CENTER OF
PLANT SCIENCES

Sant'Anna
Scuola Universitaria Superiore Pisa

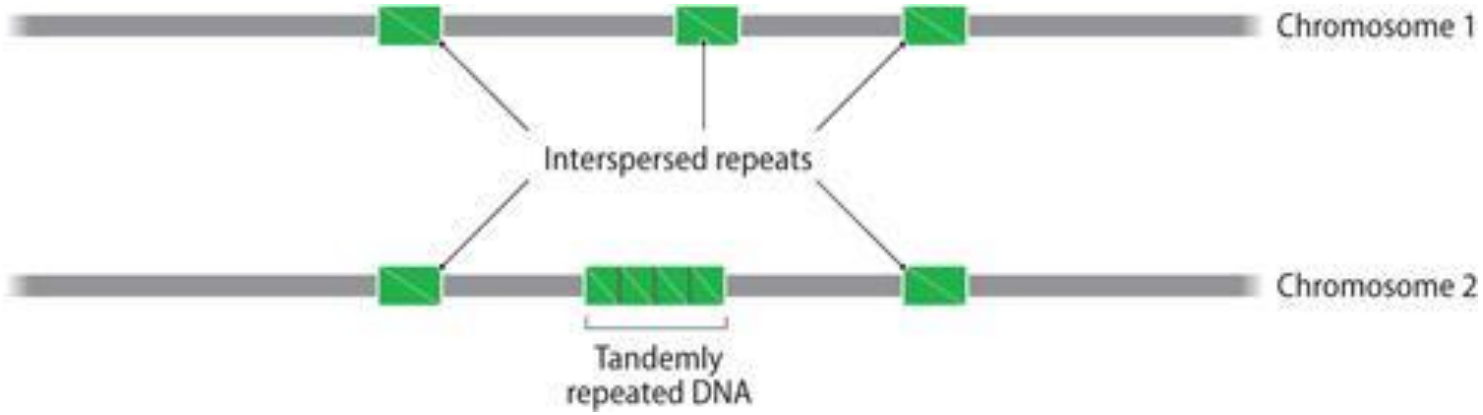
# Advanced Genomics

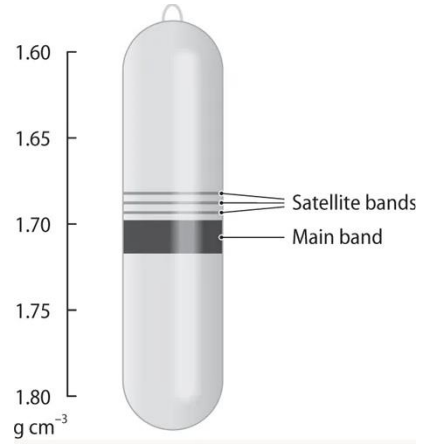
## Genome content -2

# Repetitive DNA

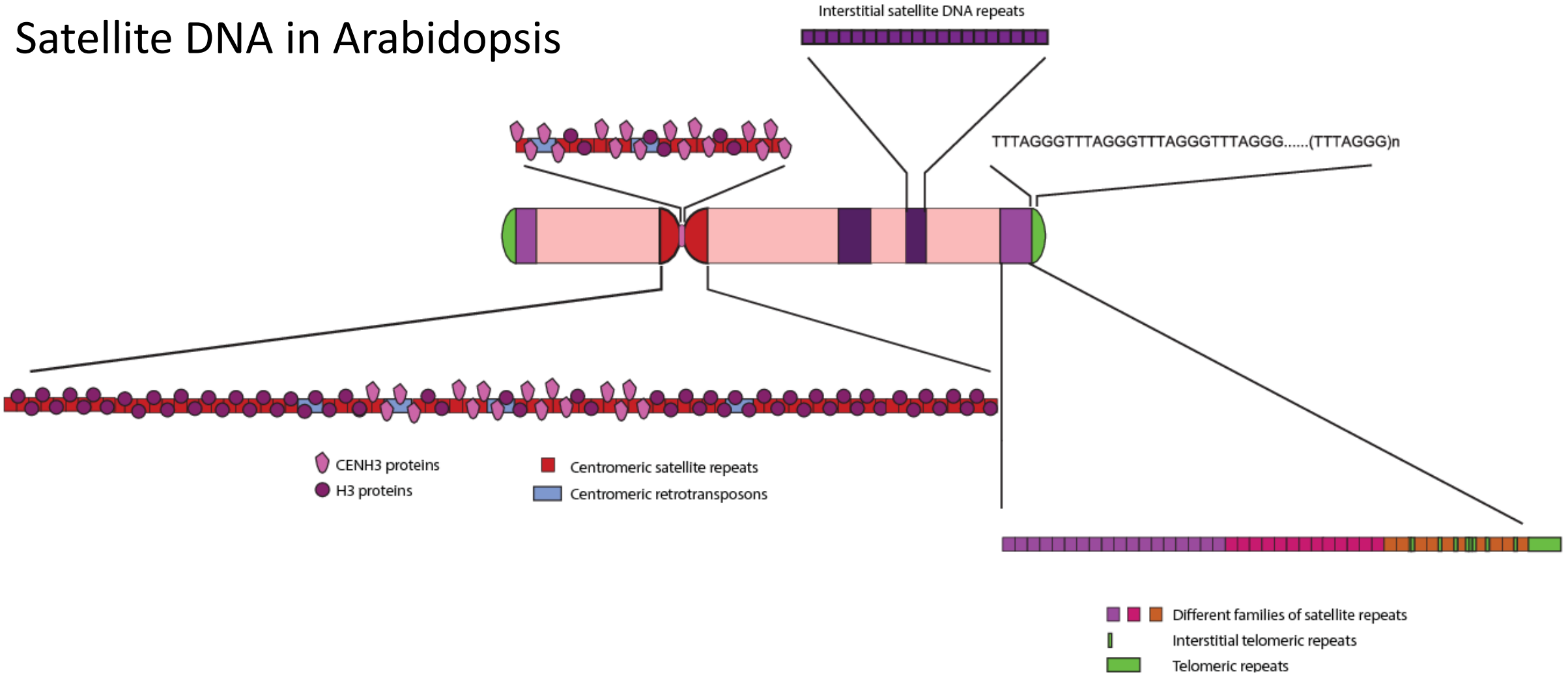
Two types:

