CHAIRE ÉPIGÉNÉTIQUE ET MÉMOIRE CELLULAIRE

Année 2016-2017 :
"Épigénétique et ADN égoïste"

2 Février, 2017

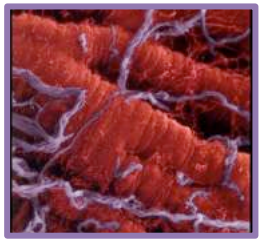
Cours I

La découverte des éléments transposables du génome : parasites ou protagonistes ?

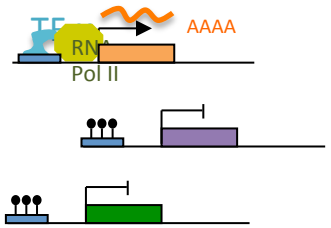
Discovery of Transposable Elements: parasites or protagonists of the genome?

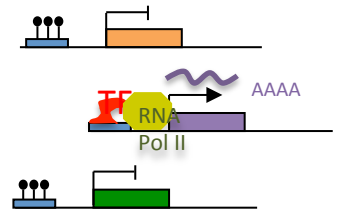
The Stable Genome

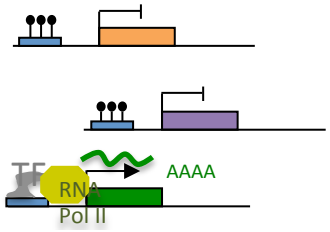
One genome: multiple gene expression patterns, multiple "epigenomes"



Lennart Nilsson ©





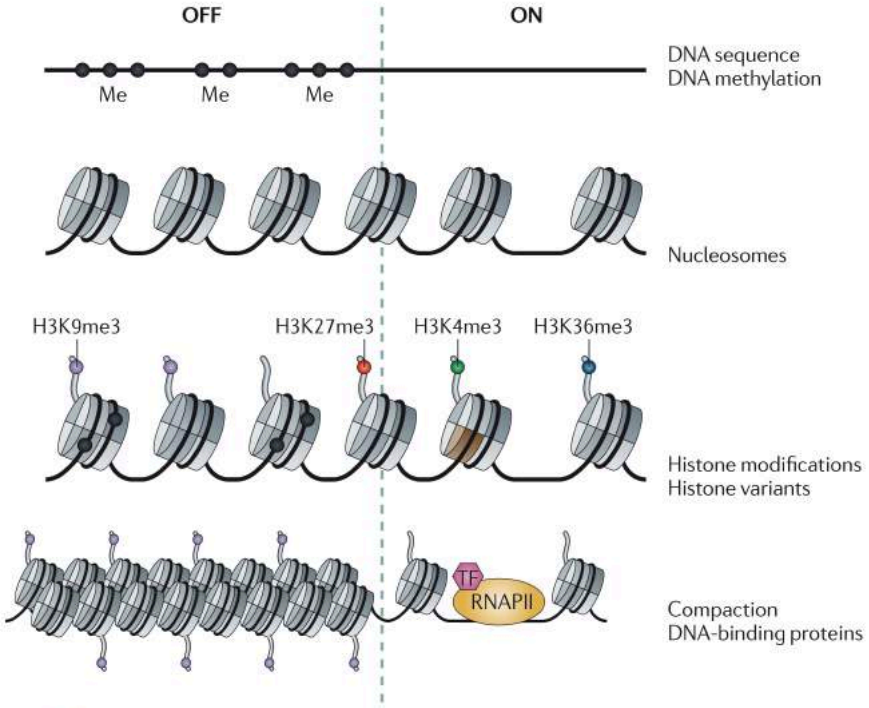


1. All cells contain same genes: cell identities depend on **which** genes are expressed/repressed.
2. Expression patterns established by **transcription factors (signalling +** positional info)
3. Genes cannot usually be activated *just* by a transcription factor
4. Changes in gene expression become stable and heritable during development.

EPIGENETICS

The Stable Genome and its Epigenomes

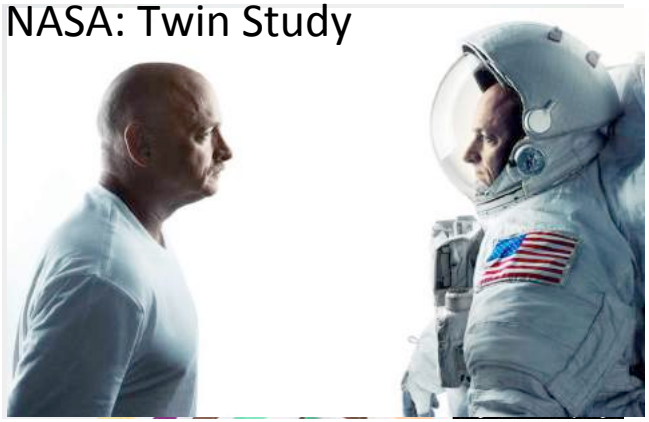
One genome: multiple gene expression patterns, multiple "epigenomes"



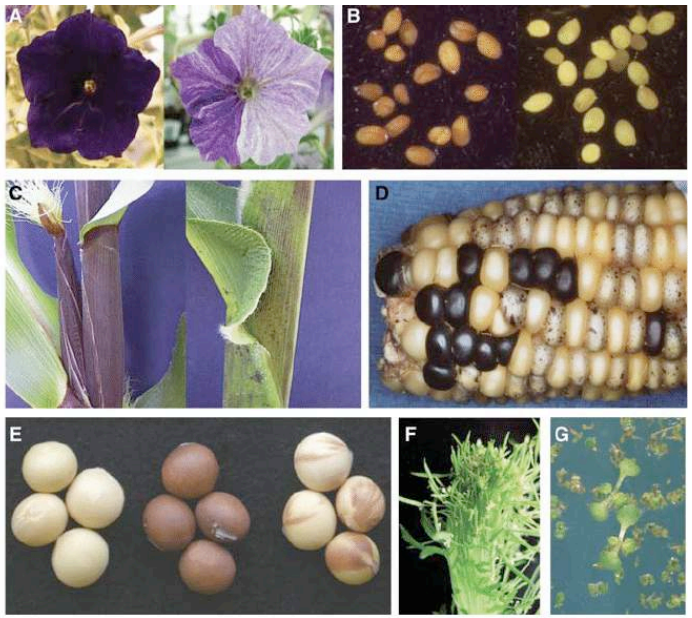
1. All cells contain same genes: cell identities depend on **which** genes are expressed/repressed.
2. Expression patterns established by **transcription factors (signalling +** positional info)
3. Genes cannot usually be activated *just* by a transcription factor
4. Changes in gene expression become stable and heritable during development.

EPIGENETICS

But how Static *is* the Genome (and Epigenomes)?



Genetically identical twins



Epigenetic changes can occur over life

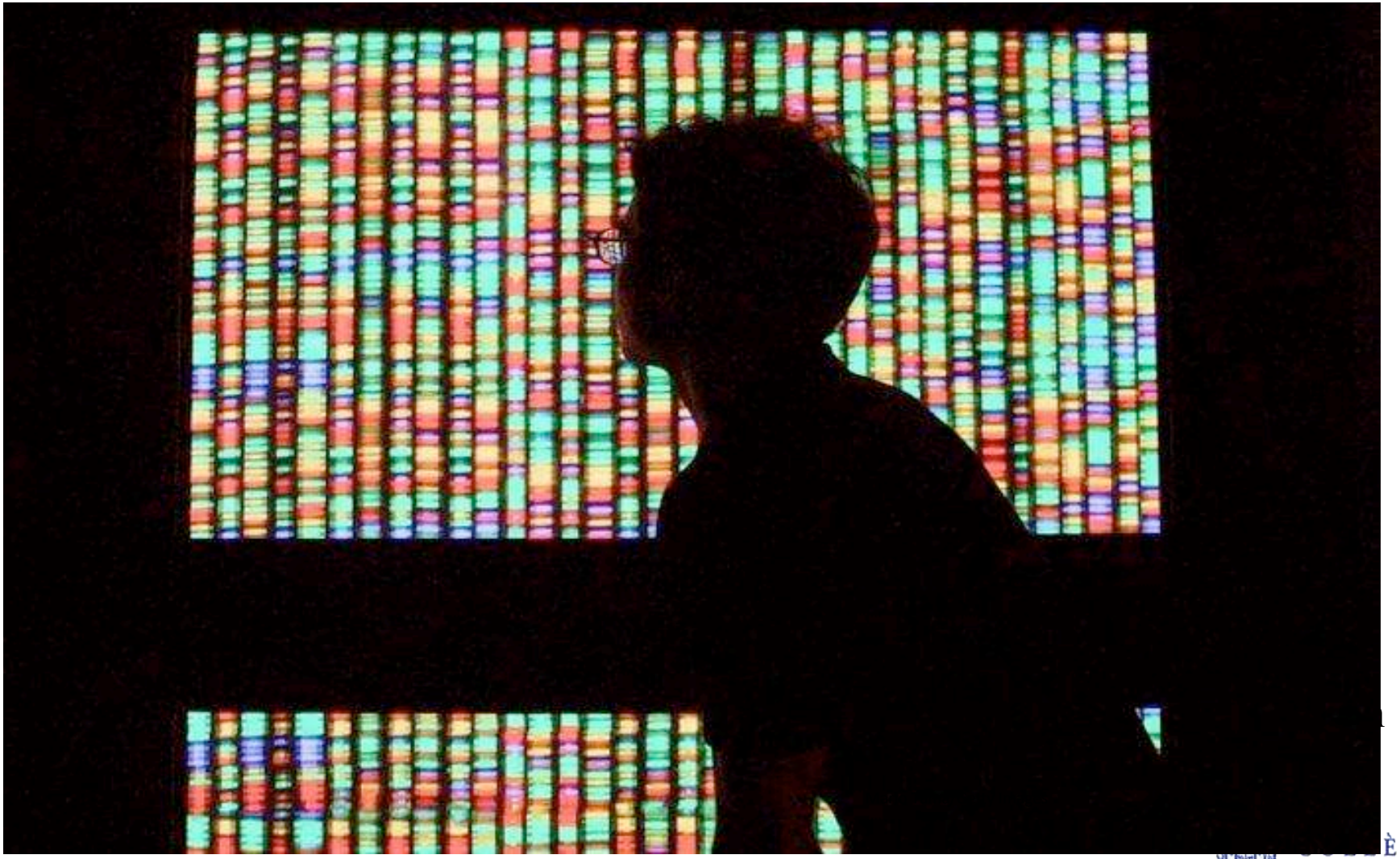
Epigenetic changes
can account for
differences
between genetically
identical individuals

In most cases, an important driving force of such phenotypic variation, either within cells of an individual, or between genetically identical individuals is thought to be **Transposable Elements (TEs)**:
DNA segments that can move around the genome
and that can attract epigenetic factors

Parasites or Protagonists?

The Dynamic Genome

- Genomes are dynamic and changeable, in a life, over evolutionary time



The Dynamic Genome

- Genomes are dynamic and changeable, in a life, over evolutionary time
- Transposable elements (TEs) contribute to genome fluidity
- TEs and their relics (fossils) are major players in genome evolution
- TEs have helped to shape the form and function of many genes
- TEs are primarily parasitic DNA - and parasites must be controlled, or they will destroy their host (both genome integrity *and* gene regulation)
- Defense strategies include epigenetic mechanisms, which may even have evolved for this purpose, and then been co-opted for other processes
- Epigenetic mechanisms enable exploitation of TEs and their relics to influence endogenous gene expression, chromosome functions, phenotypic diversity: genetic/epigenetic; soma/germ line
- Environment can influence transposon activity, which in turn may help an organism adapt to environmental changes.

COURS 2017 : Épigénétique et ADN égoïste

lundi 6 février
La découverte des éléments transposables du génome : parasites ou protagonistes ?
Discovery of Transposable Elements: parasites or protagonists of the genome?

lundi 13 février
Le rôle de l'épigénétique dans la régulation des éléments transposables.
The role of epigenetics in the regulation of transposable elements

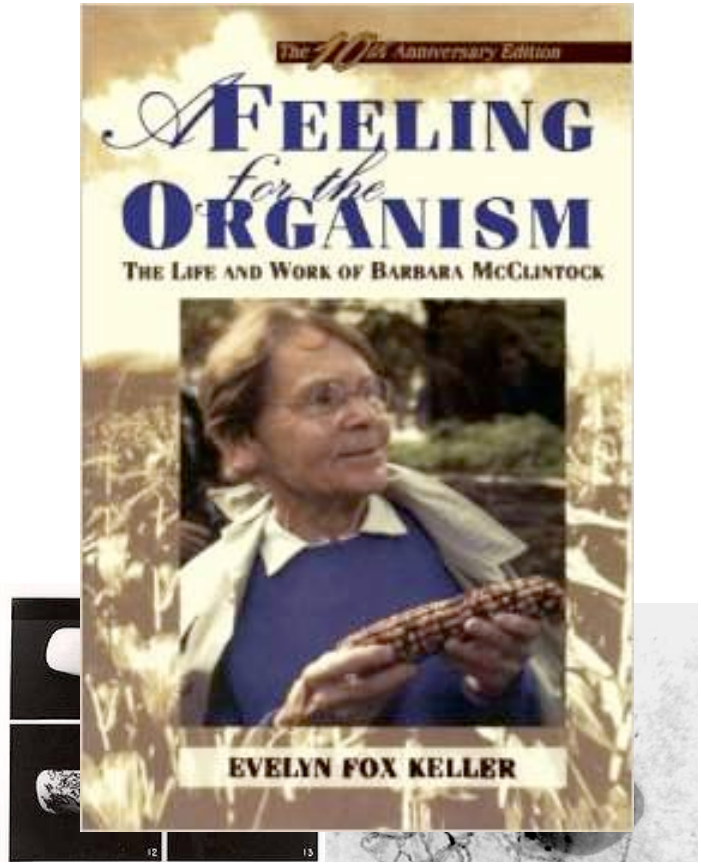
lundi 20 février
L'impact des éléments transposables et de leurs reliques sur le développement.
The impact of transposable elements and their relics on development

lundi 27 février
L'implication des éléments transposables dans les maladies : mutations et épimutations
The implication of transposable elements in disease: mutations and epimutations

mercredi 08 mars
Contribution des éléments transposables et de leur contrôle épigénétique à l'évolution
Contributions of transposable elements and their epigenetic control in evolution

Barbara McClintock

Barbara McClintock
(1902-1992)
Botanist, geneticist,
cytogeneticist



Nobel Prize in
Physiology or Medicine,
1983
"for her discovery of
mobile genetic
elements".

Her work went against the prevailing genetic theory of the time that genes were *fixed* in their positions on chromosomes

McClintock found that genes could not only move, they could also be turned on or off due to certain environmental conditions or during different stages of cell development.

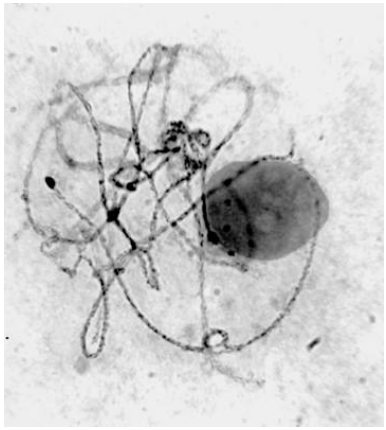
McClintock also showed that gene mutations could be reversed…

Barbara McClintock

The chromosomal basis of heredity was well established by the time McClintock started her PhD in the 1920's, in the Botany Department at Cornell University.

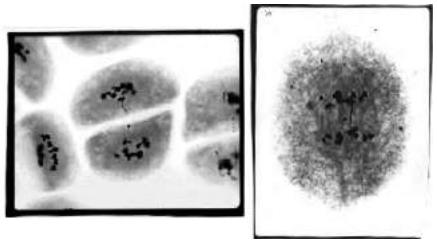
During her PhD, she developed cytological techniques that allowed her to study the **chromosome complement**, and to relate the **linkage groups of the genes** with the **physical structure of chromosomes**.

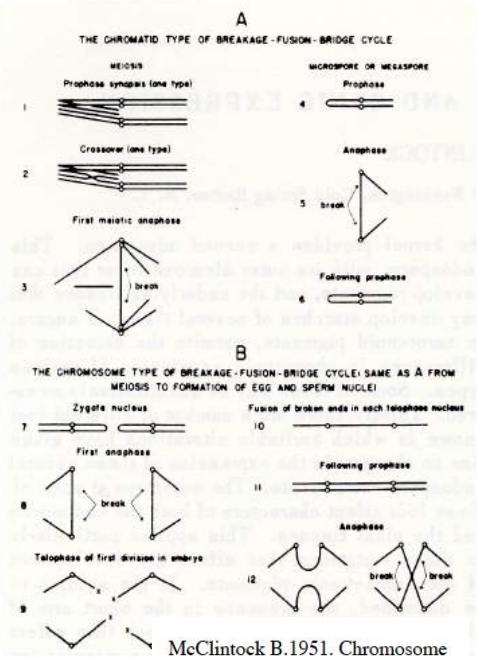
McClintock made a remarkable series of cytogenetic discoveries with the Cornell maize genetics group between 1929 and 1935:



- Identification of maize chromosomal linkage groups – 1st genetic maize map
- First ever cytological proof of genetic crossing-over and evidence of chromatid crossing-over
- Cytological determination of physical location of genes on chromosomes
- Identification of the genetic consequences of non-homologous pairing
- Establishment of the causal relationship between the instability of ring-shaped chromosomes and phenotypic variegation
- Discovery that the centromere is divisible
- Identification of chromosomal site essential for nucleolar formation (rDNA)

Barbara McClintock





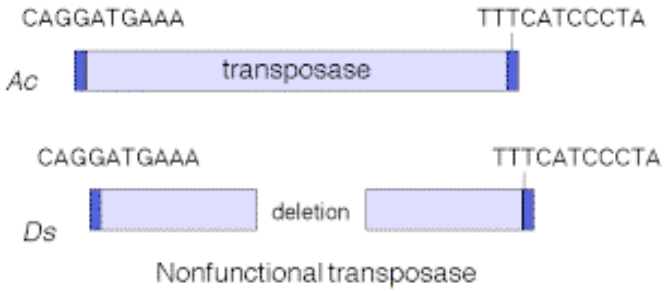
McClintock B.1951. Chromosome organization and genic expression. Cold Spring Harbor Symposia on Quantitative Biology 16: 13-47.)

