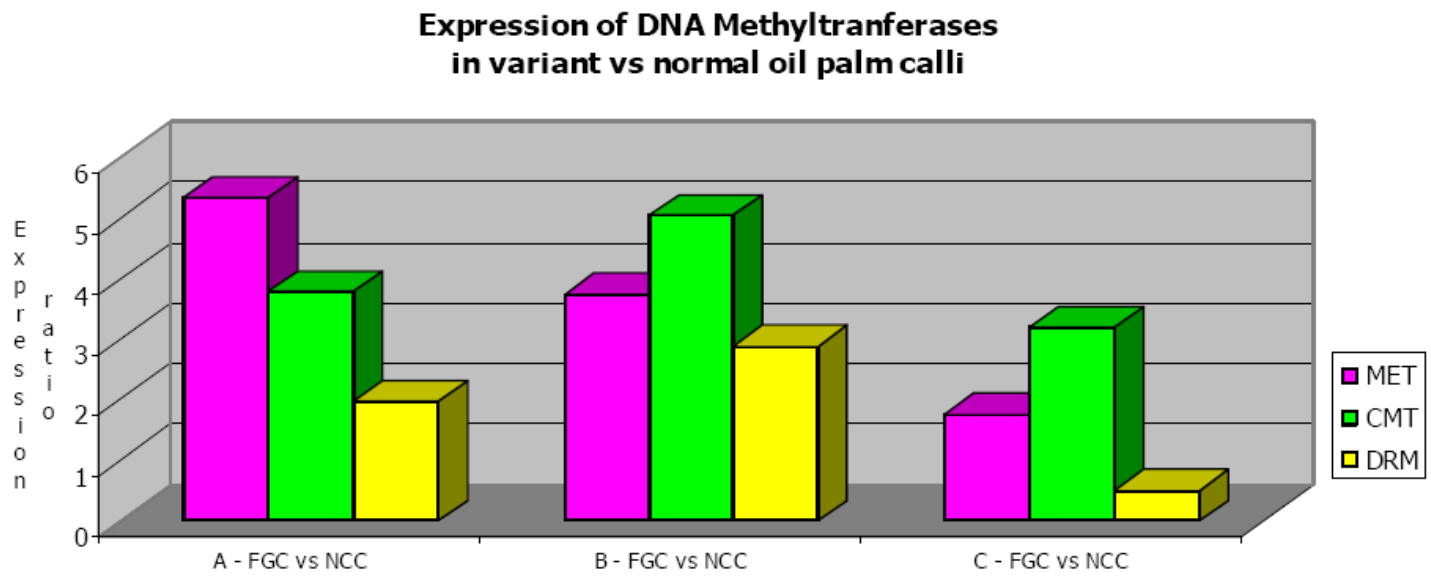
[1] *CIRAD, IRD, UMR DIAPC, BP 64501, F-34394 Montpellier, Cedex 5, France*
[2] *CSIRO Plant Industry, GPO Box 1600, ACT 2602, Canberra, Australia*