- Interspersed repeats, with individual repeat units scattered across the genome
- Tandemly repeated DNA, with repeats near each other in an array
  - The tandem repeats are also called satellite DNA; probably originated from slippage in the polymerase.



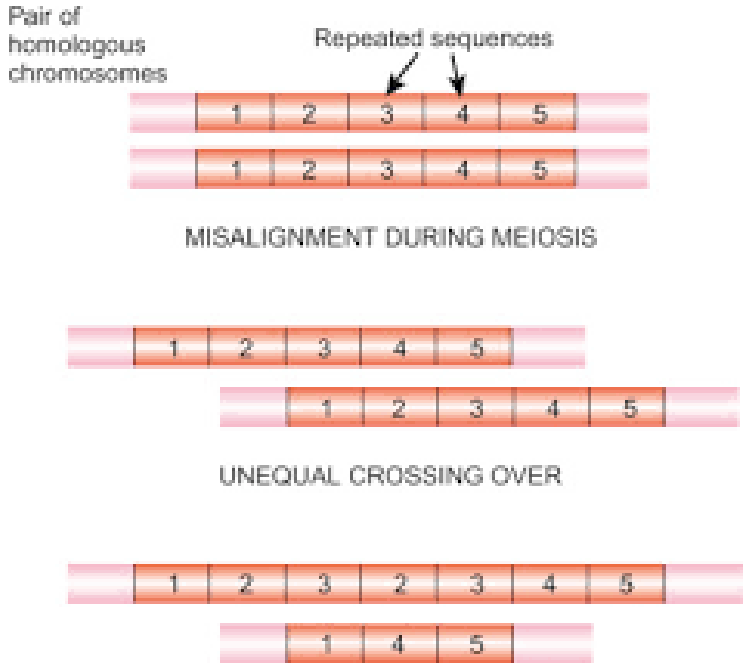



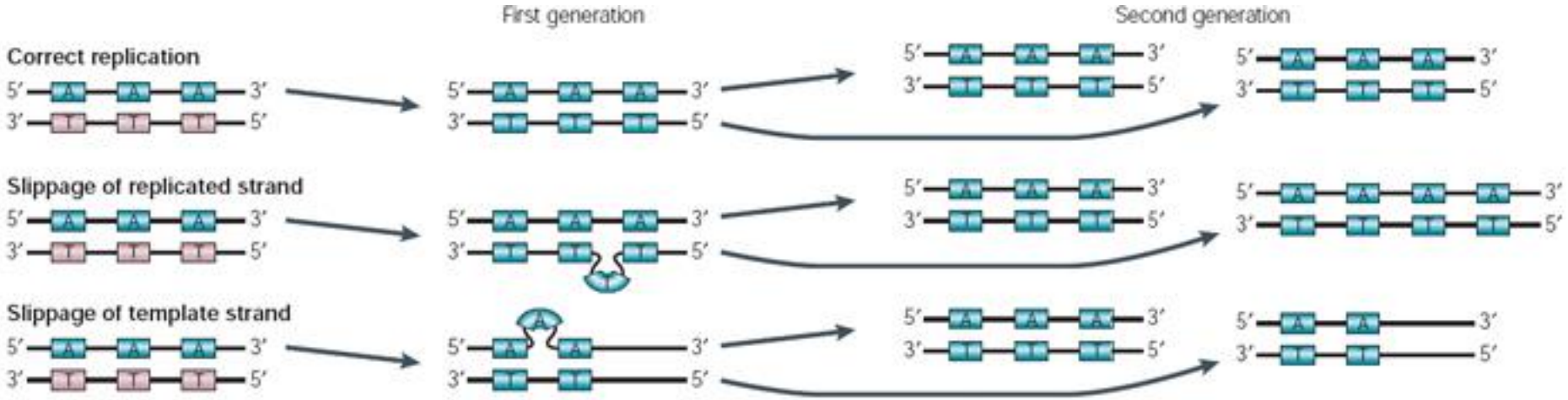
# Satellite DNA in Arabidopsis



Mechanisms for expansion and duplication:

- Unequal crossing-over
- Strand slippage

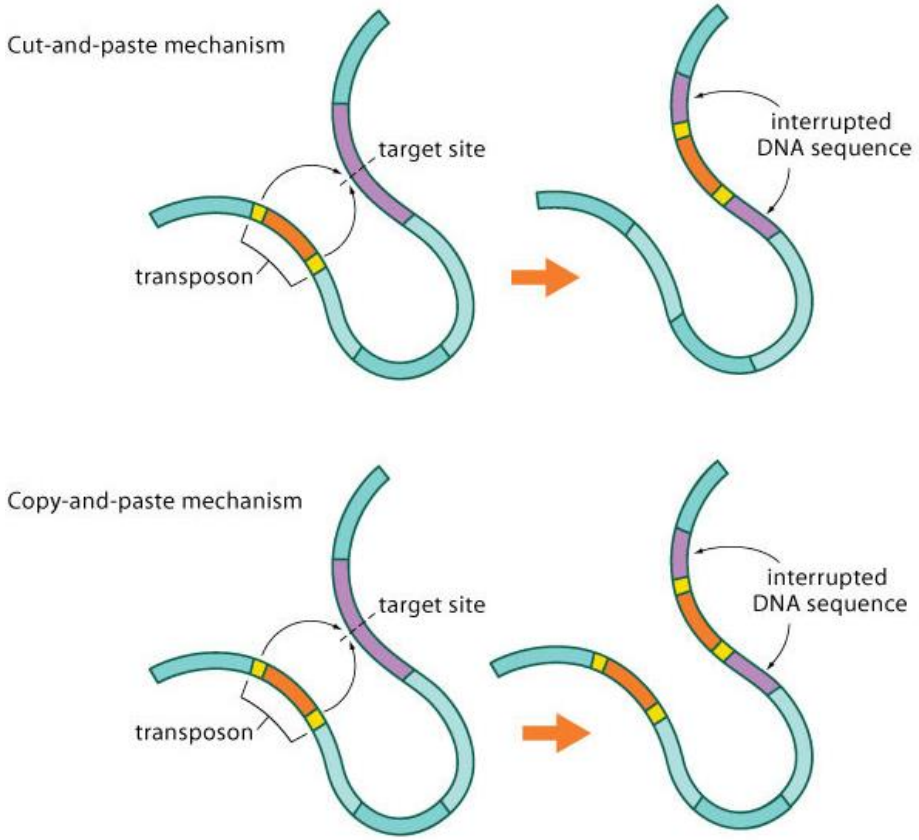


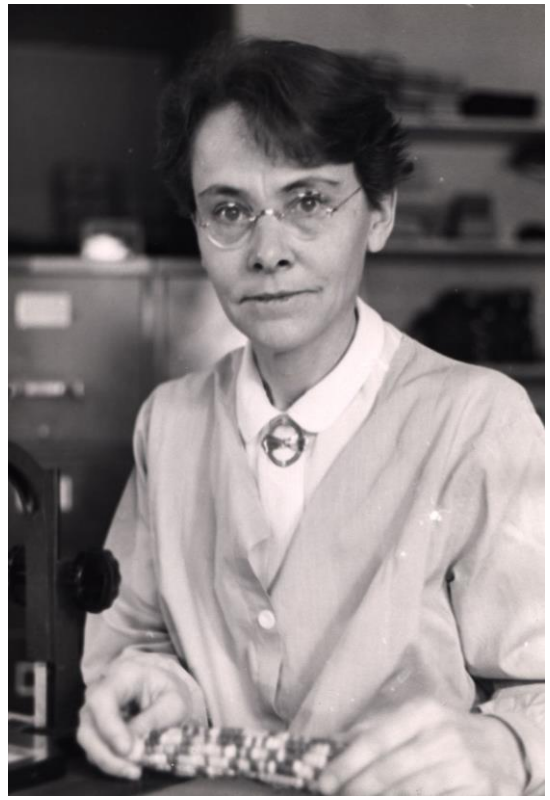



# Transposable elements

Sequences of DNA that move (or jump) from one location in the genome to another

- Infamously known as junk DNA (Ohno , 1972), selfish DNA (Dawkins , 1976), parasitic DNA (Orgel and Crick, 1980