Following her PhD, McClintock became interested in **chromosome breakage**: X-ray induced breakages would set off a cycle of chromosome instability

- she noted that chromosomes without ends were unstable and proposed the existence of telomeres
- She noted X-ray treatment can produce chromosomes with two centromeres which form a "bridge" as they attempt to separate during cell division and that the bridging stress eventually causes the chromosomes to break and the broken ends refuse to one another in the new daughter cells.
- This "Breakage-Fusion-Bridge" cycle "*cassure*-fusion-*pont*"repeats through plant development, leading to variegated (multicolored) leaves & kernels.
- She then used B-F-B as a way of generating genetic instability - and in 1944 generated plants that had received a broken chromosome from each parent
- These displayed unstable mutations at an very high frequency
- But she then noted a *unique* mutation that defined a *regular* site of chromosome breakage …
- Intensive investigation revealed in 1948, that this chromosome-breaking locus was actually able to move from one chromosomal location to another!

This was the beginning of her discovery of transposition and transposable elements
(1948)

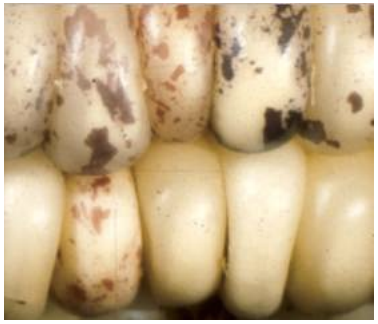
Barbara McClintock: The discovery of transposable elements

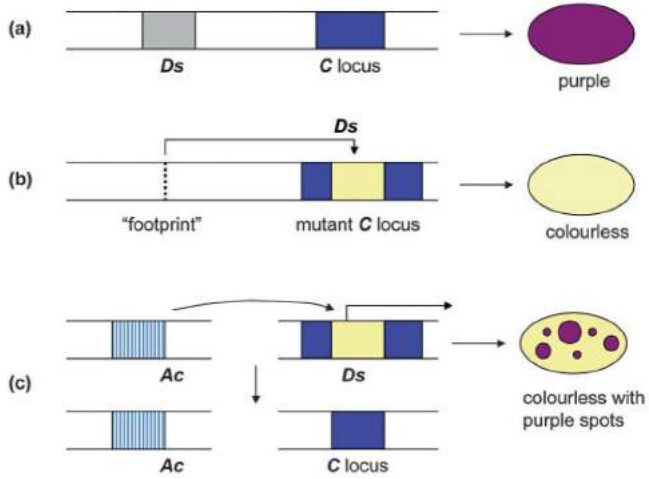


Maize kernel colour controlled by multiple genes

• Classical genetics: a mutation in any of these genes leads to a *colorless* kernel.

• McClintock studied unstable mutations - spots of purple pigment on white kernels





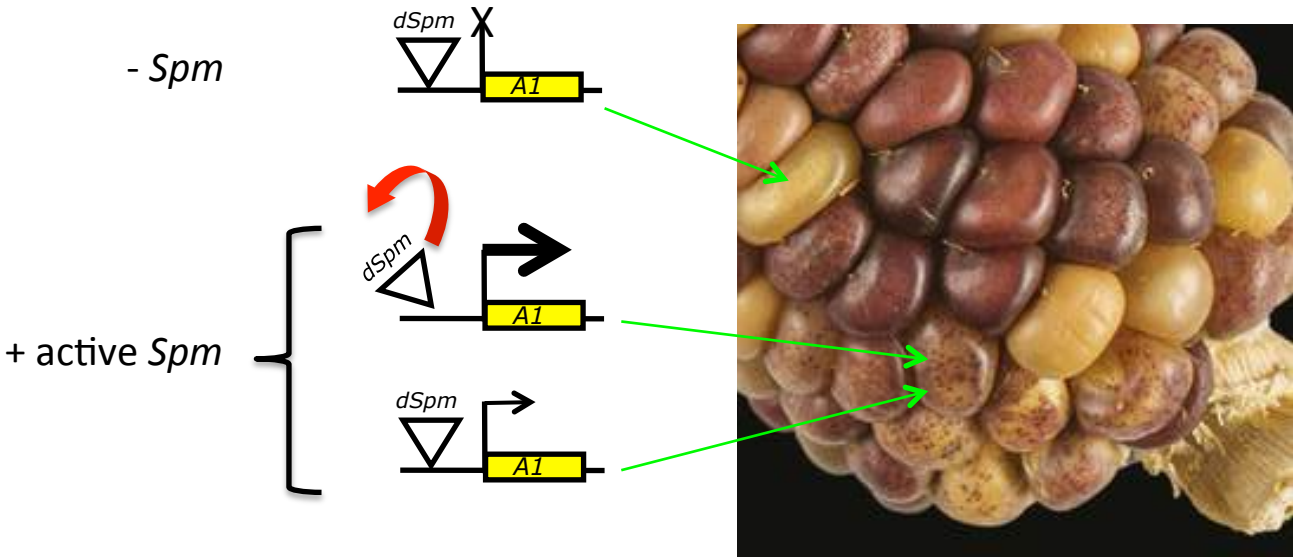
McClintock concluded that the *c* allele resulted from insertion of a "mobile controlling element" into the *C* allele.

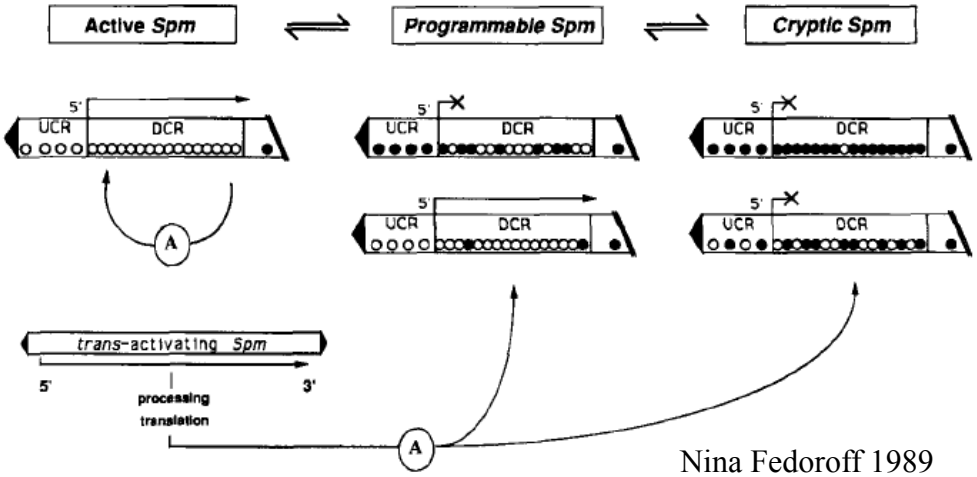
(1) The element is *Ds* (dissociation) was a non-autonomous transposon.

(2) Its transposition is controlled by *Ac* (activator), an autonomous transposon

A reversion of *c* to *C* in a cell leads to purple pigment, and hence a spot. The earlier in development the reversion occurs, the larger the spot.

…and of their epigenetic control

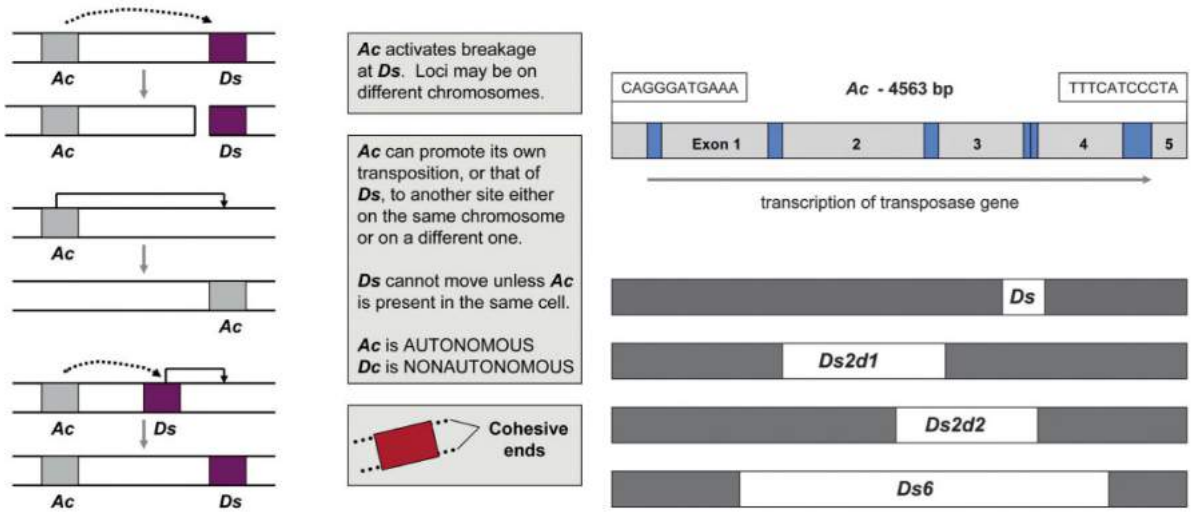




Nina Fedoroff 1989

Summary of McClintock's Discoveries on Mobile Elements

- **Transposable elements** :
  - Induce chromosome breaks (in some cases)
  - Mutate genes
  - Move in and out of genes (revertants)
  - Can be sensitive to genomic "stress" (eg BFB cycles)
  - Can be autonomous (Ac) or non-autonomous (Ds) – and non-autonomous elements require autonomous ones to cause breaks or to move
  - Display **changes of "state"** (influencing gene expression) in a dynamic and highly controlled way during development
  - Can **change "phase"** (Spm) being more or less active at different times during development / growth of the organisms



From R.N Jones, Cytogenet Genome Res 109:90–103 (2005)

Summary of McClintock's Discoveries on Mobile Elements

- **Transposable elements** :
  - Induce chromosome breaks (in some cases)
  - Mutate genes
  - Move in and out of genes (revertants)
  - Can be sensitive to genomic "stress" (eg BFB cycles)
  - Can be autonomous (Ac) or non-autonomous (Ds) – and non-autonomous elements require autonomous ones to cause breaks or to move
  - Display **changes of "state"** (influencing gene expression) in a dynamic and highly controlled way during development
  - Can **change "phase"** (Spm) being more or less active at different times during development / growth of the organisms

- **Effects of these elements on gene expression during development**

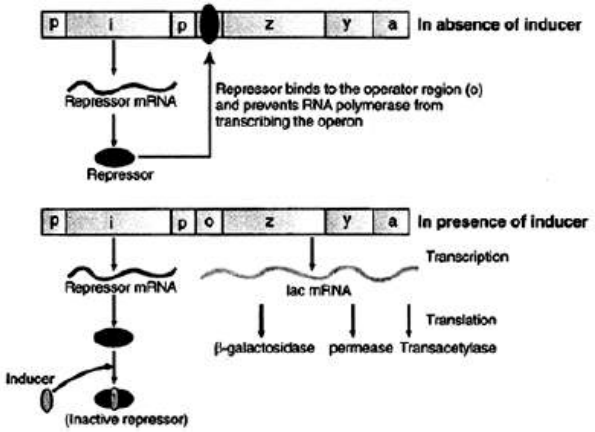
Her proposal that differential gene expression might be controlled, during development, by such elements and their activities ("**Controlling Elements**"), seemed incomprehensible or unacceptable at the time – and were largely ignored:

> *"It is now known that controlling elements may modify gene action in a number of different ways. They may influence the time of gene action in the development of a tissue and also determine the cells in which it will occur".*

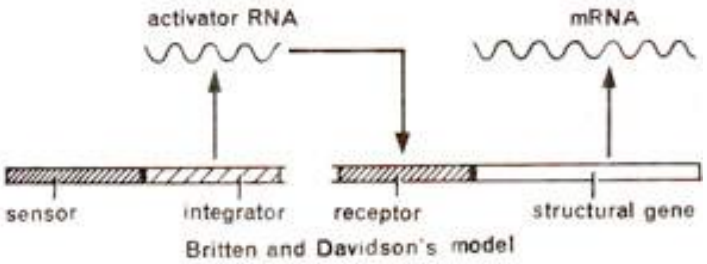
The Lac Operon extended to Developmental Gene Regulation

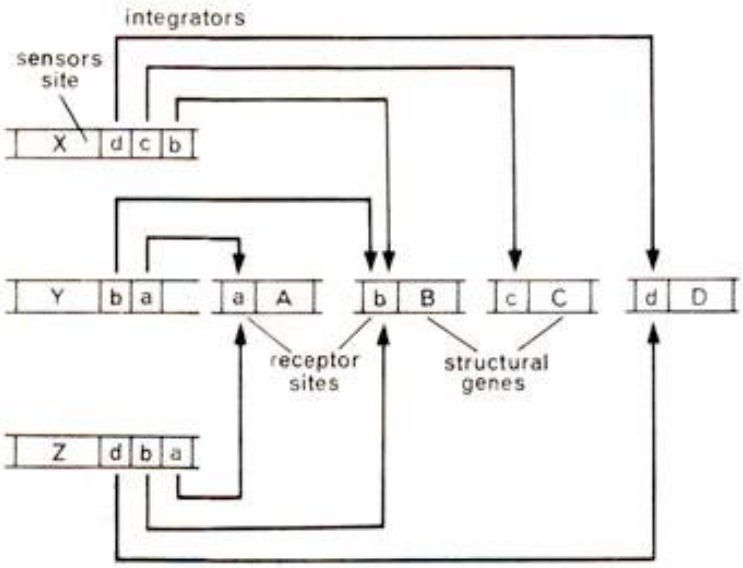
Jacob and Monod, 1961: Lac operon model
Fundamental concept that gene control relies on specific repressors and activators and the DNA sequence elements they recognize.

T.H. Morgan in 1934 had already evoked the idea of "gene batteries", or sets of genes that are expressed at different stages during development.



In 1969, Britten and Davidson proposed a theoretical model for how gene regulatory networks might work during differentiation, with integrated activation of large numbers of noncontiguous genes upon external signals.





Britten R.J. and Davidson E. (1969) "Gene Regulation for Higher Cells: A Theory", *Science* 165: 349-357.

Parallels between McClintock's Controlling Elements and the Lac Operon

THE
AMERICAN NATURALIST

Vol. XCV September-October, 1961 No. 884

SOME PARALLELS BETWEEN GENE CONTROL SYSTEMS IN MAIZE AND IN BACTERIA

BARBARA McCLINTOCK
Department of Genetics, Carnegie Institution of Washington, Cold Spring Harbor, New York

It has been realized for some time that, although the gene is necessary for expression of a certain phenotype, it may not in itself be sufficient for such expression and mechanisms may exist that control its action. Genetic systems that serve this purpose in maize were recognized some years ago, and studies conducted with a number of them have been reported (for references, see Brink, 1958, 1960; McClintock, 1956a and b; Peterson, 1960). Without adequate confirmation of similar systems in other organisms, it could be considered that the systems in maize may not reflect a type of control of gene action that is common to organisms in general. Recently, however, genetic systems that control gene action have been discovered in bacteria (Jacob and Monod, 1959, 1961; Jacob *et al.*, 1960) and it is now apparent that a relationship may exist between the bacterial and the maize control systems. The bacterial control systems, described by Jacob *et al.*, are composed of two genetic elements, each distinct from the "structural" gene. One of them, designated the "operator," is located adjacent to the structural gene (or sequence of structural genes) and controls its activation. The structural gene, when activated, is responsible for the production of a particular sequence of amino acids and thus for the specificity of a protein. The second element of this system, termed the "regulator," may be located close to the structural gene, or it may be located elsewhere in the bacterial chromosome. The regulator is responsible for the production of a repressor substance—not a protein—that appears in the cytoplasm. The operator element responds in some yet unknown manner to changes in degree of effective action of the repressor substance by "turning on" or "turning off" the action of the structural gene in accordance with such changes. Each operator-regulator system is specific, in that an operator will respond only to the specific product of the regulator of its system.

In maize likewise, some of the control systems are composed, basically, of two elements. One is closely associated with the structural gene and directly controls its action; it may be likened to the operator element in bacteria. The other element may be located near the first or may be independently located in the chromosome complement. It establishes the conditions

In maize likewise, some of the control systems are composed, basically, of two elements. One is closely associated with the structural gene and directly controls its action; it may be likened to the operator element in bacteria. The other element may be located near the first or may be independently located in the chromosome complement. It establishes the conditions to which the gene-associated element responds, a particular change in these conditions being reflected in a particular change in action of the gene, and thus is comparable to the regulator element in bacteria. In maize, as in bacteria, each "operator-regulator" system is quite specific: an "operator" element will respond only to the particular "regulator" element of its own system.

Courtesy of Cold Spring Harbor Laboratory Archives.
Noncommercial, educational use only.