Received 23 April 2008; Revised 9 June 2008; Accepted 10 June 2008

IRD Institut de recherche pour le développement

CIRAD

Real Time qPCR analysis on embryogenic calli



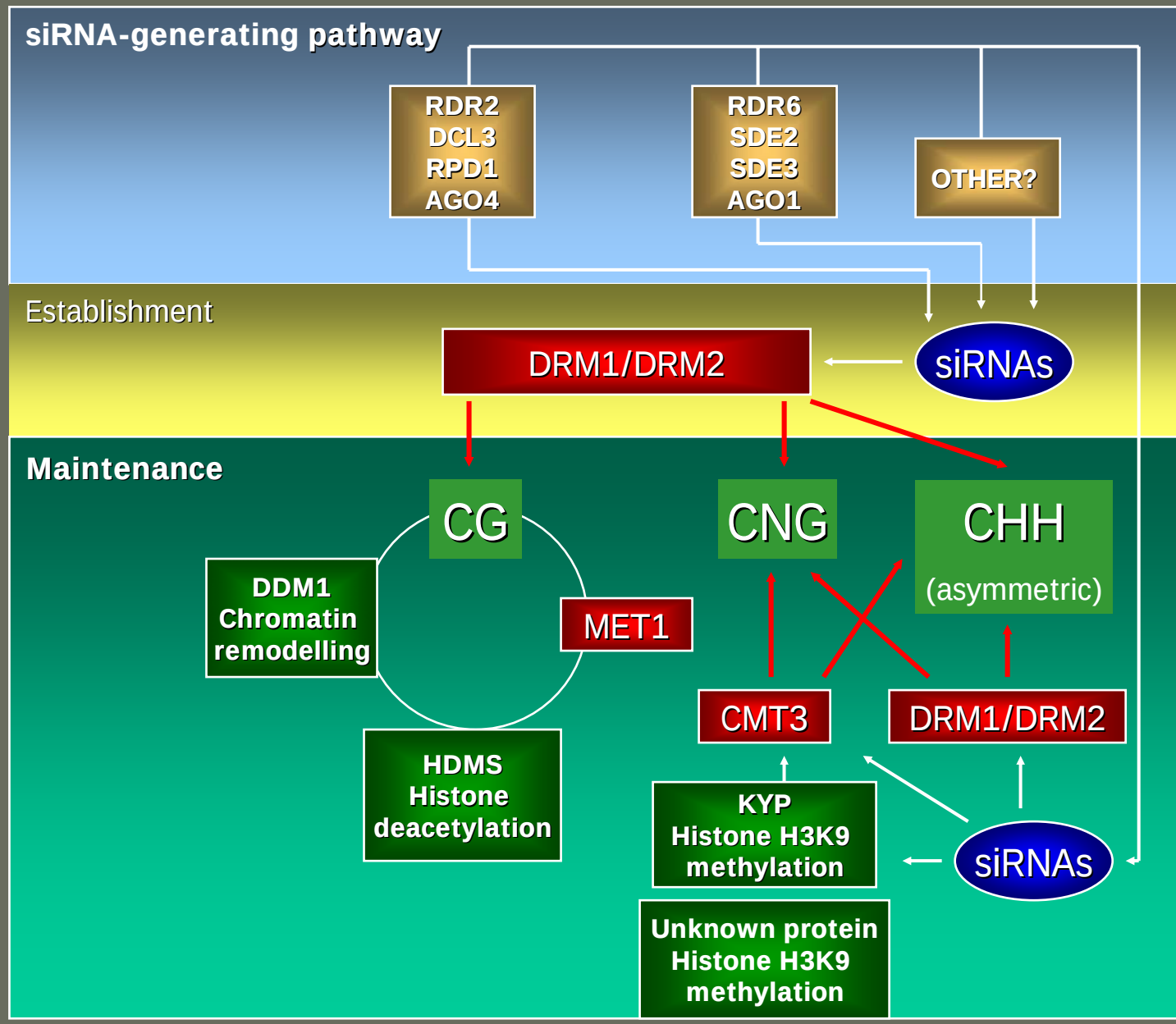
Quantitative Real Time PCR was performed on total RNAs extracted from embryogenic calli from three different genotypes (A, B and C). For each genotype, Fast-Growing Calli (FGC, generating 100% of "mantled" palms) were compared to Nodular Compact Calli (which yield on average 5% of variant palms). Gene expression was represented by the ratio of expression in FGC by expression in NCC.
For each methyltransferases gene, expression levels were standardized using control Elongation Factor EgEF1-α1 sequence from oil palm (accession no. AY550990) as a reference.