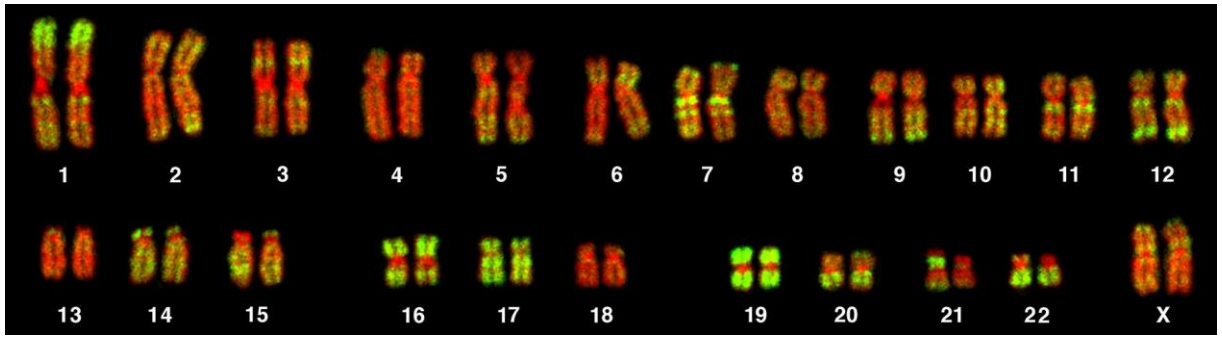


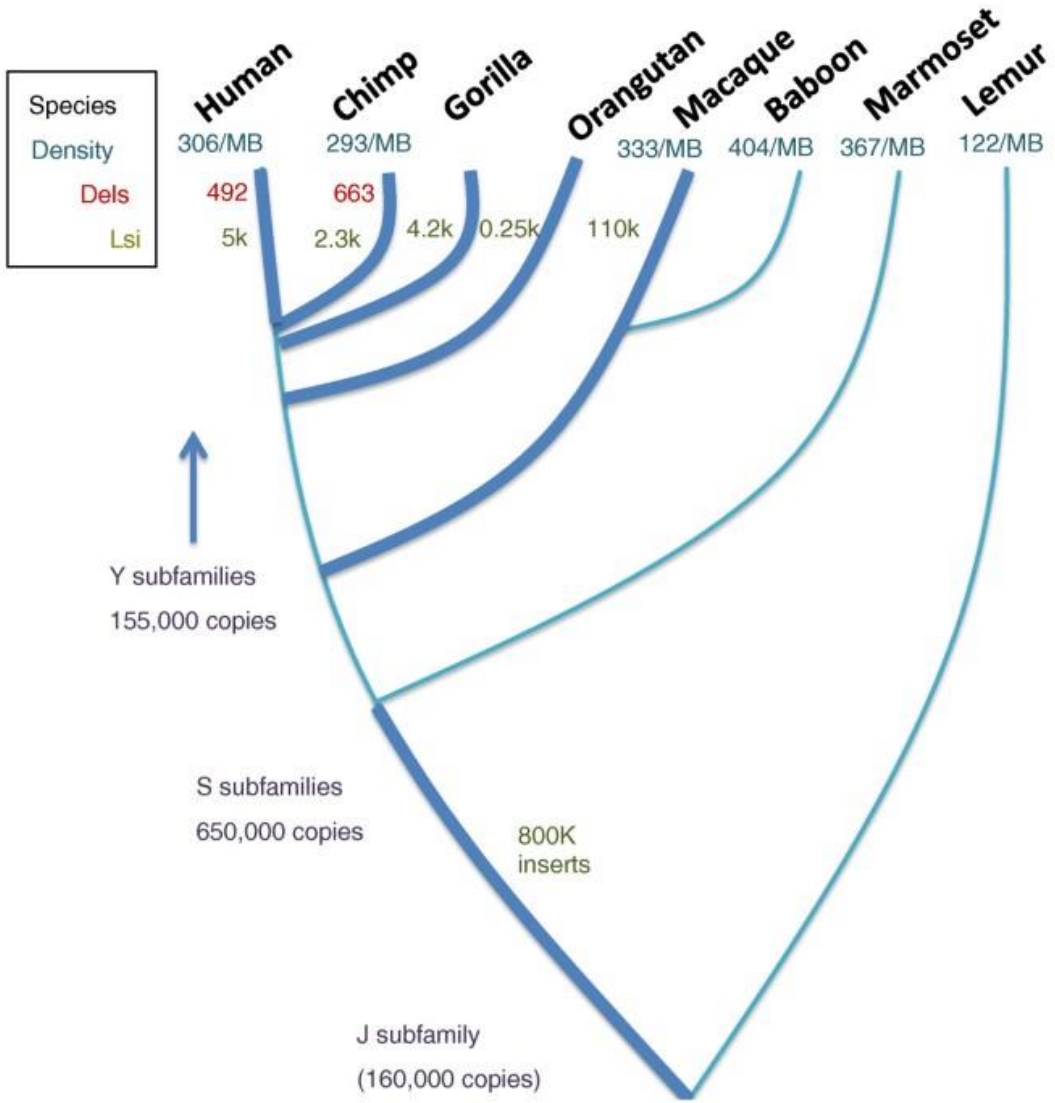



Barbara McClintock, Nobel Prize for Physiology and Medicine 1983

The *Alu* element

- is 300 bp long
- Belongs to SINE (short interspersed nuclear elements)
- about 11% of the human genome
- About 1 million copies



# INDUCTION OF INSTABILITY AT SELECTED LOCI IN MAIZE

BARBARA McCLINTOCK
*Department of Genetics, Carnegie Institution of Washington, Cold Spring Harbor, N. Y.*

Received April 14, 1953

IN previous reports (McCLINTOCK 1950, 1951), studies of the origin and expression of genic instability at a number of known loci in the maize chromosomes were summarized. It was concluded that changes in genic expression result from chromosome alterations at the locus of a gene and these are initiated by units other than those composing the gene itself. The mutations are considered, therefore, not as changes in the potentials of genic action, but rather chromosomal modifications at the locus effecting the kind and the degree of genic expression. The extragenic chromatin units have specificity in that differences among them may be recognized. Each exerts a specific type of control of the action of the gene with which it becomes associated. These units may be transposed from one location to another within the chromosome complement. When incorporated at a new location, each expresses its mode of control of the action of the associated gene, and in a manner similar to that which occurred at the former location. These conclusions have been supported by extensive examination of the action of one particular system that has modified genic action at a number of different loci. It is the so-called Dissociation-Activator (*Ds-Ac*) two-unit system.

Unnumbered figure pg 489a
*Introduction to Genetic Analysis, Ninth Edition*
© 2008 W.H. Freeman and Company