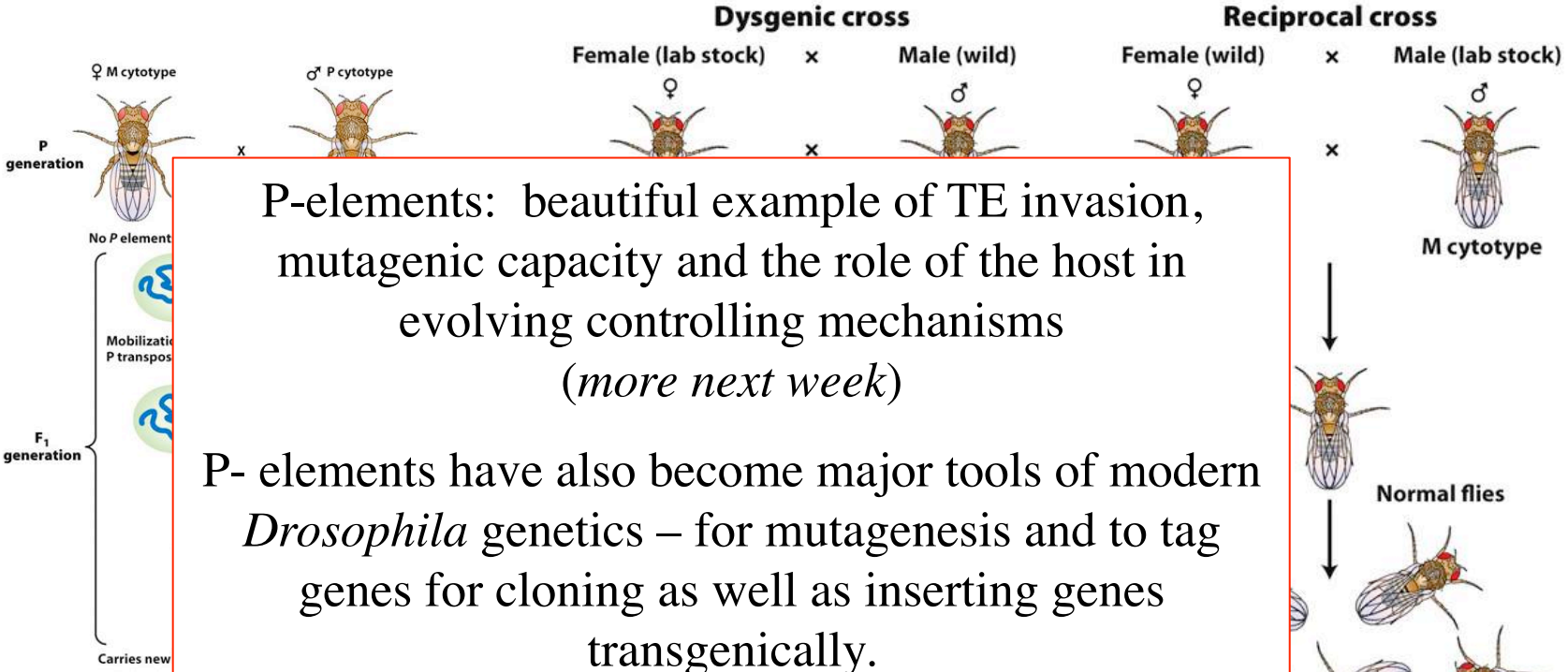
Barbara McClintock at CSH with Jacques Monod, 1946

COLLÈGE DE FRANCE 1530

Mobile Elements: not just an oddity of Maize…

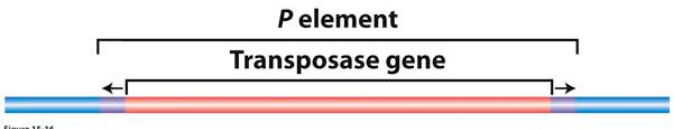
• **Bacterial Transposons** were originally detected in bacteria in 1950's, and later found to be mobile genetic elements that confer drug resistance. They can jump from the chromosome to plasmid DNA and back, transferring antibiotic resistance and resulting in bacterial strains that are multi-antibiotic resistant.

• **Hybrid dysgenesis** results from the mobilization of DNA sequences called P elements in Drosophila embryos. When a sperm from a P-carrying strain fertilizes an egg from a non-P-carrying strain, the P elements transpose throughout the genome, usually disrupting vital genes



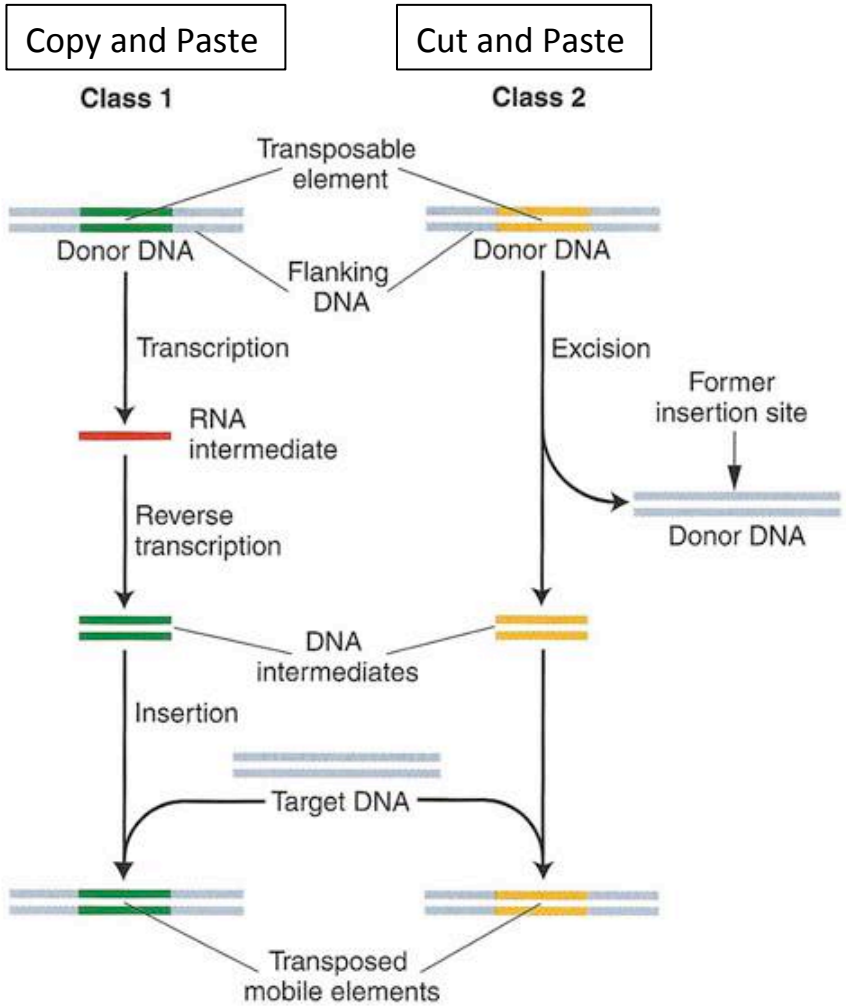
P-elements: beautiful example of TE invasion, mutagenic capacity and the role of the host in evolving controlling mechanisms
(*more next week*)

P- elements have also become major tools of modern *Drosophila* genetics – for mutagenesis and to tag genes for cloning as well as inserting genes transgenically.



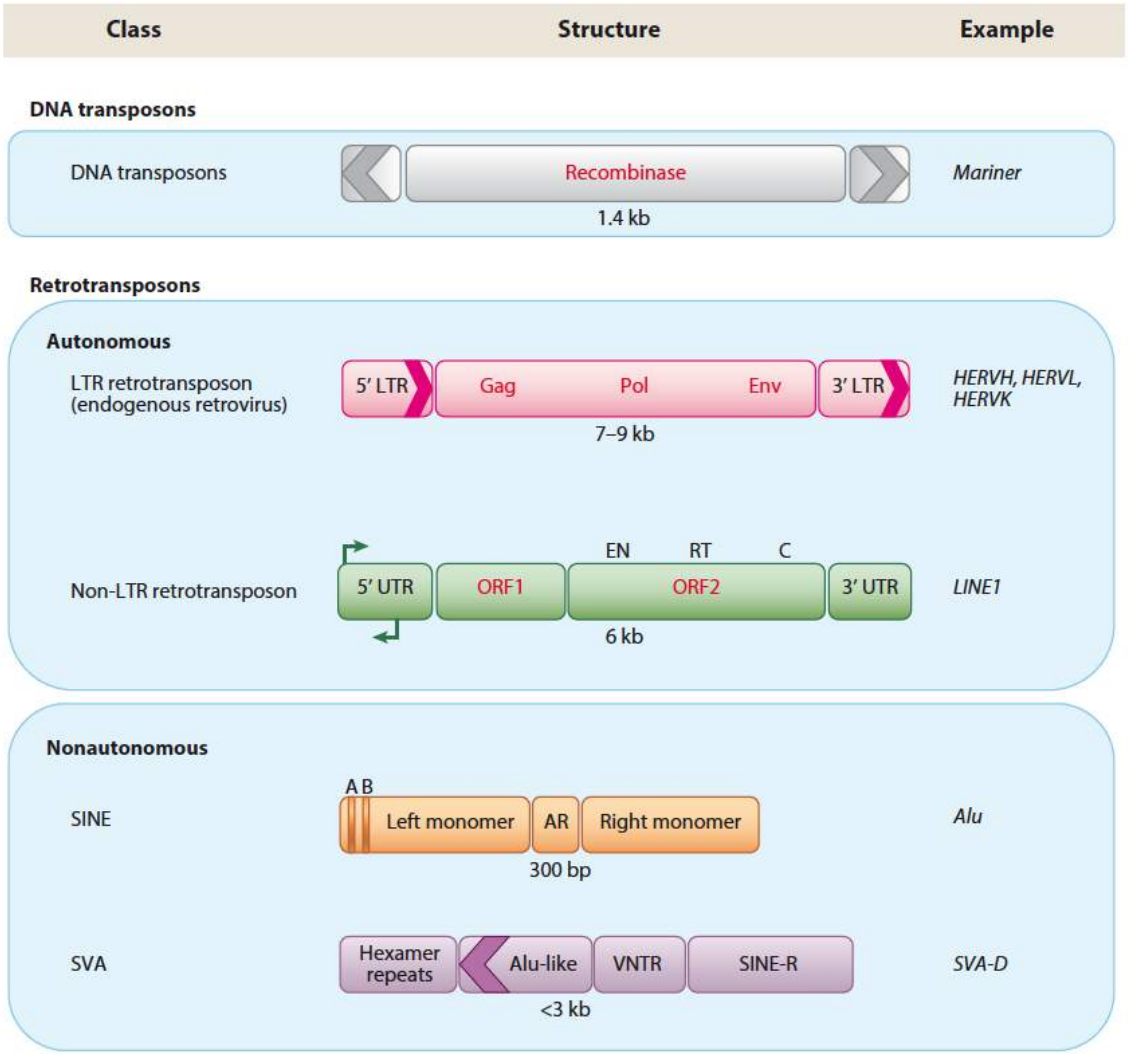
Classes of Transposable Elements

Transposons = DNA sequences that copy or cut themselves out of one part of the genome and reinsert themselves somewhere else.



Haig H. Kazazian Jr. Science 2004;303:1626-1632

Classes of Transposable Elements

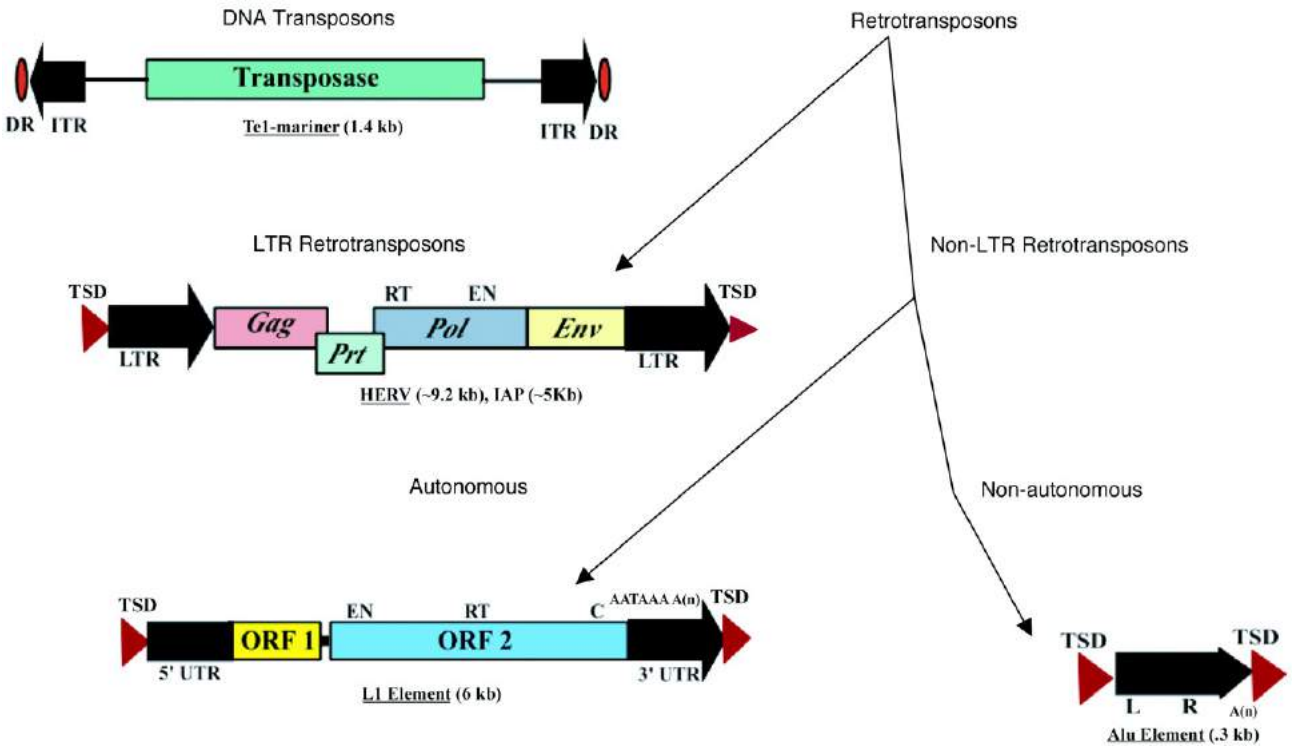


Endogenous Retroviruses (ERVs) share replication strategies reminiscent of viruses - but without the extracellular phase usually => no potential for horizontal transmission.

A significant fraction of TEs present in higher species are endogenous retroviruses (ERVs). ERVs are derived from exogenous relatives that integrated into the germ line, becoming inherited in a Mendelian fashion, and forfeited their ability to spread from cell to cell – usually by losing Env protein functions (Dewannieux & Heidmann 2013).

From Friedl and Trono, 2014

Classes of Transposable Elements



Phylogenetic comparisons of the reverse transcriptase sequences of endogenous retroviruses (ERVs) and exogenous retroviruses point to a common ancestor hundreds of million years ago

(Eickbush & Jamburuthugoda 2008, Xiong & Eickbush 1990).

Haig H. Kazazian Jr. Science 2004;303:1626-1632

Classes of Transposable Elements

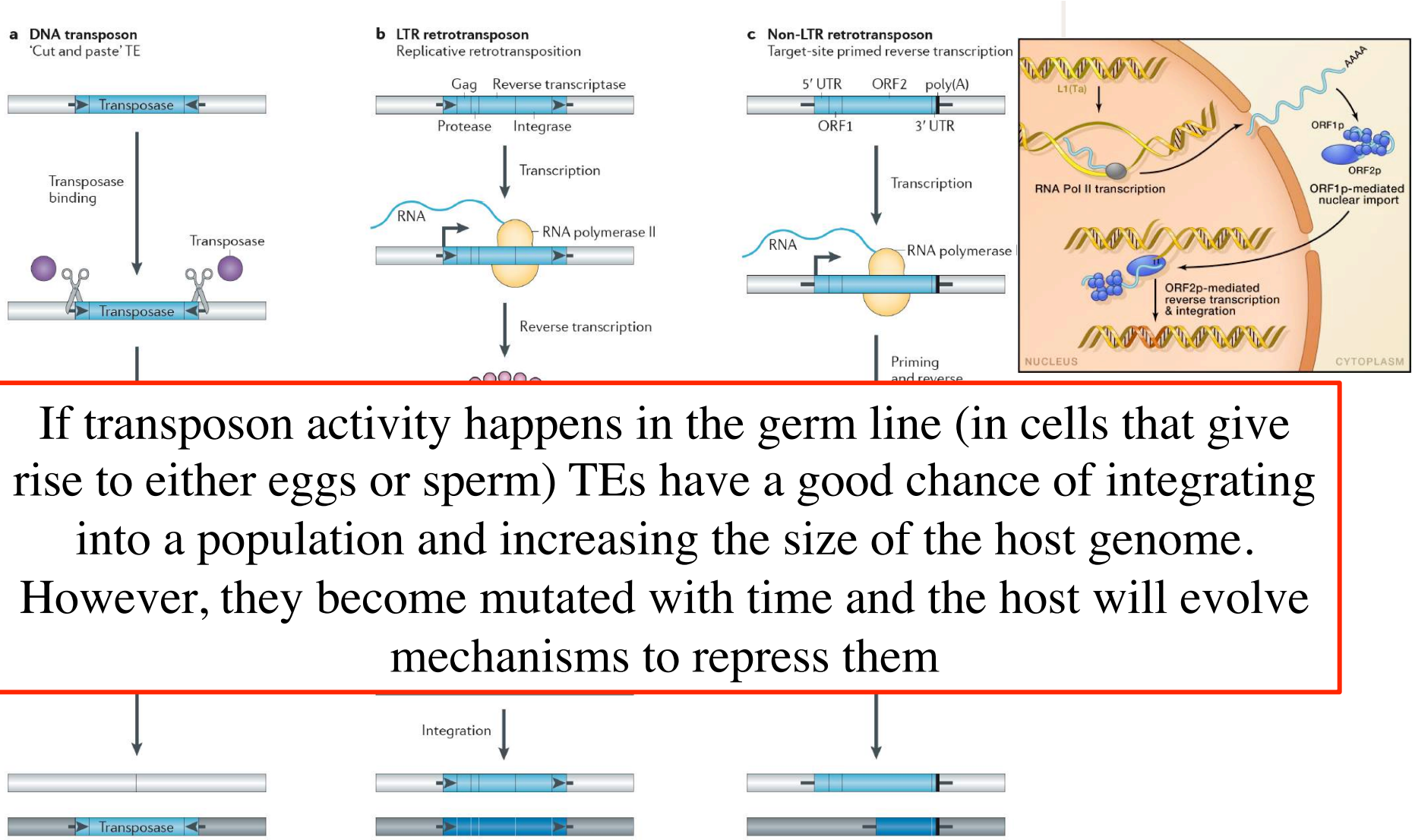
Table 1 | **Classes of transposable elements and their mobility mechanisms**

Class of TE	Structural features	Replication mechanism	Variant forms	Active examples
DNA transposons	• TIRs • Transposase	Transposase-mediated excision of donor dsDNA followed by insertion into the target site	• Some DNA transposons also mobilize via replicative mechanisms • ssDNA transposons lack TIRs: donor ssDNA is inserted into target-site ssDNA, such as for IS*608* of *Helicobacter pylori*	• Tn*7* in *Escherichia coli* • *P* elements in *Drosophila melanogaster* • *Tc1* elements in *Caenorhabditis elegans*
LTR retrotransposons	• LTRs • Gag, protease, reverse transcriptase and integrase	Within virus-like particles, reverse transcriptase copies the mRNA of the TE into a full-length cDNA; integrase inserts the cDNA into target sites	• Solo LTRs are commonly found in genomes and are a result of LTR–LTR recombination	• *Ty1*, *Ty3* and *Ty5* in *Saccharomyces cerevisiae* • *Tf1* and *Tf2* in *Schizosaccharomyces pombe* • *Tnt1* in tobacco
Non-LTR retrotransposons	• One or two ORFs • 5′ truncations and inversion/deletion (for mammalian L1 elements) • Some end in poly(A) tails (for example, L1s); others do not (for example, R2)	An element-encoded endonuclease mediates TPRT. The endonuclease nicks the DNA at the target site and uses the 3′ nicked end for the primer as it reverse transcribes TE mRNA	• Non-autonomous, non-LTR retrotransposons (for example, *Alu* and SVA elements, as well as other eukaryotic SINEs) rely on the endonuclease and reverse transcriptase of an autonomous non-LTR retrotransposon to mediate retrotransposition • The L1 retrotransposition machinery can also mobilize mRNAs (to generate processed pseudogenes) and certain non-coding RNAs (for example, the U6 snRNA)	• L1 in human, mouse, and other mammals • *I* factor in *D. melanogaster* • Zorro3 in *Candida albicans* • R1 and R2 in insects

L1, long interspersed element 1; LTR, long terminal repeat; SINE, short interspersed element; snRNA, small nuclear RNA; SVA, SINE-R–VNTR–*Alu*; TE, transposable element; TIR, terminal inverted repeat; TPRT, target-site-primed reverse transcription.