Journal of Experimental Botany
www.jxb.oxfordjournals.org

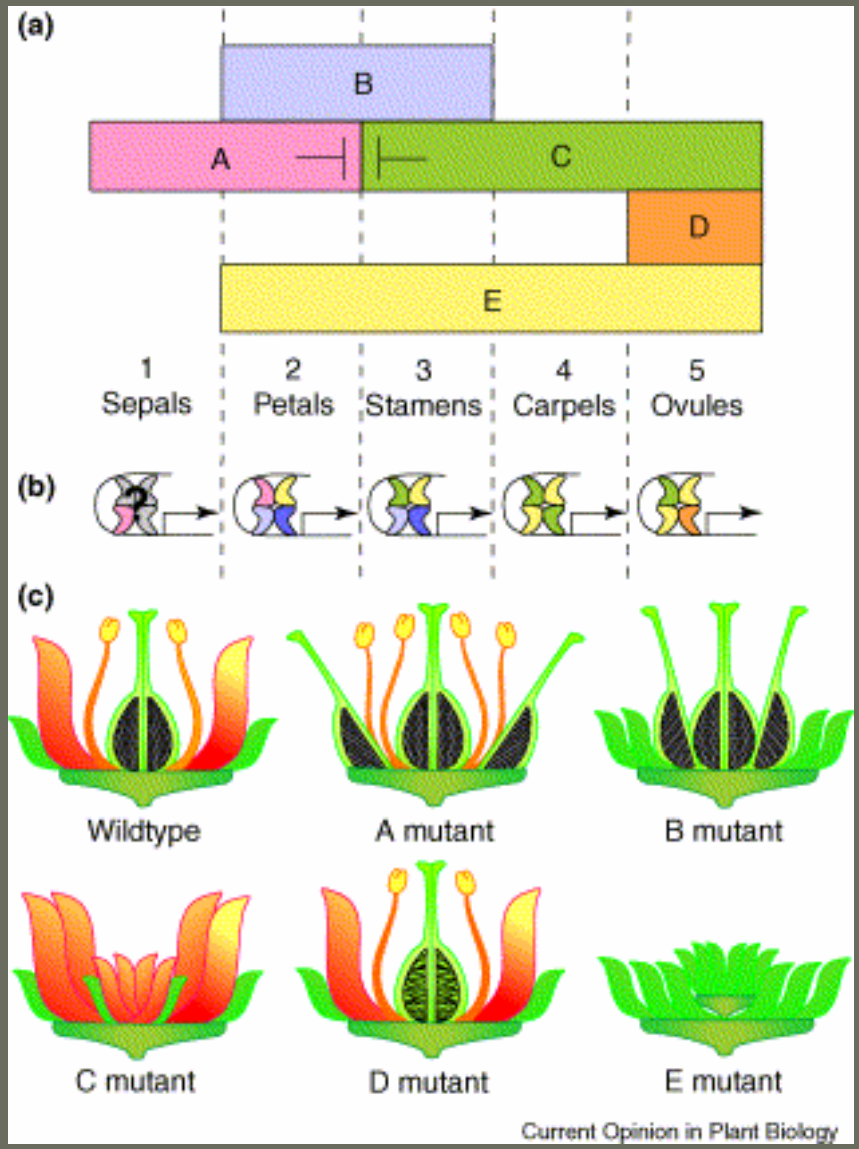
Establishment and Maintenance of DNA methylation

GARDENING THE GENOME
DNA METHYLATION IN
ARABIDOPSIS THALIANA

Simon W.-L. Chan, Ian R. Henderson* and Steven E. Jacobsen*



Methylation around MADS Box candidate genes



Several oil palm MADS box genes have shown differential expression patterns according to the presence of *mantled* abnormality