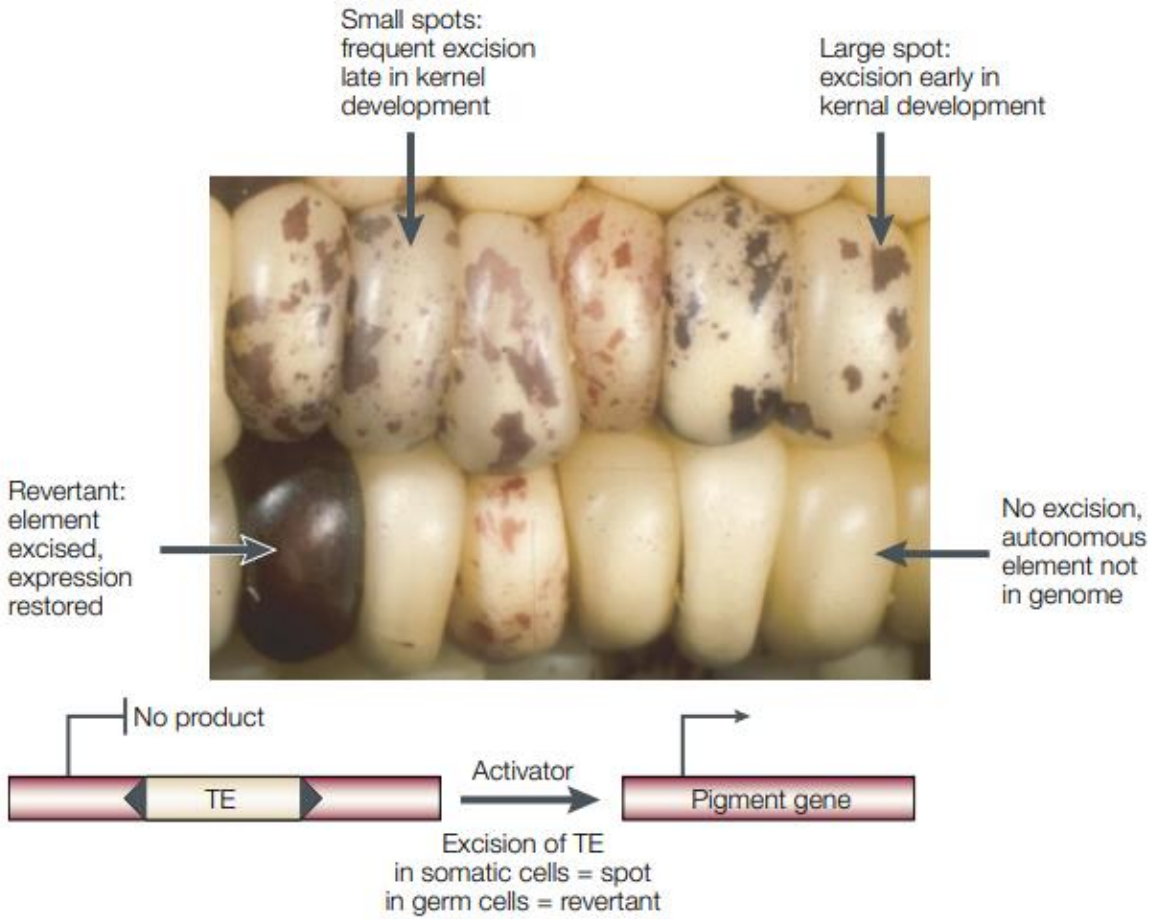
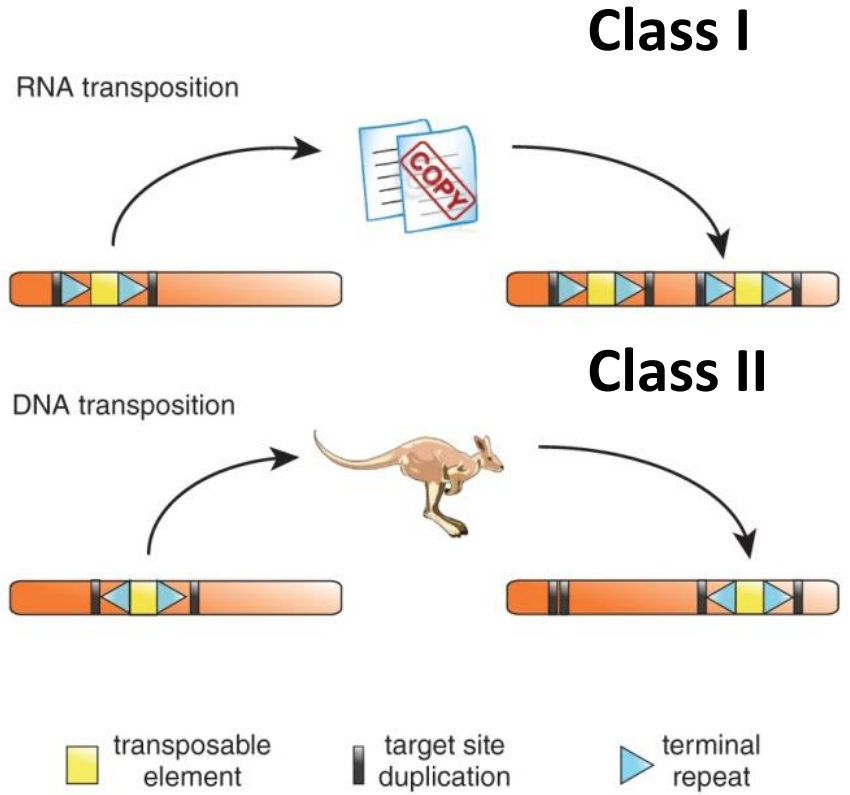



Figure 1 | **Using kernel phenotypes to study transposon behaviour.** Kernels on a maize ear show unstable phenotypes due to the interplay between a transposable element (TE) and a gene that encodes an enzyme in the anthocyanin (pigment) biosynthetic pathway. Sectors of revertant (pigmented) aleurone tissue result from the excision of the TE in a single cell. The size of the sector reflects the time in kernel development at which excision occurred. An understanding of the genetic basis of this and similar mutant phenotypes led to the discovery of TEs and to an amazingly detailed description of the behaviour of what we now call class 2 (DNA) elements