From Friedl and Trono, 2014

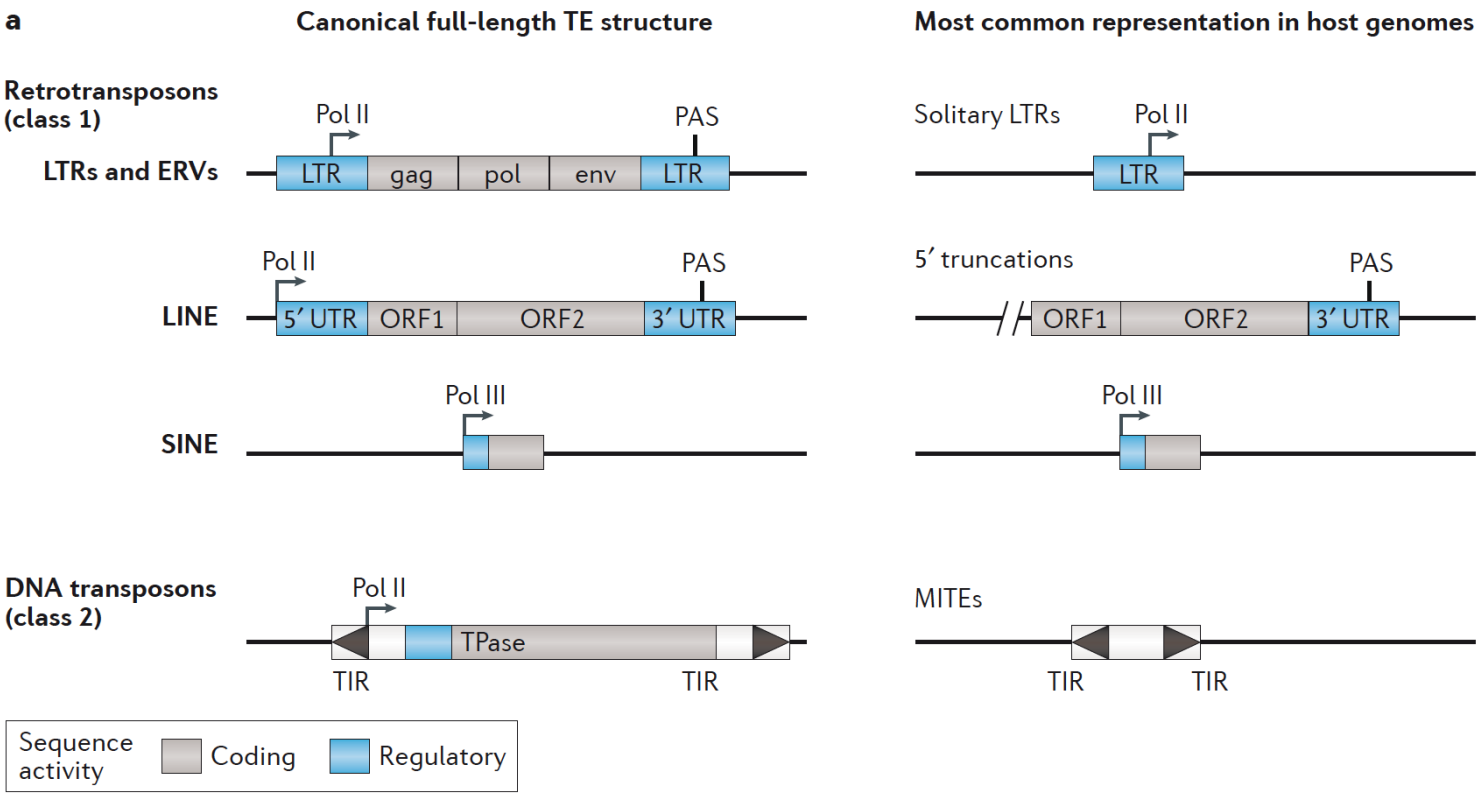
Transposable Elements: How, When and Where do they Move?



If transposon activity happens in the germ line (in cells that give rise to either eggs or sperm) TEs have a good chance of integrating into a population and increasing the size of the host genome. However, they become mutated with time and the host will evolve mechanisms to repress them

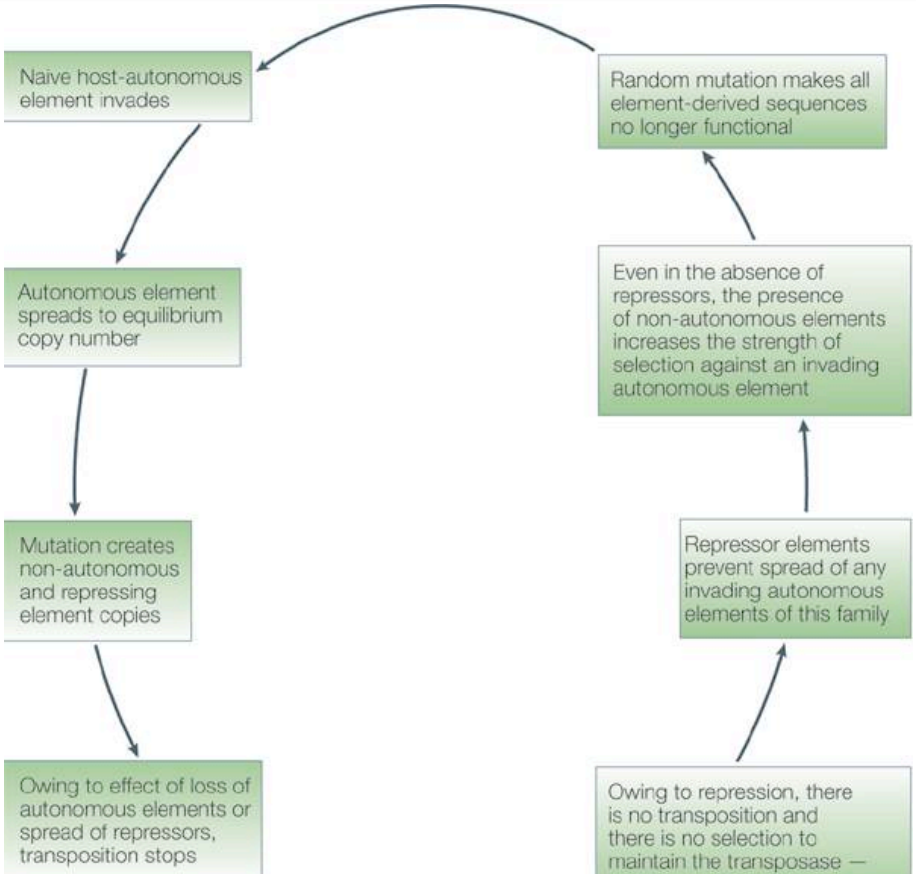
From Levin and Moran, 2011

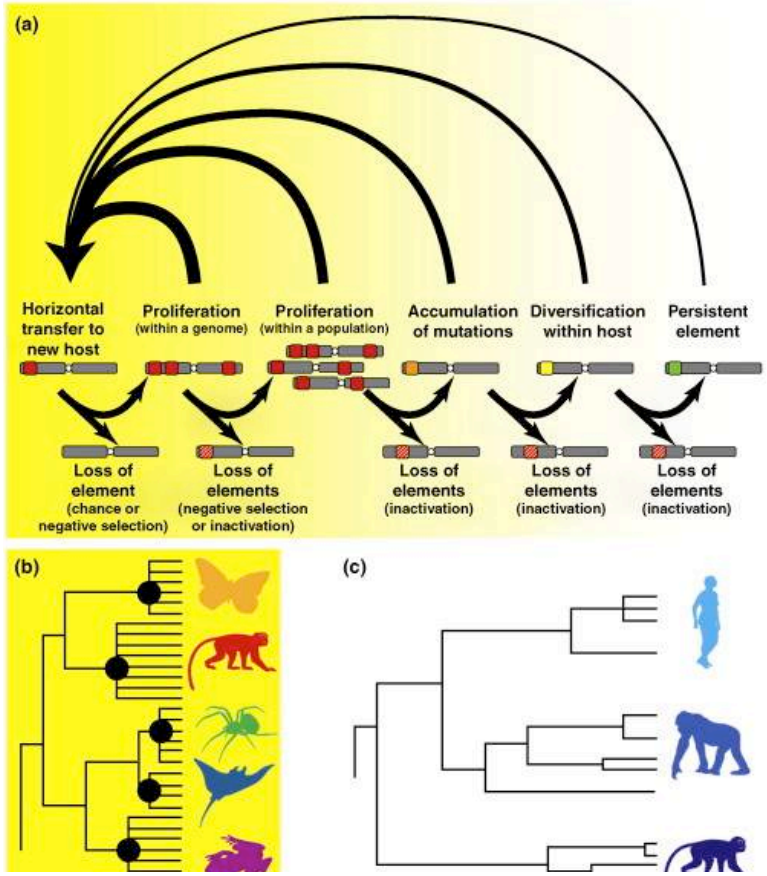
Transposable Elements and their Relics in Genomes



From Chuong et al, 2016

The Lifecycle of a TE family over Evolutionary Time



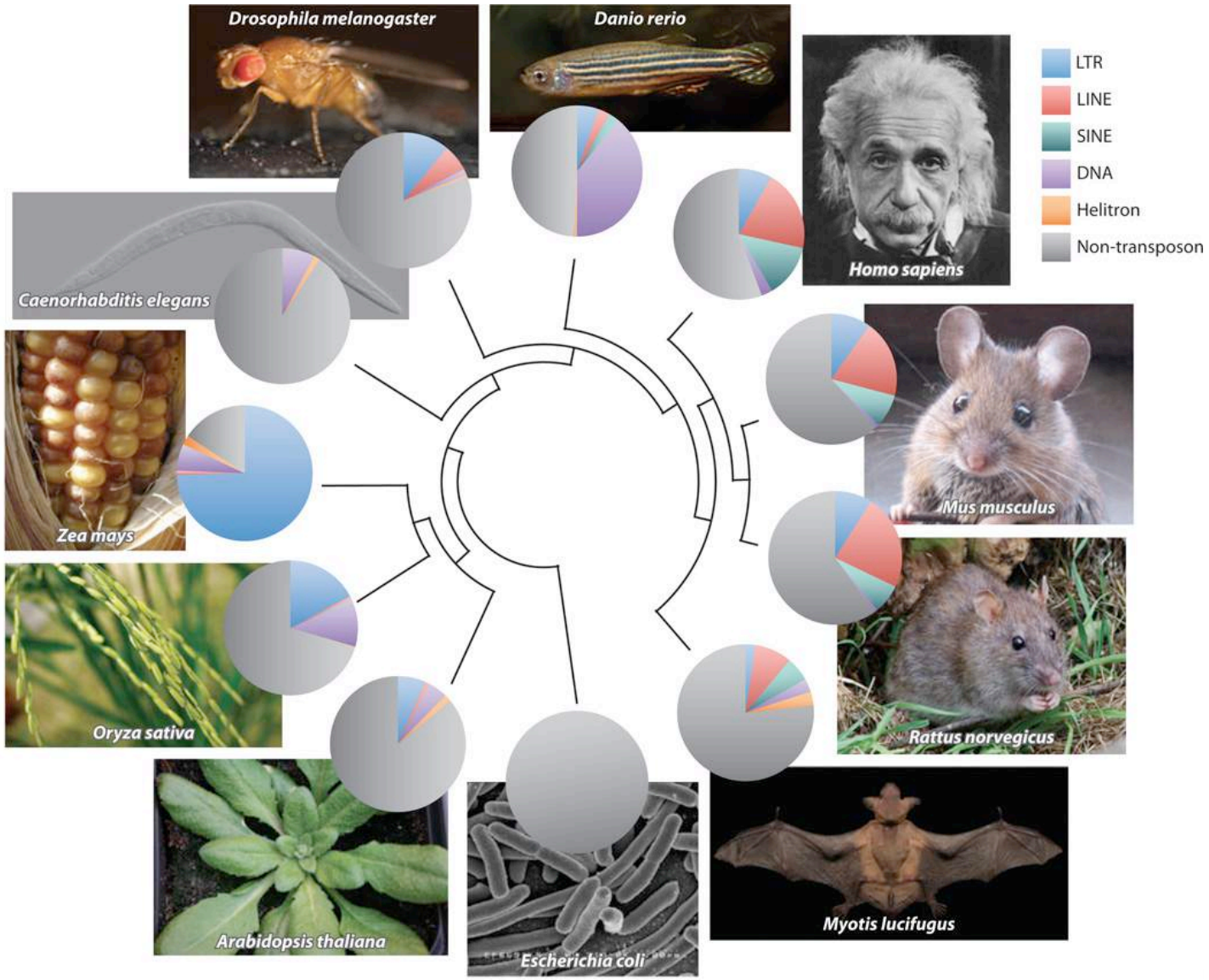


Birth-and-death process: A new TE family is born when an active copy colonizes a novel host genome and it dies when all copies in a lineage are lost (by chance or negative selection) or inactivated, a process which may be driven by **host defense mechanisms** and/or **by the accumulation of disabling mutations in the TE sequence**.

There are two major ways for TEs to **escape extinction**: the first is to **horizontally transfer to a new host genome** prior to inactivation and the second is **to inflict minimal harmful effects** (e.g. low replication rate), so as to evade the eye of selection in their current host. (from *Schaack et al, 2010*).

The LINE-1 element of mammals provides an exceptional example of vertical endurance, having persisted and diversified over the past 100 My with no evidence of horizontal tranfer.

Transposon compositions in different species

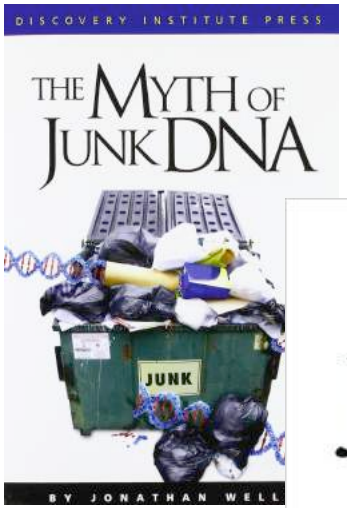


Huang, Burns and Boeke, Ann. Rev. Genetics, 2014

TEs as "Junk" DNA?

"Junk" DNA originally coined by Susumu Ohno (1972), and actually referred to repetitive satellite DNA – but quickly became a generic term for all non-protein coding DNA.

The fact that TEs and their relics were dumped into this category of "junk", and the difficulties implicit in investigating TEs at the molecular level, given their repetitive nature, led to a general lack of interest in transposons for several decades.

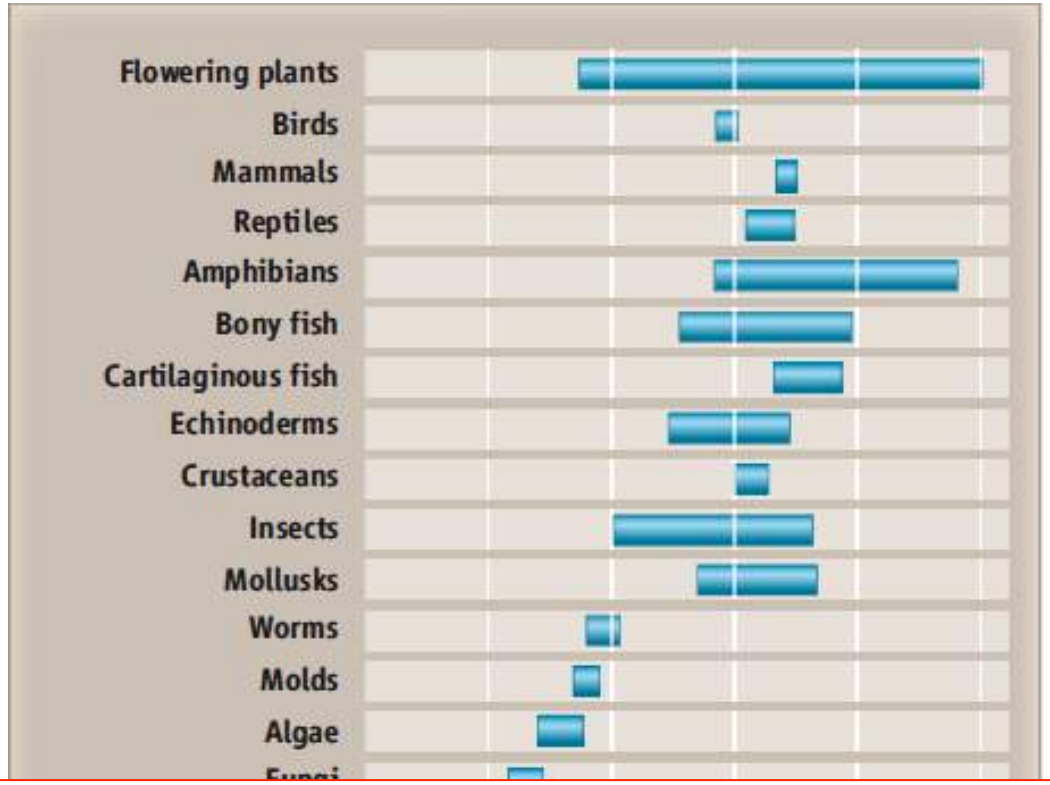


Nevertheless, the high copy number of TEs that were present in most eukaryotic genomes and the fact that they could propagate themselves "selfishly" - potentially leading to massive increases in genome size – meant that they were considered as the explanation for the C-value Paradox…

The C Paradox and "Junk" DNA

The size of an organism's genome does not reflect gene number and is not correlated with its obvious complexity.