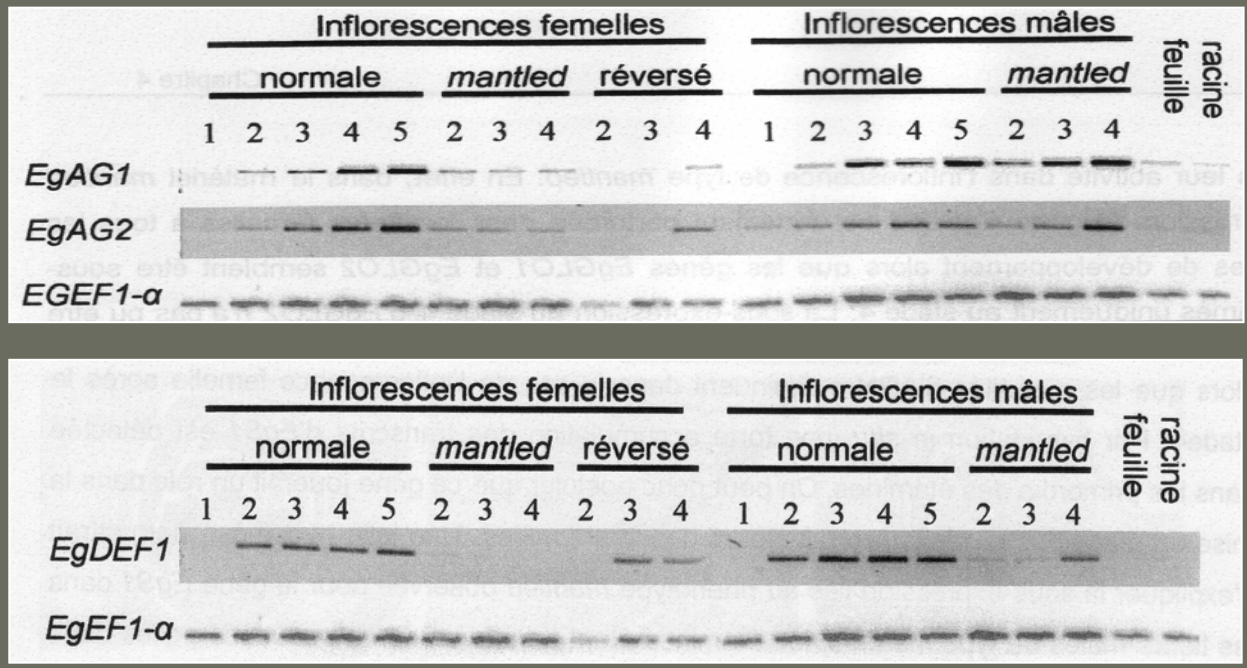
Alterations in expression affect not only B-type, but also C ,D and E-type genes