Freshotte et al 2002

There's three broad groups of transposable elements

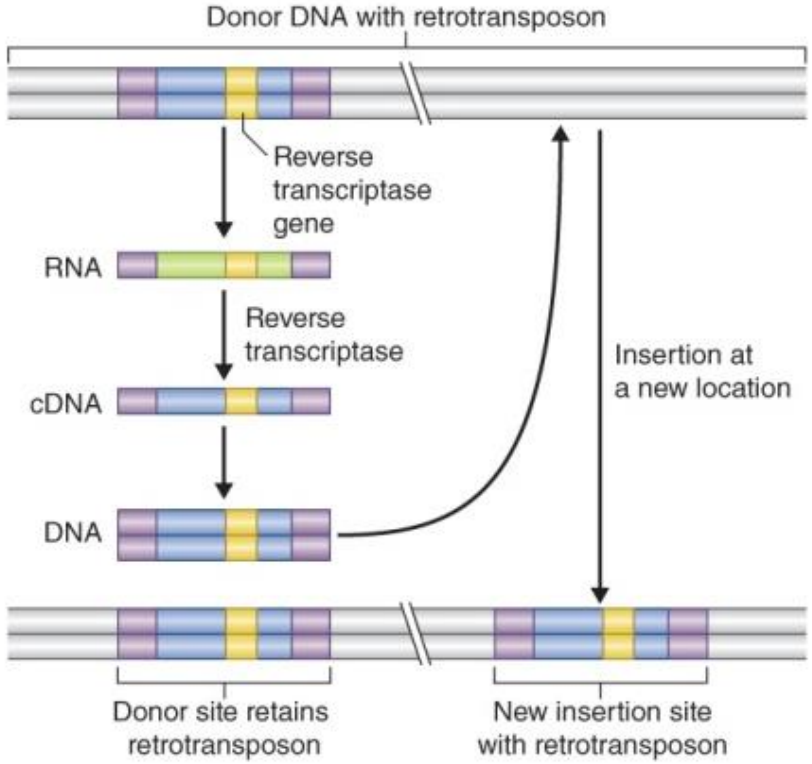
- **Retrotransposons (class I):** (aka RNA elements): they copy themselves in RNA, then back to DNA, then are inserted back in the genome
- **DNA elements (class II)**: they are excised from the DNA and reintegrated in another region of the genome
- **Helitrons (class II)**: eukaryotic rolling-circle transposable elements




Can be **autonomous** or **non-autonomous**, depending on whether they are able to cause their own movement or not (non-autonomous having lost their capacity through evolution)

## Retrotransposons

- Replicate with a copy-paste mechanism; are the main responsible for genome size expansion
- Structurally similar (=related) to retrovirus. The main difference being that they cannot leave the cell.
- Fundamental content of a retrotransposon:
  - Reverse transcriptase (RT) gene to go from RNA to DNA
  - RNase H gene needed for replication
  - *gag* protein gene(s)