C-value is the amount, in picograms, of DNA within a haploid nucleus



Solution to the C-paradox = Selfish DNA?

ie DNA existing only for itself without contributing to an organism's fitness.

The Selfish DNA debate

Selfish DNA: the ultimate parasite

L. E. Orgel & F. H. C. Crick
The Salk Institute, 10010 N. Torrey Pines Road, La Jolla, California 92037

The DNA of higher organisms usually falls into two classes, one specific and the other comparatively nonspecific. It seems plausible that most of the latter originated by the spreading of sequences which had little or no effect on the phenotype. We examine this idea from the point of view of the natural selection of preferred replicators within the genome.

The familiar neo-darwinian theory of natural selection is concerned with the competition between organisms in a population. At the level of molecular genetics it provides an explanation of the spread of 'useful' genes or DNA sequences within a population. Organisms that carry a gene that contributes positively to fitness tend to increase their representation at the expense of organisms lacking that gene. In time, only those organisms that carry the useful gene survive. Natural selection also predicts the spread of a gene or other DNA sequence within a single genome, provided certain conditions are satisfied. If an organism carrying several copies of the sequence is fitter than an organism carrying a single copy, and if mechanisms exist for the multiplication of the relevant sequence, then natural selection must lead to the emergence of a population in which the sequence is represented several times in every genome.

The idea of selfish DNA is different. It is again concerned with the spread of a given DNA within the genome. However, in the case of selfish DNA, the sequence which spreads makes no contribution to the phenotype of the organism, except insofar as it is a slight burden to the cell that contains it. Selfish DNA sequences may be transcribed in some cases and not in others. The spread of selfish DNA sequences within the genome can be compared to the spread of a not-too-harmful parasite within its host.

REVIEW ARTICLES

Selfish genes, the phenotype paradigm and genome evolution

W. Ford Doolittle & Carmen Sapienza

Necessary and unnecessary explanations

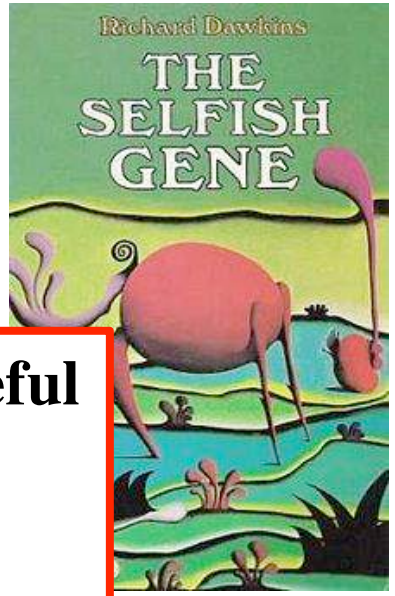
We do not deny that prokaryotic transposable elements or repetitive and unique-sequence DNAs not coding for protein in eukaryotes may have roles of immediate phenotypic benefit to the organism. Nor do we deny roles for these elements in the evolutionary process. We do question the almost automatic invocation of such roles for DNAs whose function is not obvious, when another and perhaps simpler explanation for their origin and maintenance is possible. It is inevitable that natural selection of the special sort we call non-phenotypic will favour the development within genomes of DNAs whose only 'function' is survival within genomes. When a given DNA, or class of DNAs, of unproven phenotypic function can be shown to have evolved a strategy (such as transposition) which ensures its genomic survival, then no other explanation for its existence is necessary. The search for other explanations may prove, if not intellectually sterile, ultimately futile.

Limits on the selfishness of DNA

SIR — Doolittle, Kirkwood and Dempster recently argued (*Nature* **307**, 501; 1984) that some "selfish" DNAs (for example, elements which transpose duplicatively) may restrain their own reproduction within genomes to avoid driving their hosts, and hence themselves, into extinction. They suggested that such self-restraint should be expected to evolve in elements that impose a cost on their host's fitness which increases with copy number. That is, copy number should be limited so that host mortality does not exceed host reproduction. This view is fatally flawed, as it ignores a fundamental constraint on the evolution of "selfish" DNA.

The Selfish DNA debate

The notion of Selfish DNA to explain the C-value paradox (in line with the "Selfish Gene" by R. Dawkins) countered the neo-Darwinian views that all DNA of an organism's genome must have been kept



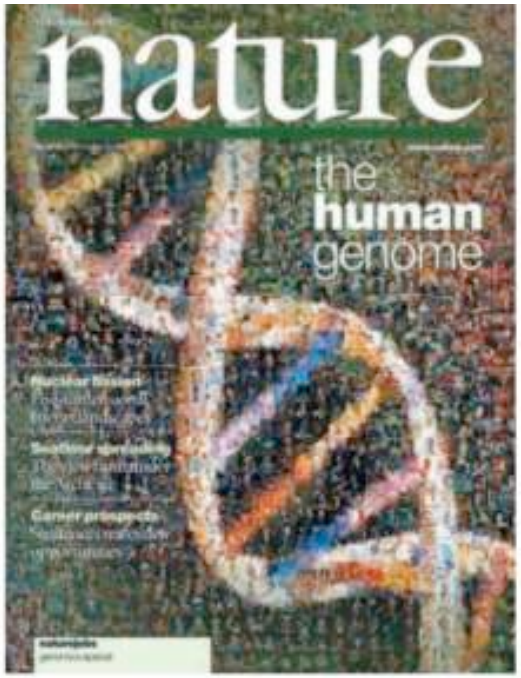
TEs may be selfish – but they might also be useful to their hosts… Otherwise why would they accumulate in such great quantities?

Sydney Brenner

I said it was 'junk' DNA, not 'trash'. Everyone knows that you throw away trash. But junk we keep in the attic until there may be some need for it…

Hence
becar
of jun
by mu
were
themselves that were usually inserted harmlessly back into the genome.

Sequencing of the Human Genome



Three billion DNA base pairs

20,000 protein-coding genes

Less than 2 % of the genome… > 98% non-coding DNA!

Much of this contains repeats and relics / fossils of transposable elements

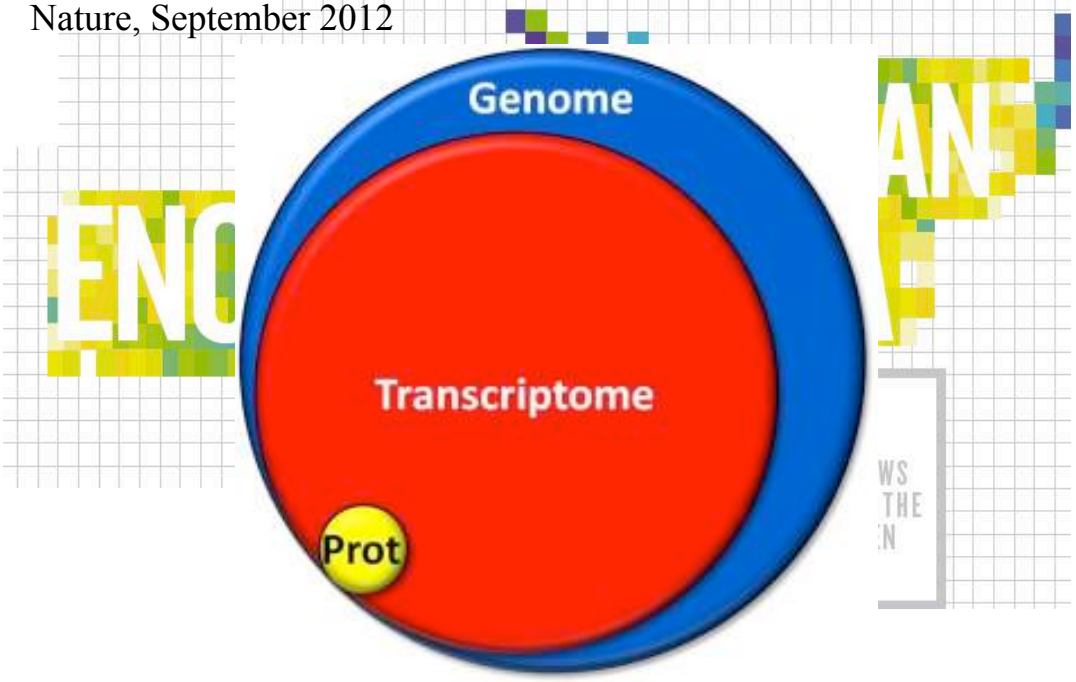
However, TE numbers are still vastly underestimated owing to the bioinformatics challenge of accurately detecting repeats - especially small/degenerate relics!

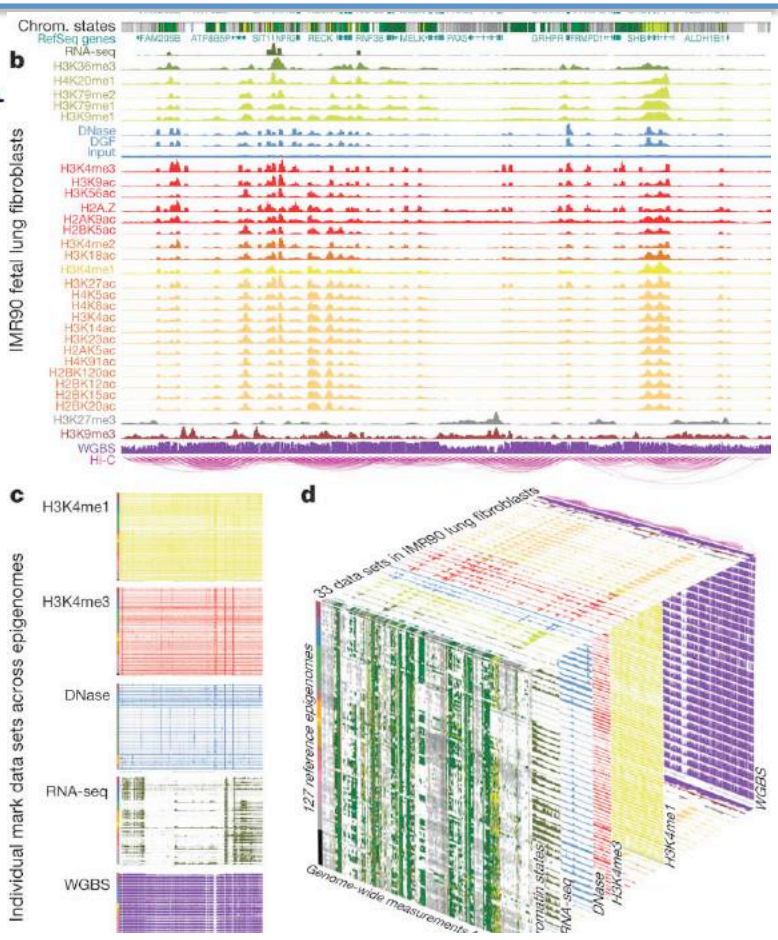
The Coding (1.5%) and Non-Coding (98.5%) Genome

An integrated encyclopedia of DNA elements in the human genome

The ENCODE Project Consortium*

Nature, September 2012





According to ENCODE's analysis, 80% of the genome has a "biochemical function" *

It's not all "junk"…

* "biochemical function" = sequences that have proteins (specifically) attached to them, those that affects how DNA is packaged and those that are transcribed…

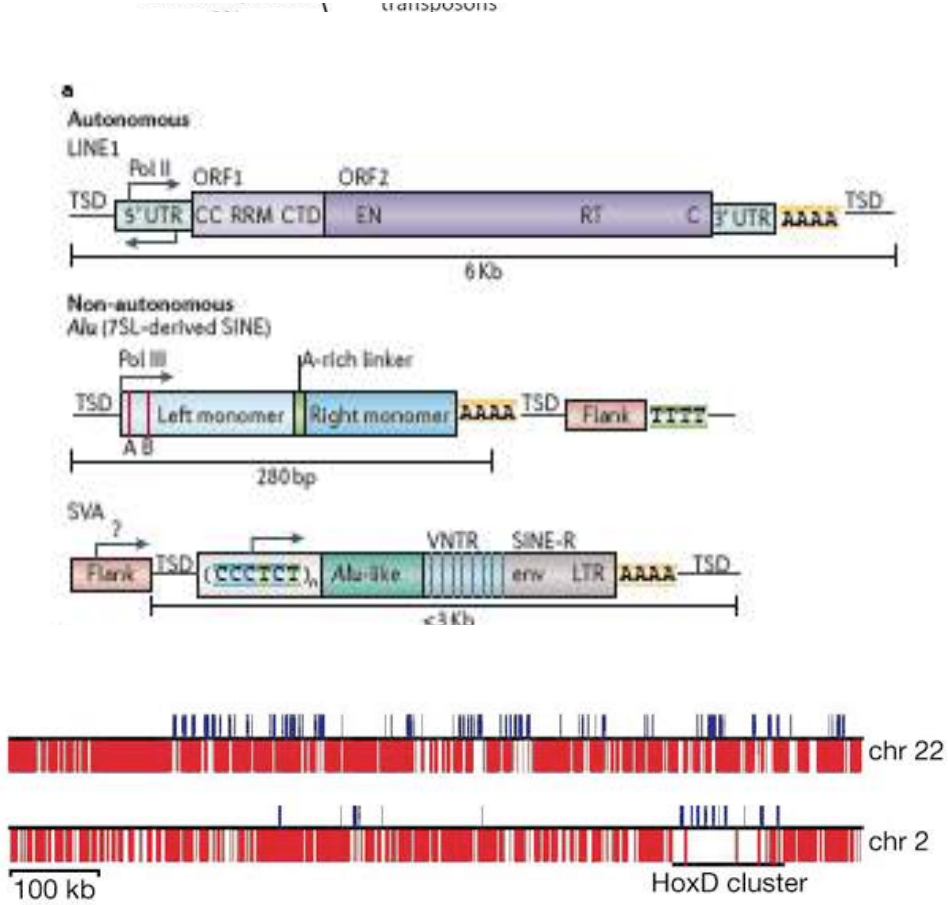
Diversity and Distribution of TEs in the Human Genome

Table 11 Number of copies and fraction of genome for classes of interspersed repeat

	Number of copies (× 1,000)	Total number of bases in the draft genome sequence (Mb)	Fraction of the draft genome sequence (%)	Number of families (subfamilies)
SINEs	1,558	359.6	13.14	
Alu	1,090	290.1	10.60	1 (~20)
MIR	393	60.1	2.20	1 (1)
MIR3	75	9.3	0.34	1 (1)
LINEs	868	558.8	20.42	
LINE1	516	462.1	16.89	1 (~55)
LINE2	315	88.2	3.22	1 (2)
LINE3	37	8.4	0.31	1 (2)
LTR elements	443	227.0	8.29	
ERV-class I	112	79.2	2.89	72 (132)
ERV(K)-class II	8	8.5	0.31	10 (20)
ERV (L)-class III	83	39.5	1.44	21 (42)
MaLR	240	99.8	3.65	1 (31)
DNA elements	294	77.6	2.84	
hAT group				
MER1-Charlie	182	38.1	1.39	25 (50)
Zaphod	13	4.3	0.16	4 (10)
Tc-1 group				
MER2-Tigger	57	28.0	1.02	12 (28)
Tc2	4	0.9	0.03	1 (5)
Mariner	14	2.6	0.10	4 (5)
PiggyBac-like	2	0.5	0.02	10 (20)
Unclassified	22	3.2	0.12	7 (7)
Unclassified	3	3.8	0.14	3 (4)
Total interspersed repeats		1,226.8	44.83	

The number of copies and base pair contributions of the major classes and subclasses of transposable elements in the human genome. Data extracted from a RepeatMasker analysis of the draft genome sequence (RepeatMasker version 09092000, sensitive settings, using RepBase Update 5.08). In calculating percentages, RepeatMasker excluded the runs of Ns linking the contigs in the draft genome sequence. In the last column, separate consensus sequences in the repeat databases are considered subfamilies, rather than families, when the sequences are closely related or related through intermediate subfamilies.

Human Retrotransposons still mobile?