Whorls 2, 3 and 5 are affected by homeotic changes

IRD
Institut de recherche pour le développement

CIRAD

Methylation around MADS Box candidate genes

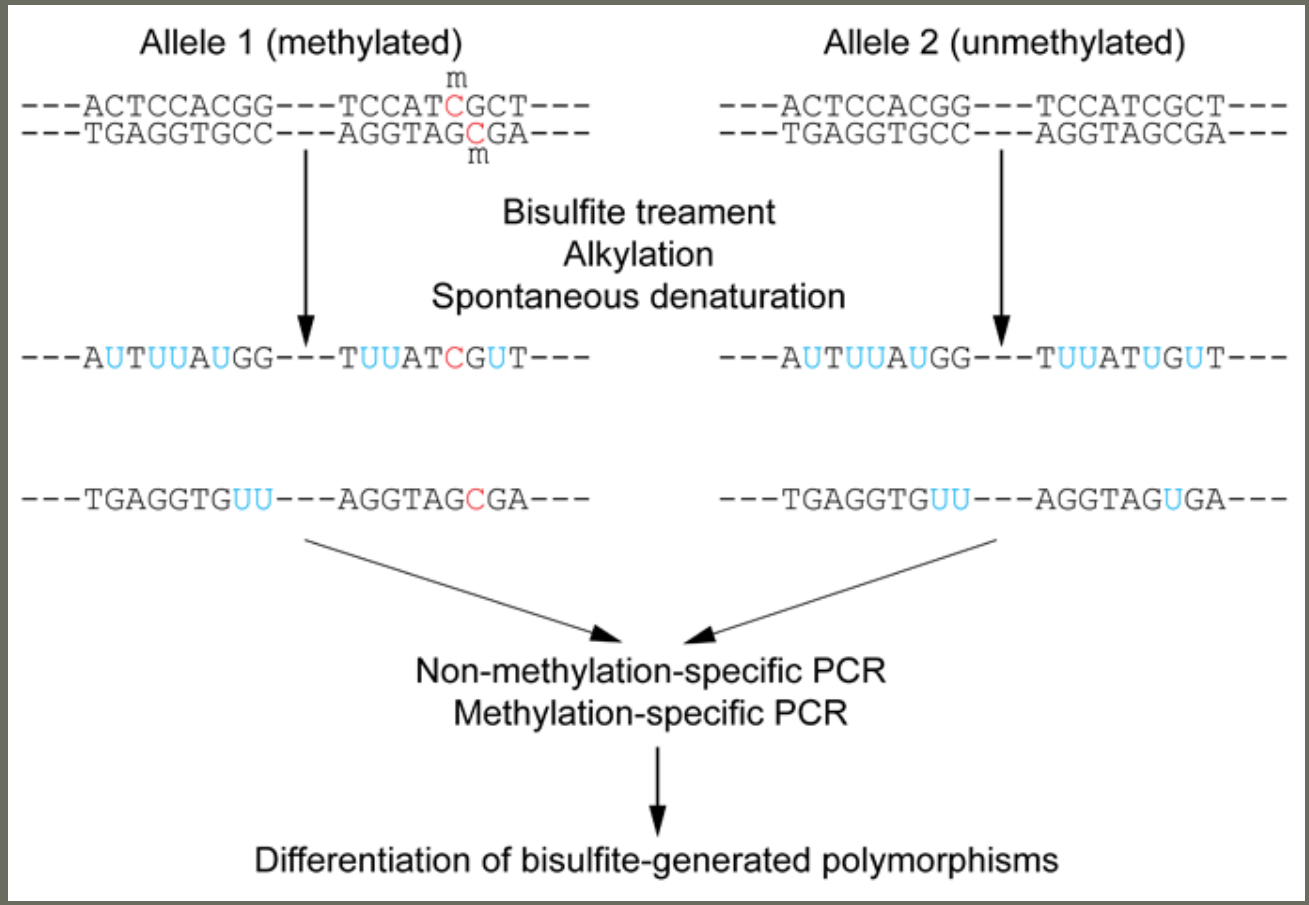


Reduced expression of genes Eg DEF1 and EgGLO2 (B type), EgAG2 (C and D type) and EgS1 (type E) in abnormal oil palm flowers

IRD Institut de recherche pour le développement

CIRAD

Methylation-specific PCR: towards MS markers...



Relevant literature

JALIGOT E., RIVAL A., BEULÉ T., DUSSERT S. & VERDEIL J.-L. (2000) Somaclonal variation in Oil Palm (*Elaeis guineensis* Jacq.): The DNA methylation hypothesis. *Plant Cell Reports 19 (7): 684-690.*

TREGEAR J., MORCILLO F., RICHAUD F., BERGER A., SINGH R., CHEAH S.C., HARTMANN C., RIVAL A. & DUVAL Y. (2001) Characterisation of a defensin gene expressed in oil palm inflorescence: induction during tissue culture and possible association with epigenetic somaclonal variation events*. Journal of Experimental Botany*, *53 : 1387-1396.*

JALIGOT E., BEULÉ T. & RIVAL A. (2002) Methylation-sensitive RFLPs reveal a differential banding pattern associated with somaclonal variation in oil palm (*Elaeis guineensis* Jacq.).*Theoretical and Applied Genetics. 104:1263-1269.*

JALIGOT E., BEULÉ T., BAURENS F.C., BILLOTE N. & RIVAL A. (2004). MSAP screening for differentially methylated markers associated with the « mantled » somaclonal variation in oil palm (*Elaeis guineensis* Jacq.). *Genome 47:224-228.*

MORCILLO F., GAGNEUR C., ADAM H., JOUANNIC S., RICHAUD F., RAJINDER S., CHEAH S.C., RIVAL A., DUVAL Y. & TREGEAR J.W. (2005) Somaclonal variation in micropropagated oil palms: Characterization of two novel genes displaying enhanced expression in epigenetically abnormal cell lines and investigation of the influence of auxin on their activity. *Tree Physiology* : 26, 585–594.

ADAM H, JOUANNIC S, ESCOUTE J., DUVAL Y , VERDEIL J-L & J.W. TREGEAR (2005) Reproductive developmental complexity in the african oil palm (Elaeis guineensis, Arecaceae). *American Journal of Botany* 92(11): 1836–1852.

RIVAL A. , E. JALIGOT, T. BEULÉ & J. FINNEGAN (2008) Isolation and differential expression of MET, CMT and DRM methyltransferase genes from oil palm (*Elaeis guineensis* Jaq.) in relation with the “mantled” somaclonal variation. *Journal of Experimental Botany* (doi:10.1093/jxb/ern178).

Thank you for your kind attention …

alain.rival@cirad.fr