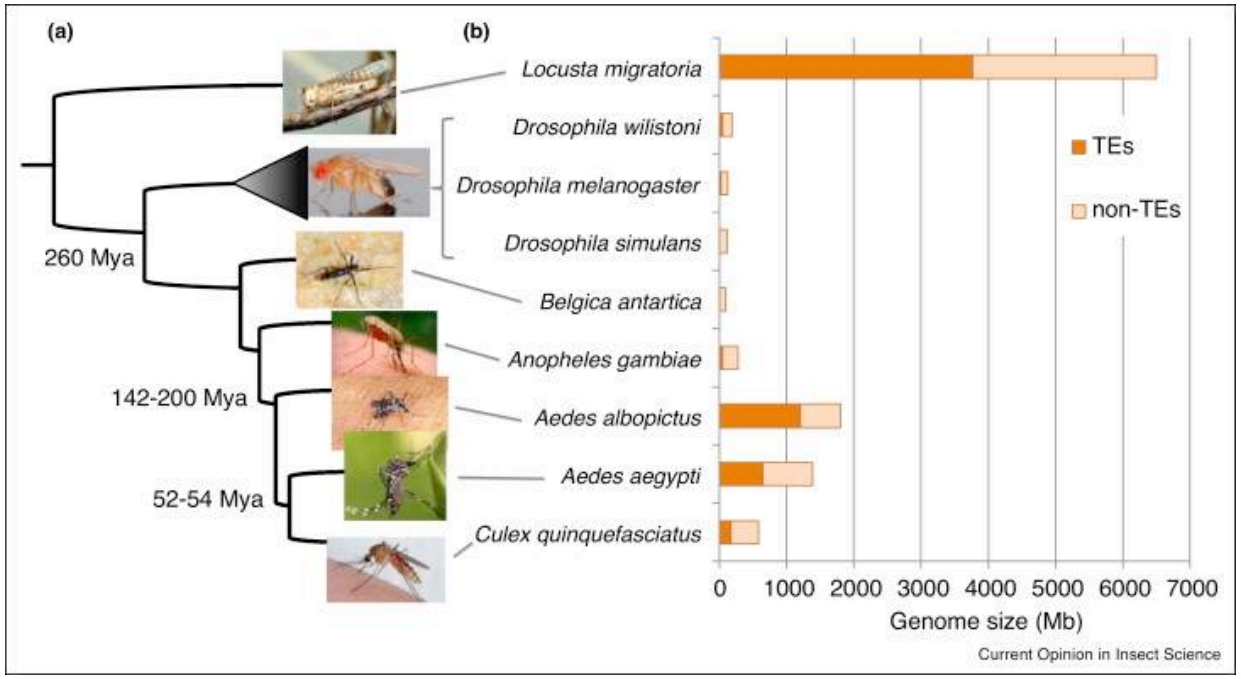


TEs are distributed throughout the genome
Depleted in protein-coding regions but still present within genes and around them in most cases - with some exceptions (eg *Hox* genes)

Lander et al, Nature 2001 – The Human

Sequencing of numerous Eukaryotic genomes has revealed that TEs profoundly influence the shape, size and functions of genomes…

TEs are paramount in genome size variation



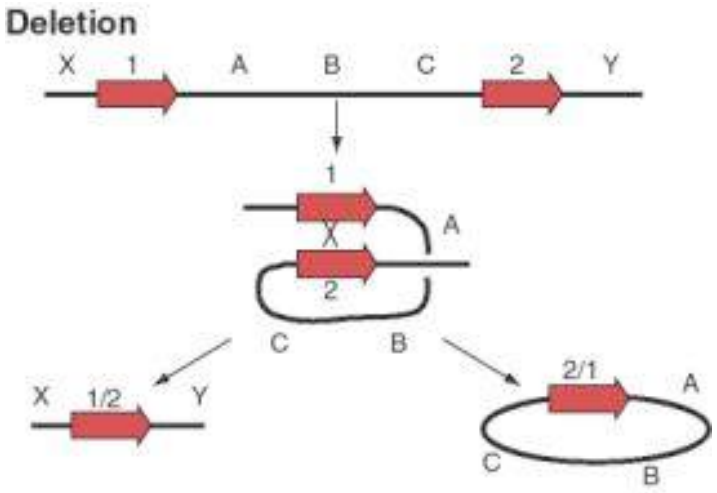
Eukaryotic Transposable Elements and Genome Evolution Special Feature: Dramatic amplification of a rice transposable element during recent domestication

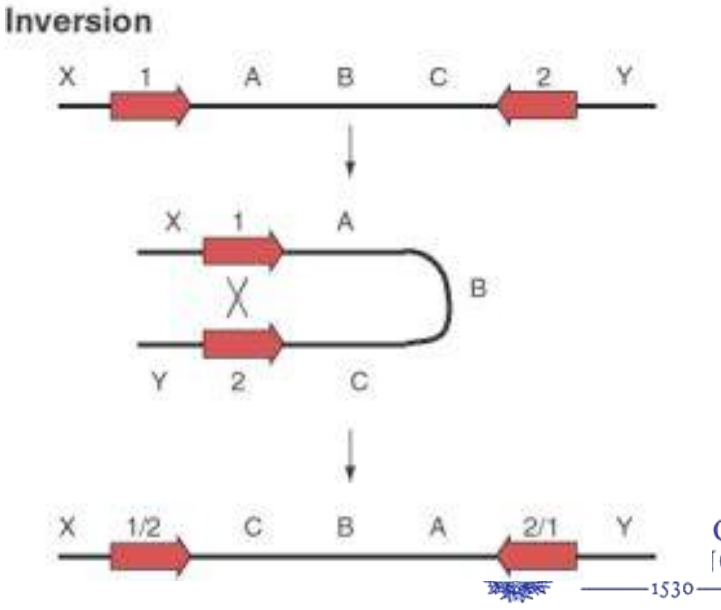
Ken Naito, Eunyoung Cho, Guojun Yang, Matthew A Campbell, Kentaro Yano, Yutaka Okumoto, Takatoshi Tanisaka, and Susan R. Wessler

PNAS 2006;103;17620-17625; originally published online Nov 13, 2006; doi:10.1073/pnas.0605421103

Lessons on TEs in the post-Genomics Era

- TEs, mobile genetic elements, or jumping genes
- Parasitic, self-replicating
- Similar to, or derived from viruses
- Move independently in a genome
- Create new copies that can trigger mutations, recombination, deletions, duplications…





Lessons on TEs in the post-Genomics Era

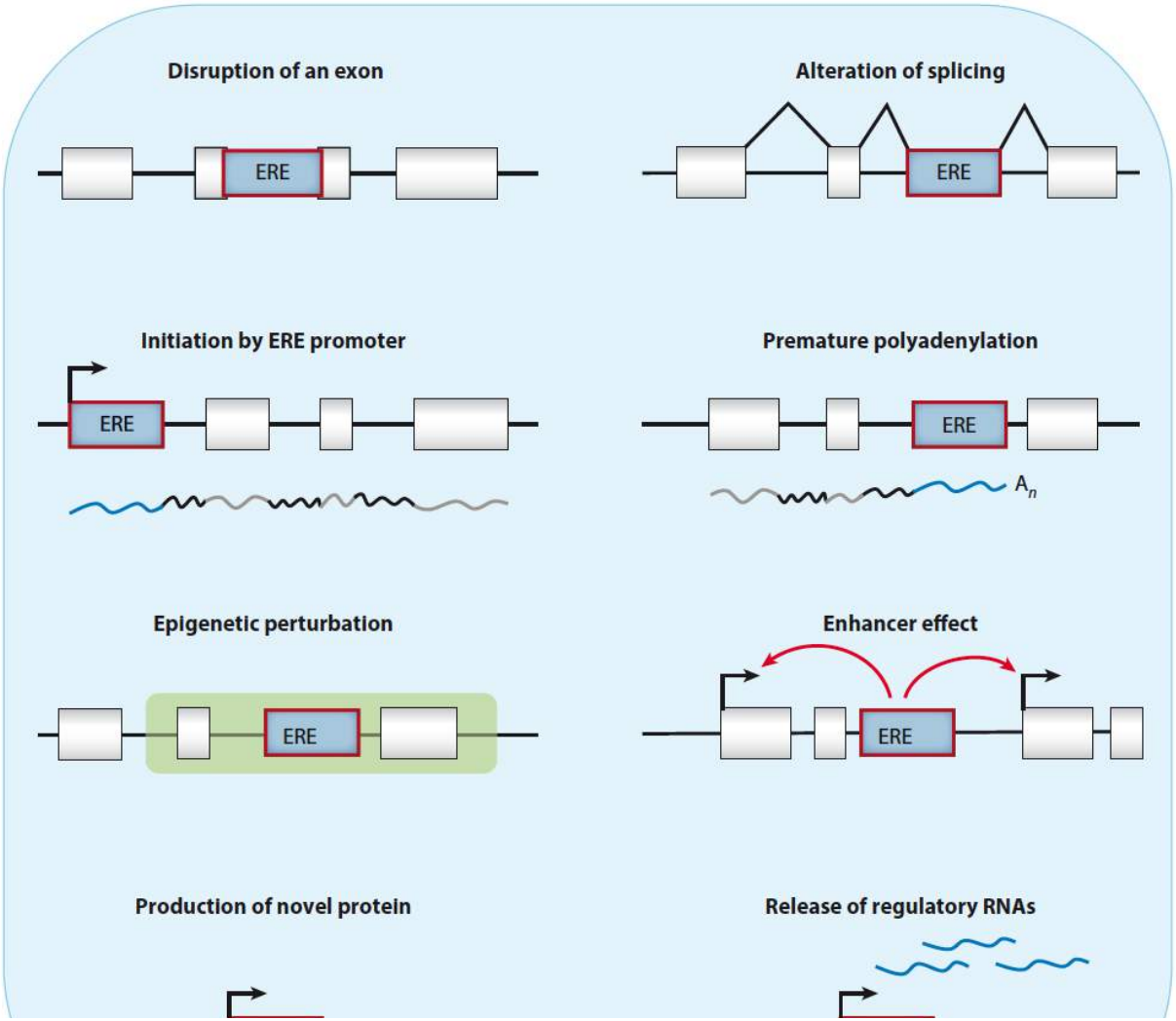
- TEs, mobile genetic elements, or jumping genes
- Parasitic, self-replicating
- Similar to, or derived from viruses
- Move independently in a genome
- Create new copies that can trigger mutations, recombination, deletions, duplications…
- Most TEs are broken (cannot tranpose; "fossils").
- Active TEs evolved to insert into "safe-havens" – but that are sensitive to environment
- Host regulates TE movement: host defense mechanisms including epigenetic strategies
- TEs can provide advantages in an evolutionary setting
- Populations of TE sequences in a genome evolve
- Surrounding genomic sequences *also* evolve
- Diseases due to TEs can occur, but are outnumbered by examples of positive impact - as expected from *millions of years of purifying selection*…

TEs are intimate components of our genome:

Rather than just being graveyards of dead TE fossils
eukaryotic genomes have a rich repository of functional and gene regulatory potential, thanks to TEs!

Transposable Elements (TEs) as Generators of Genetic Diversity and Modulators of Gene Expression

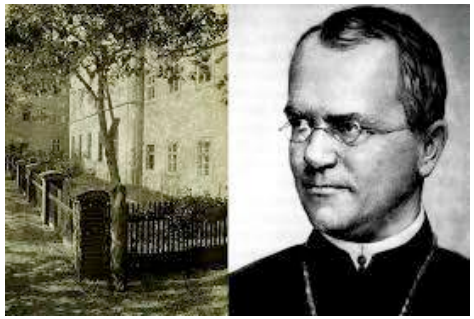
- TEs are probably the most powerful genetic force engaged in the evolution of higher species.
- They sprinkle genomes with thousands of identical sequences, paving the way for recombination events that can trigger deletions or duplications
- They can disrupt existing genes but also provide new protein-coding sequences
- They exert a wide range of transcriptional influence, either directly or via host mechanisms responsible for their control



TEs as targets and drivers of epigenetic processes

COURS II

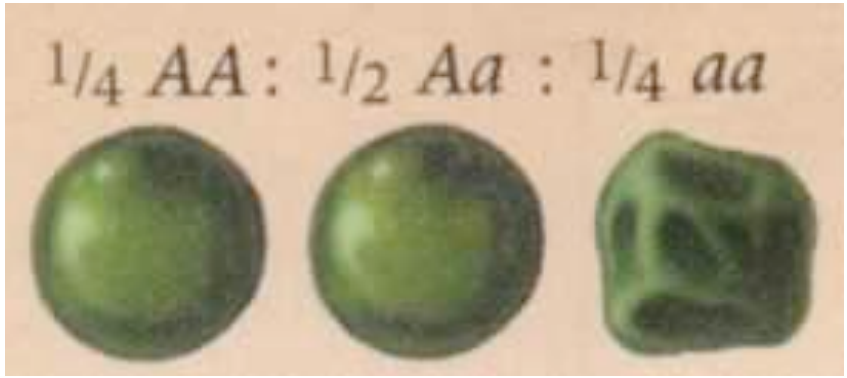
Transposable Elements (TEs) as Generators of Genetic and Phenotypic Variation

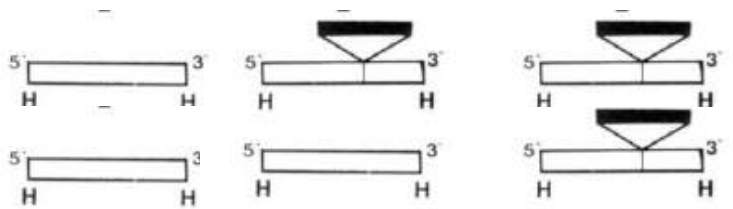


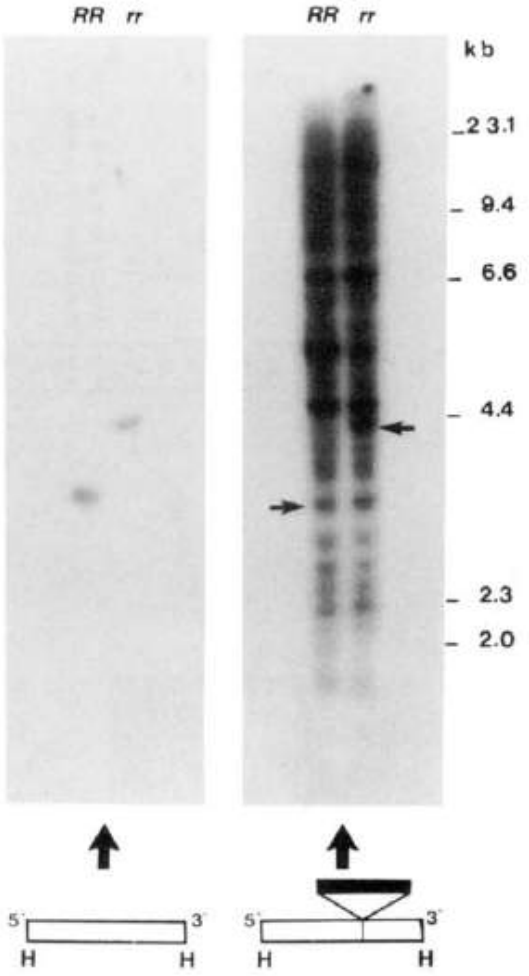
Cell, Vol. 60, 115–122, January 12, 1990, Copyright © 1990 by Cell Press

The Wrinkled-Seed Character of Pea Described by Mendel Is Caused by a Transposon-like Insertion in a Gene Encoding Starch-Branching Enzyme

Madan K. Bhattacharyya, Alison M. Smith,
T. H. Noel Ellis, Cliff Hedley, and Cathie Martin



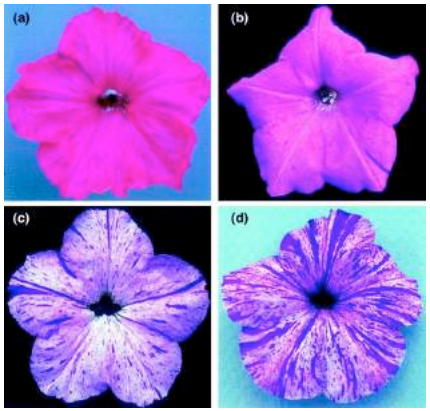




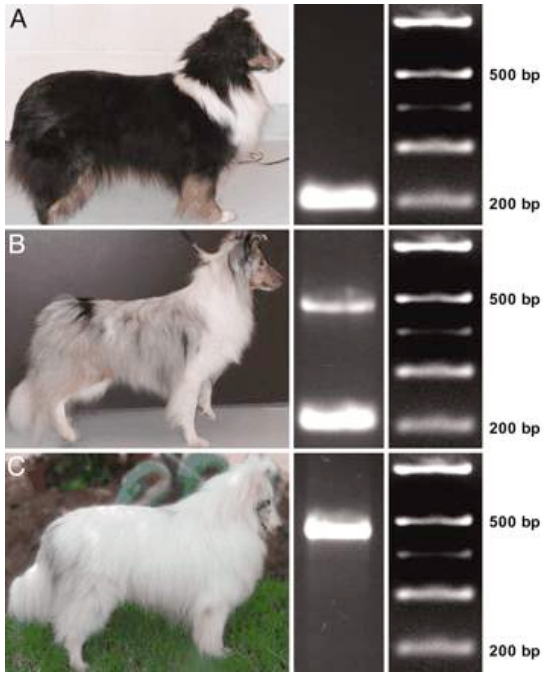
COLLÈGE DE FRANCE 1530

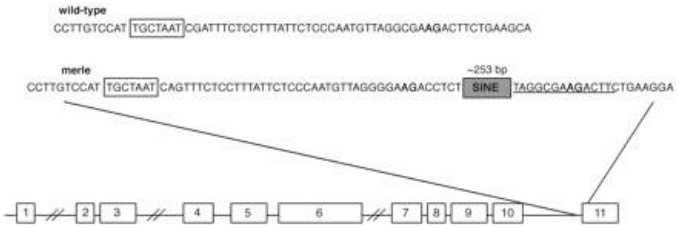
Transposable Elements (TEs) as Generators of Genetic and Phenotypic Variation

Petunia flowers
Psl transposon in *Hf1* gene



Shetland Sheepdogs
SINE insertion in *SILV* gene



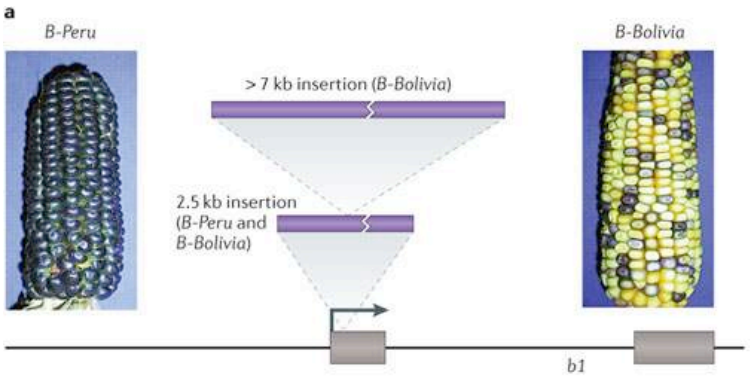


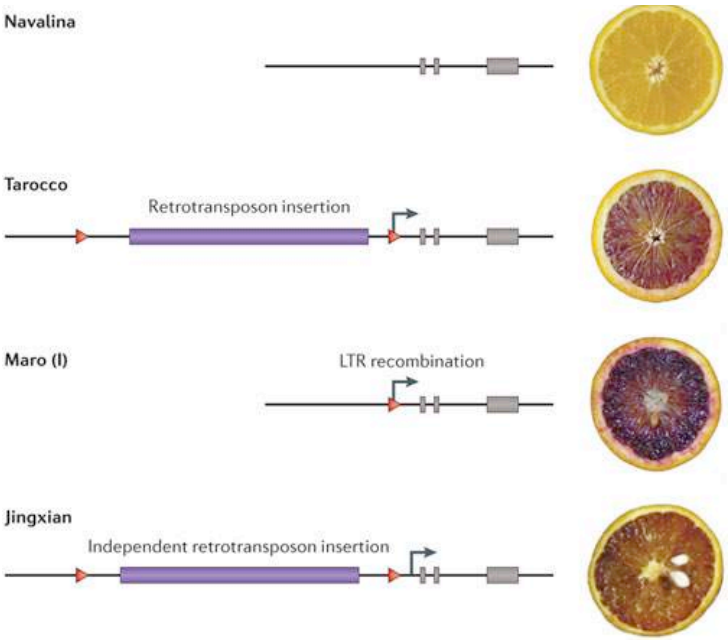
Clark et al., PNAS 2006

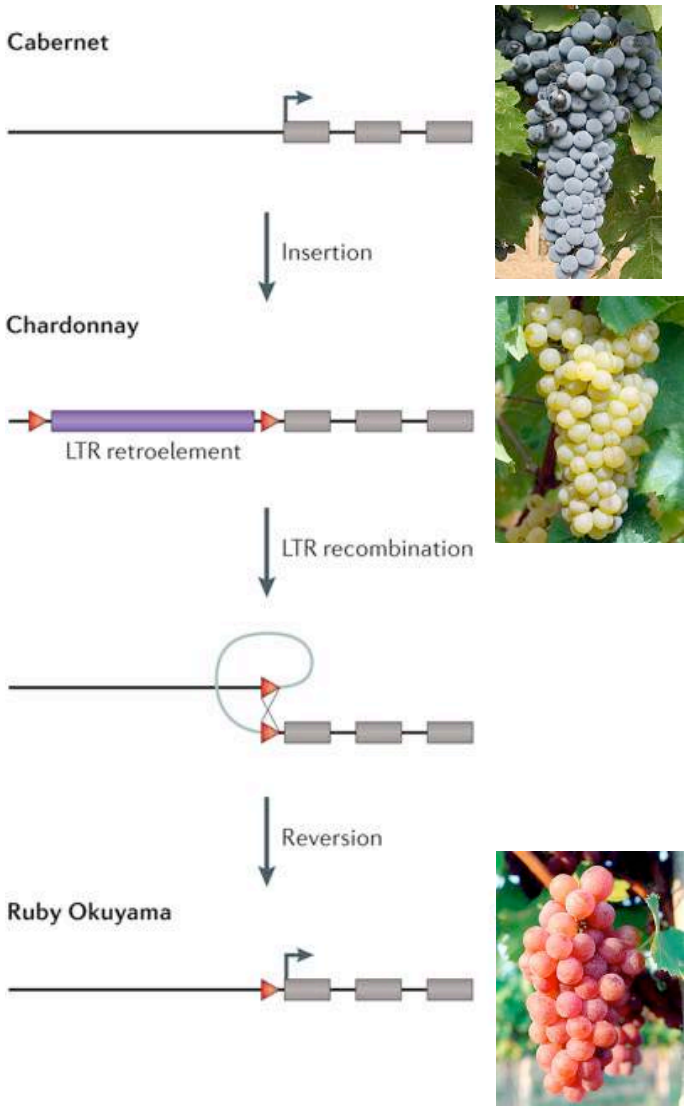
Peppered moth
Insertion of a type II transposon the *cortex* gene

Van't Hof et al., Nature, 2016

Transposable Elements (TEs) as Generators of Genetic and Phenotypic Variation







Copia-like retrotransposon adjacent to a gene encoding Ruby, a MYB transcriptional activator of anthocyanin production.

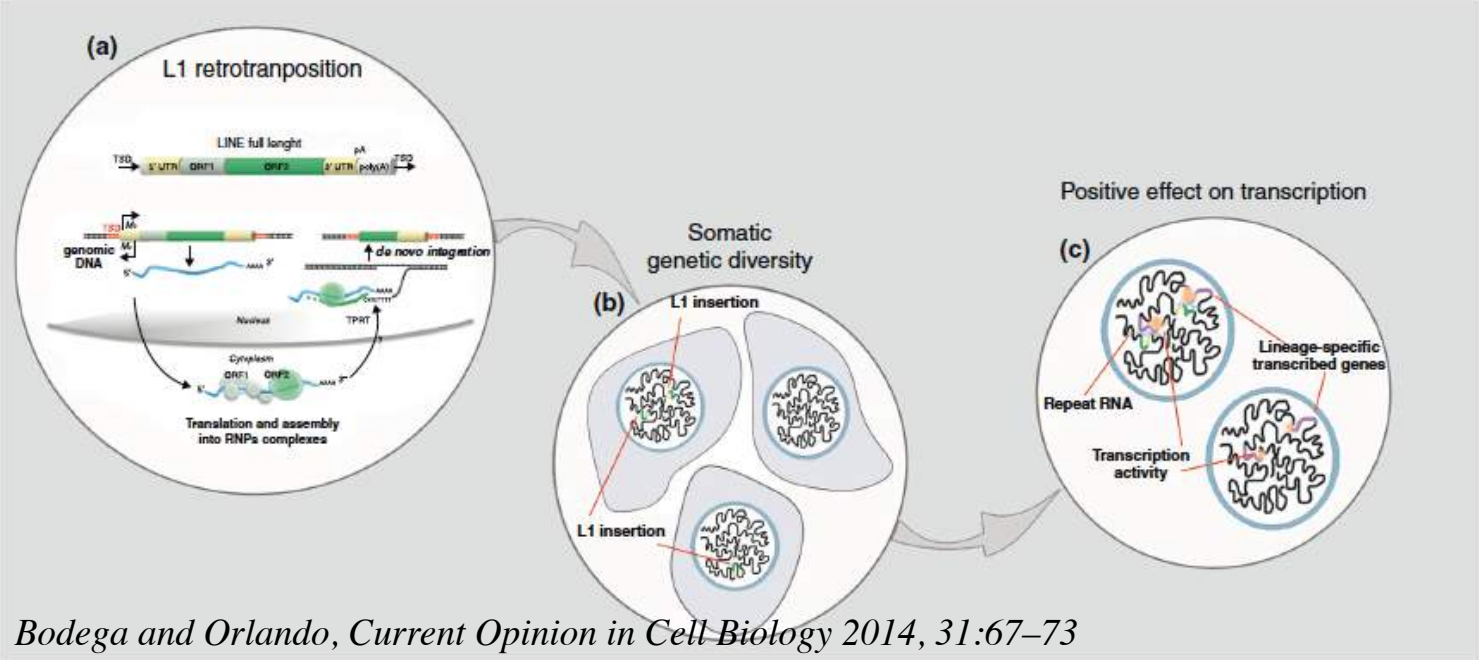
The TE controls *Ruby* expression, and cold dependency reflects the induction of the retroelement by stress.

Nature Revi Genetics

COLLÈGE DE FRANCE 1530

E. Heard, February 6th, 2017

Transposable Elements (TEs) as Generators of Somatic Genetic and/or Epigenetic Variation

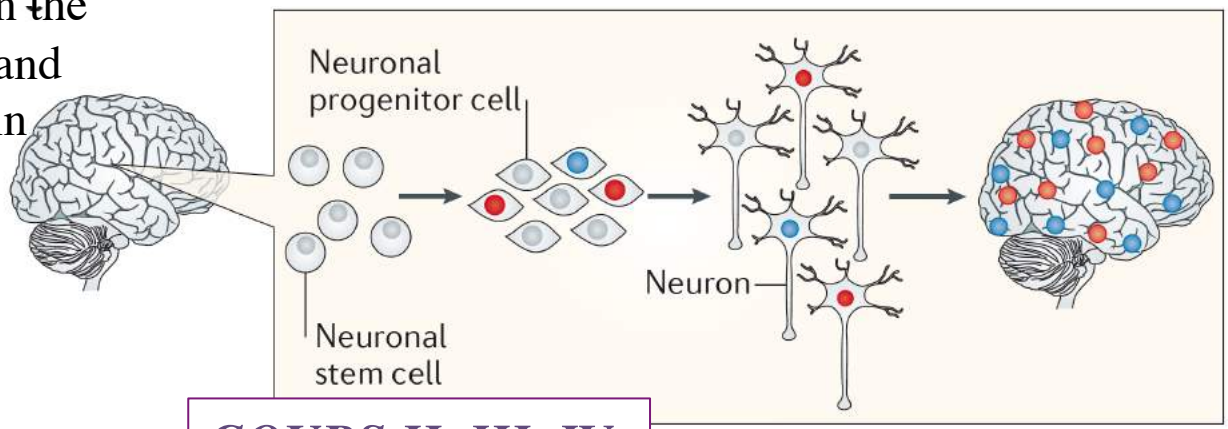


Bodega and Orlando, Current Opinion in Cell Biology 2014, 31:67–73

Current Opinion in Cell Biology

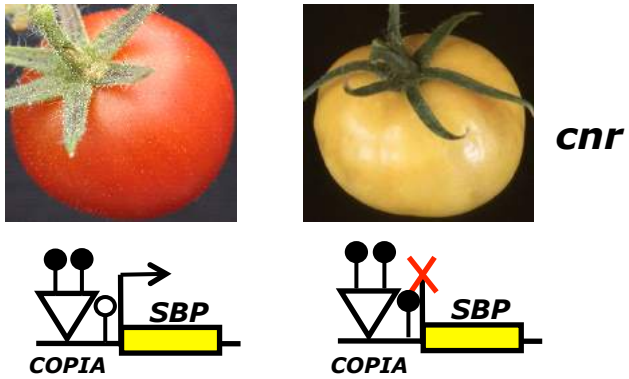
Mobile DNA elements in the generation of diversity and complexity in the brain

Muotri et al, 2005



COURS II, III, IV

Transposable Elements (TEs) as Generators of Epigenetic Phenotypic Variation (*Epialleles*)



(Manning et al, Nat Genet, 2006)

SBP encodes a transcription factor that allows ripening (red). In the *cnr* mutant a TE is integrated upstream of the promoter of *SBP*. The TE is constitutively methylated but its methylation can spread to the promoter of the gene and correlates with its silencing preventing ripening (yellow).

Silencing is **metastable** in somatic tissues – but fully stable through meiotic transmission.

(Rakyan et al, PNAS 2003)

(Morgan et al, Nat. Genet. 1999)

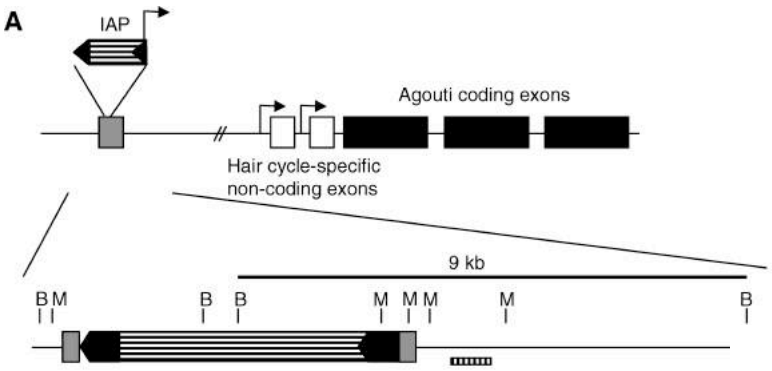
Two *bone fide* trans-generational **epimutations** in mammals (Axin-fused and Agouti) are also associated with IAPs (retrotransposons)

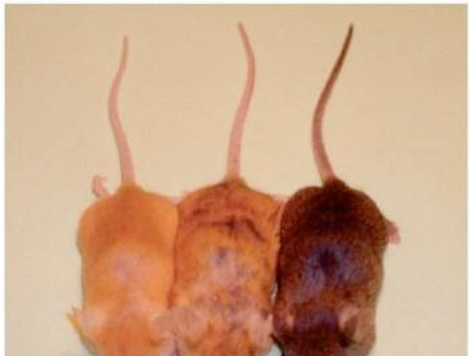
Ineffecient reprogramming in the germ line?

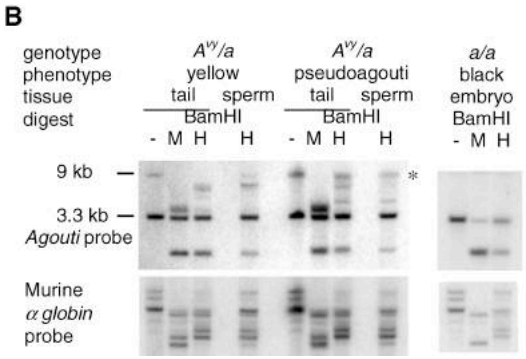
COURS III

Transposable Elements (TEs) as Generators of Epigenetic Phenotypic Variation (*Epialleles*)

TEs attract epigenetic marking, providing phenotypic variation in absence of genotypic variation







And - these states can be influenced by maternal diet

Adult siblings – essentially *identical* genomes
Differ by DNA methylation at just one TE locus…

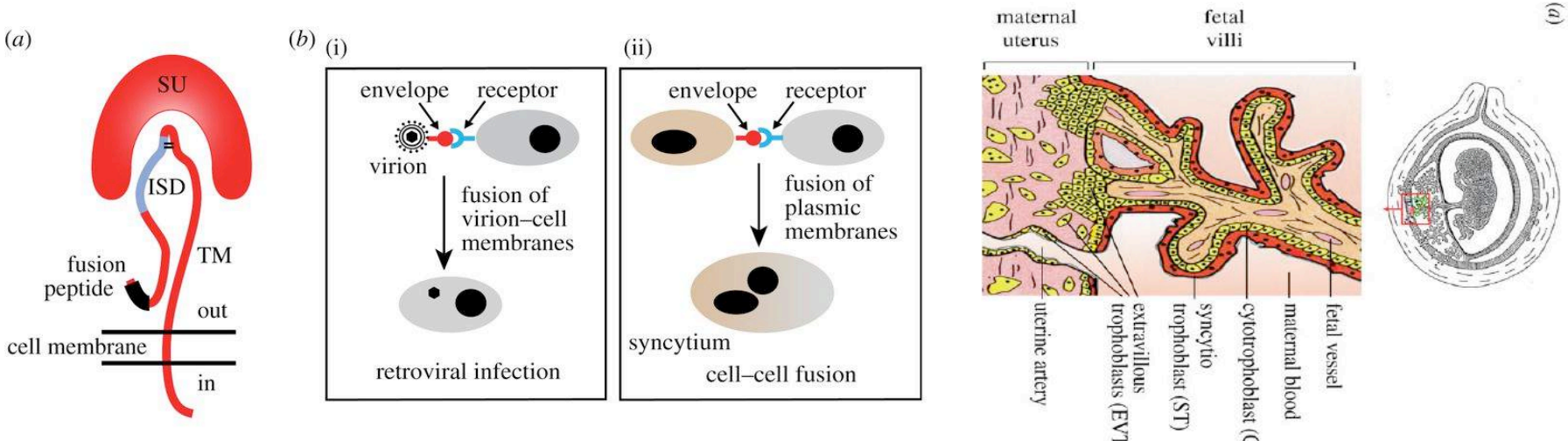
Exaptation

The role of TEs in the evolution of:

New genes (Syncitin, RAG1/2)
New functions (Placentation; Immune systems)
New gene regulatory elements and networks
Epigenetic processes (imprinting, X inactivation…)

Domestication of TEs to generate new genes and functions

• **Placentation**: Syncytin genes are required for the formation of the syncytiotrophoblast double layer. Humans and mice have two *syncytin* genes, all four corresponding to the *env* gene of ERVs, which entered these species on separate occasions a few tens of Mya (Dupressoir et al. 2012). Exaptation of a founding retroviral env gene may have led to emergence of mammalian ancestors with a placenta from egg-laying animals; subsequently replaced in diverse mammalian lineages by new env-derived syncytin genes, each providing its host with a selective advantage (Lavialle, Heidmann et al, 2012)



• **Immune System**: Recombination-activating genes *RAG1* and *RAG2*, which are critical for V(D)J recombination and immune system development, probably originated from domestication of a member of the *Transib* family of DNA transposons, approximately 500 Mya (Kapitonov & Jurka 2005, Zhou et al. 2004).

• **Viral Defense System**: he murine Fv1 restriction factor is also of retroviral origin and has been co-opted by the host as an antiretrovirus defense mechanism (Best et al. 1996)

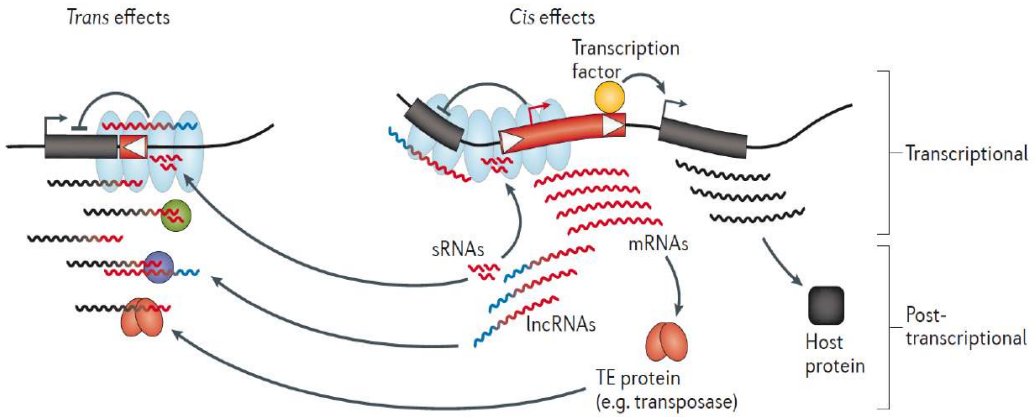
TEs as a source of regulatory potential

TEs provide a dynamic source of of transcriptional regulators:

• In *cis*: TEs contain *cis*-acting regulatory sequences that can influence gene expression through promoter, enhancer, or insulator effects (eg exaptation of ERV enhancers appears to have contributed to the rapid evolutionary diversification of the placenta (Chuong 2013, Chuong et al. 2013).

• In *trans*: Many TEs produce small RNAs [e.g., PIWI-interacting RNAs (piRNAs), endogenously produced small interfering RNAs, or microRNAs (miRNAs)], long noncoding RNAs (lncRNAs), or enhancer-overlapping RNAs : capable of altering transcription in *trans*.

• Host mechanisms have evolved to control spread of TEs & also contribute to altered host gene expression eg TE-targeting repressors trigger formation of heterochromatin, that can spread and be stably propagated. This may also have been the starting point for epigenetic phenomena such as Genomic Imprinting and X-chromosome inactivation…



Bejerano et al. 2006; Bo
2010; Lynch et al. 2011;
Martienssen 2007; Yoder

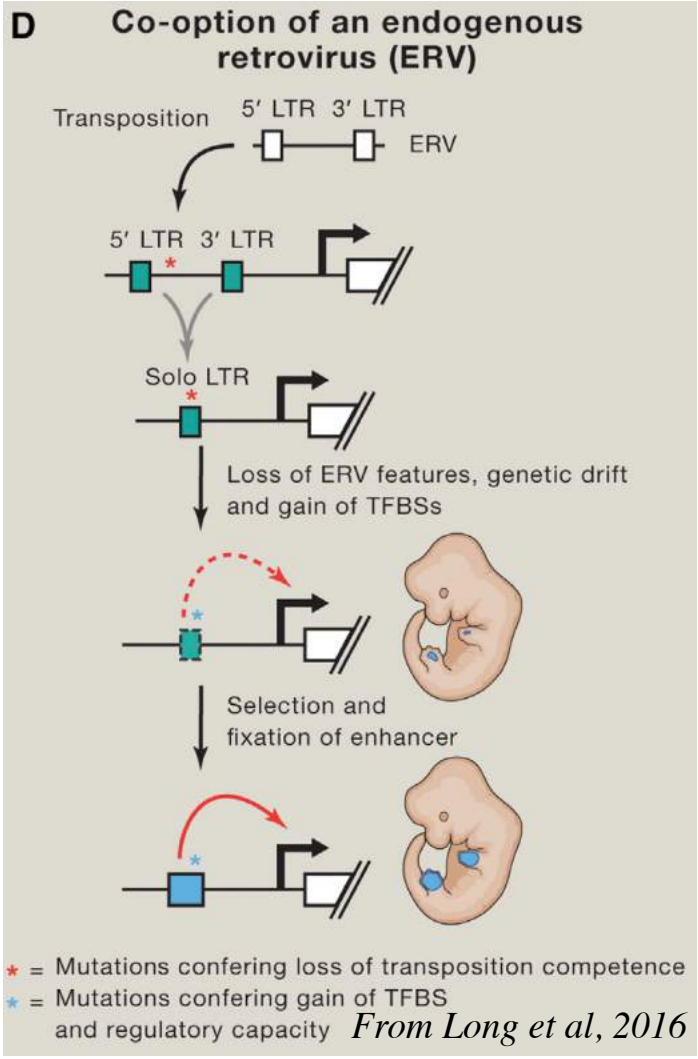
2003; Kunarso et al.
& Trono 2011, Slotkin &

TE Relics co-opted as Modulators of Gene Expression

When transcriptionally active, TEs not only produce transcripts, some of which can have long-range regulatory functions, but can also stimulate the expression of nearby genes through promoter or enhancer effects.

ERE-mediated, tissue-specific expression during early embryogenesis.

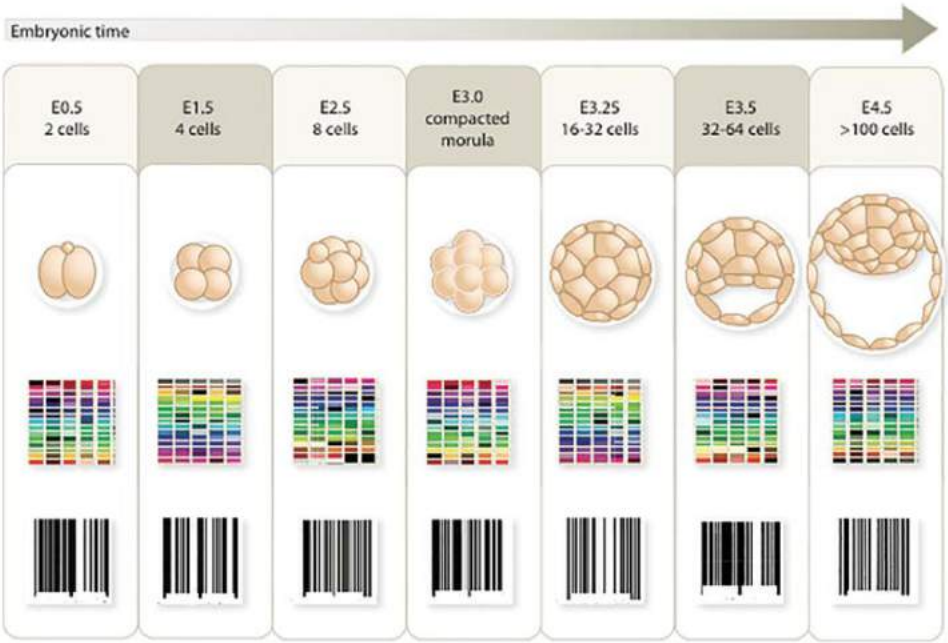
In human embryonic stem (ES) cells, 30% of transcripts are ERE-associated.
Fort et al. 2014, Lu et al. 2014, Santoni et al. 2012

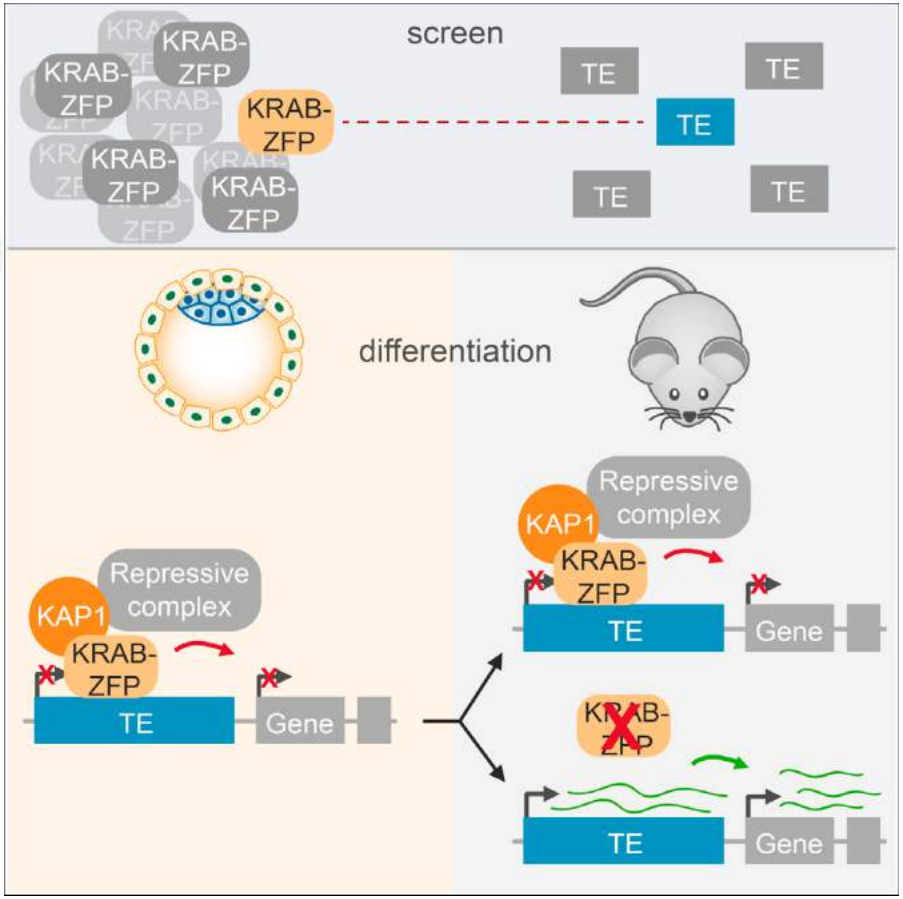


From Long et al, 2016

McClintock's "Controlling Elements" Today

KRAB-ZFPs and KAP1 are embryonic controllers of transposable elements (TEs) thought to irreversibly silence Tes. These modulators continue to control TE expression in adult tissues, where they also act to control expression of neighboring cellular genes.

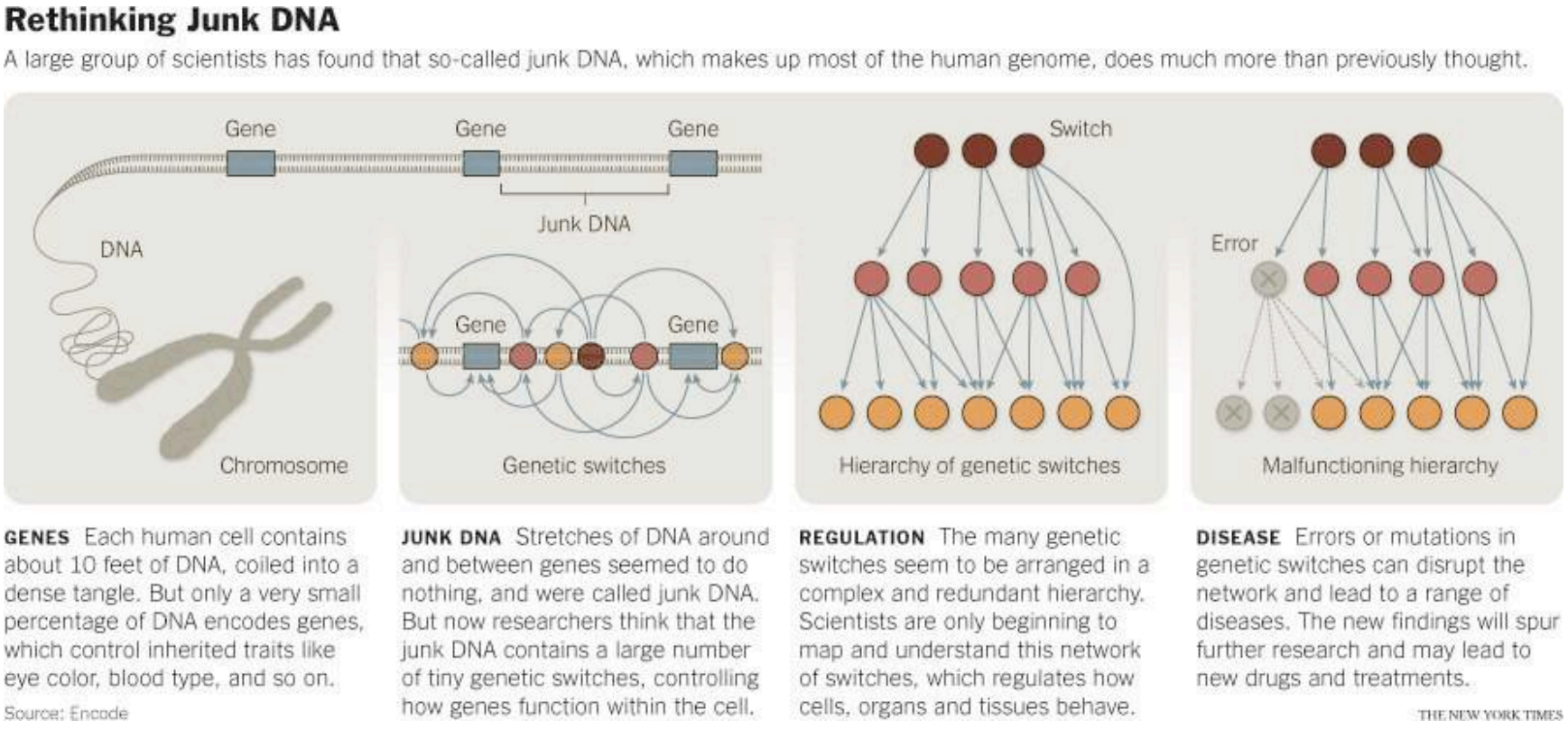




Ecco et al, Dev. Cell, 2016

Not "Junk" DNA: an Encyclopedia of Regulatory Elements

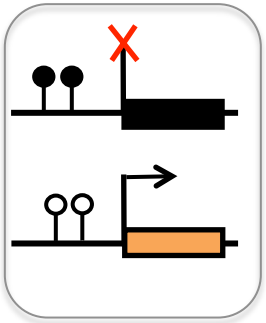
More than 80% of the human genome examined to date has a known biological function– *not* junk DNA

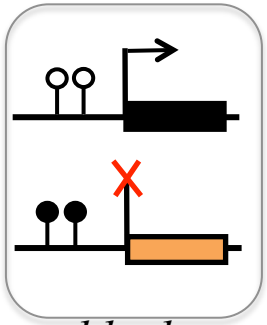


Barbara McClintock's visionary conclusions that mobile elements are the basis for controlling elements in development have finally accepted almost 70 years later

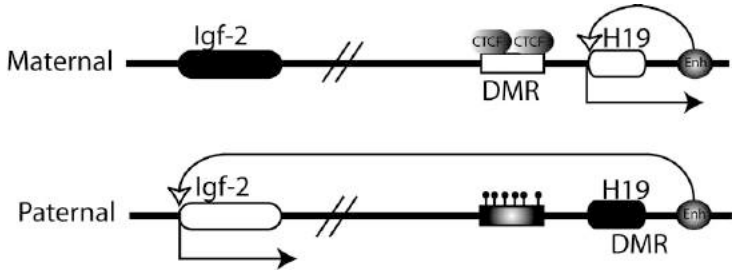
And TEs provide a Repertoire for Epigenetic Processes

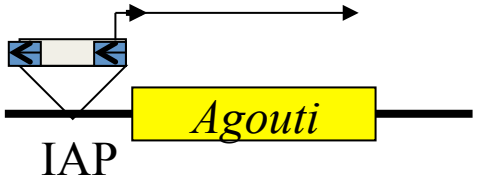
X inactivation





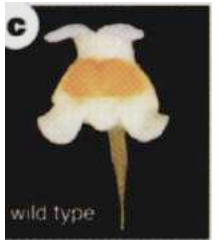
Imprinting

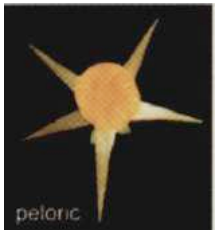


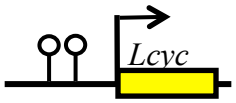


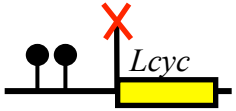
methylation

Morgan et al., 1999





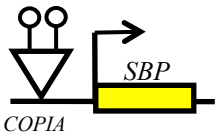


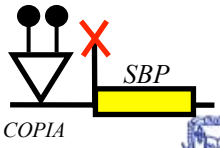


peloric

(Cubas et al, Nature, 1999)

cnr

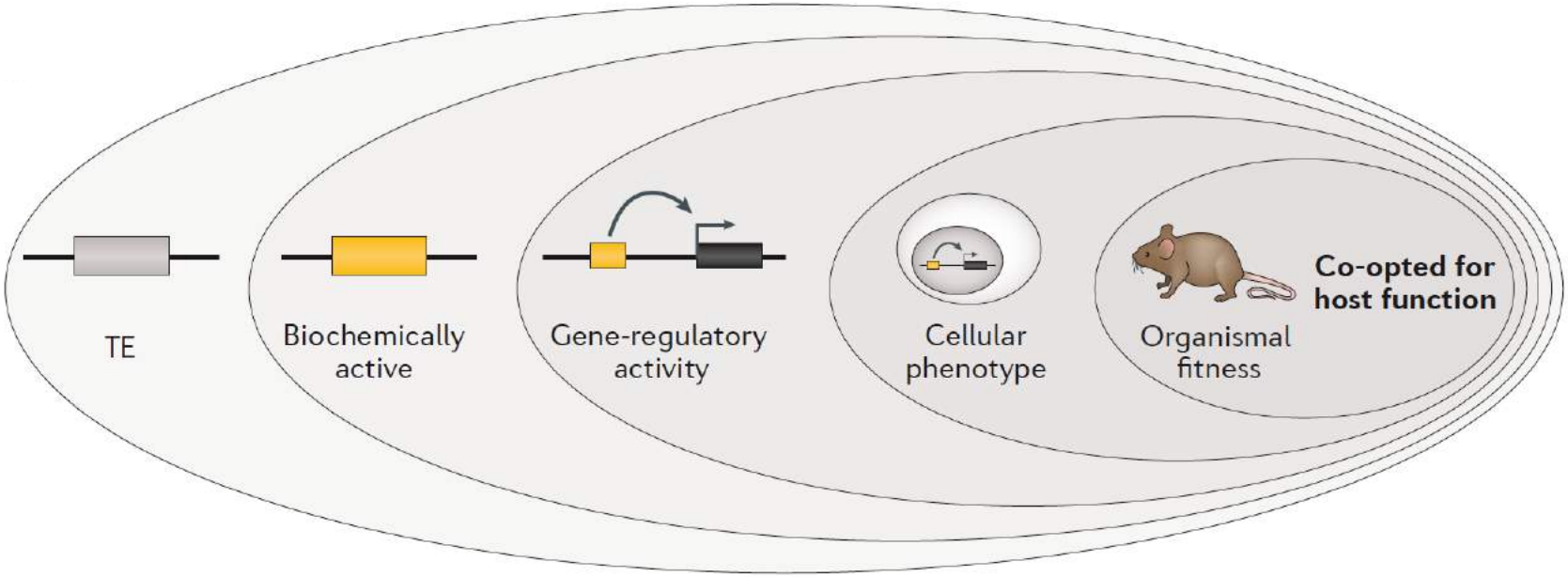




(Manning et al, Nat Genet, 2006)

Conclusions

From Dangerous mutagens to helpful Parasites
From Junk DNA to Regulator Networks
From Selfish to Altruistic DNA



CHAIRE ÉPIGÉNÉTIQUE ET MÉMOIRE CELLULAIRE

Année 2016-2017 :
"Épigénétique et ADN égoïste"

13 Février, 2017

Cours II

Le rôle de l'épigénétique dans la régulation des éléments transposables.

The role of epigenetics in the regulation of transposable elements