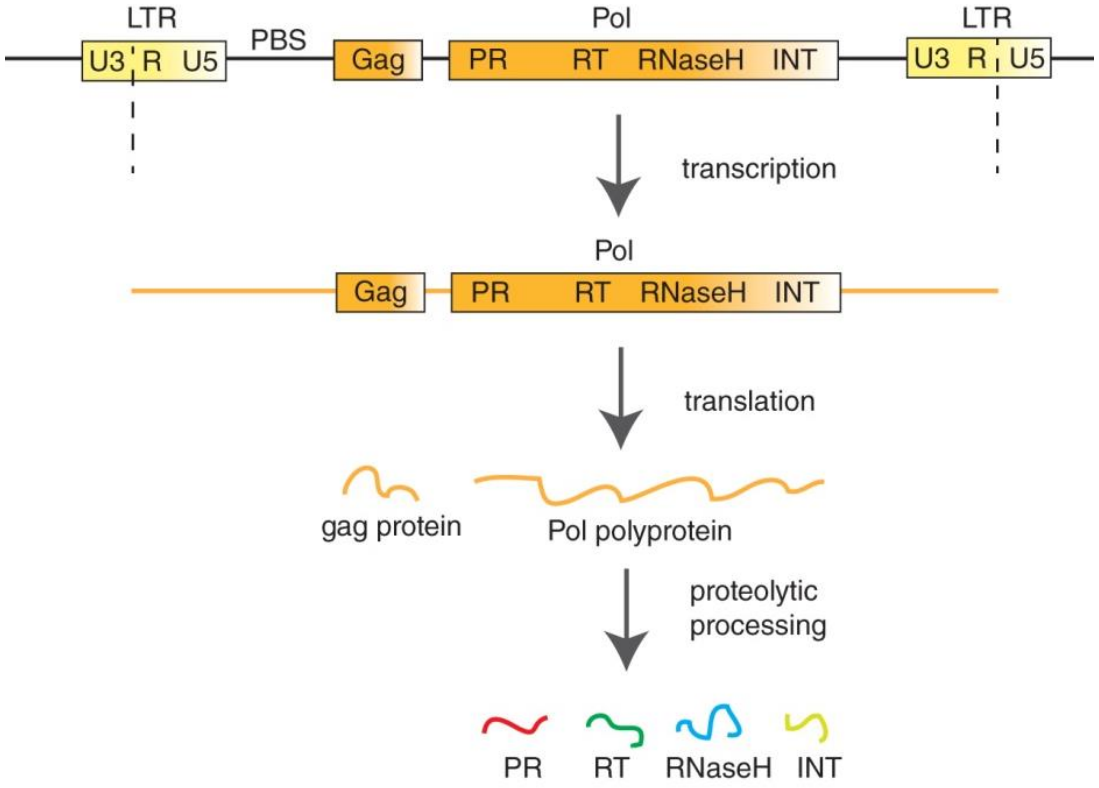



Depending on end repeats, can be either of two types:

- Long-terminal-repeats (LTR) retrotransposons
- non-LTR retrotransposons

## LTR retrotransposons

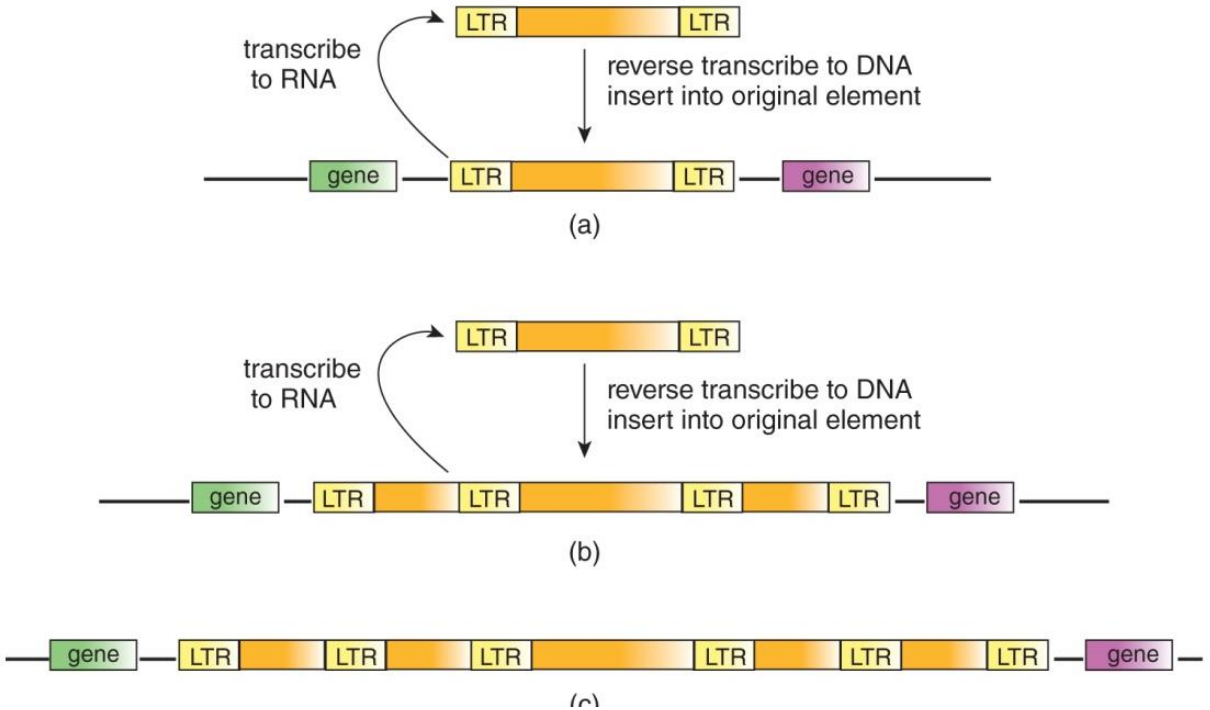
- More common in plants than in animals
- Each organized in a single transcription unit, with transcription starting downstream the 5’ end of the first LTR and ending upstream the 3’ end of the second LTR
- The transcription unit contains: gag, RT, Rnase H, an integrase (INT), a protease (PR) to cleave the precursor protein in functional units
- DNA to RNA happens in the nucleus; RNA to DNA in the cytoplasm; the cDNA gets back in the nucleus to be reintegrated by INT
- LTRs contain regions U3, R, U5 which are necessary for transcription and reverse transcription

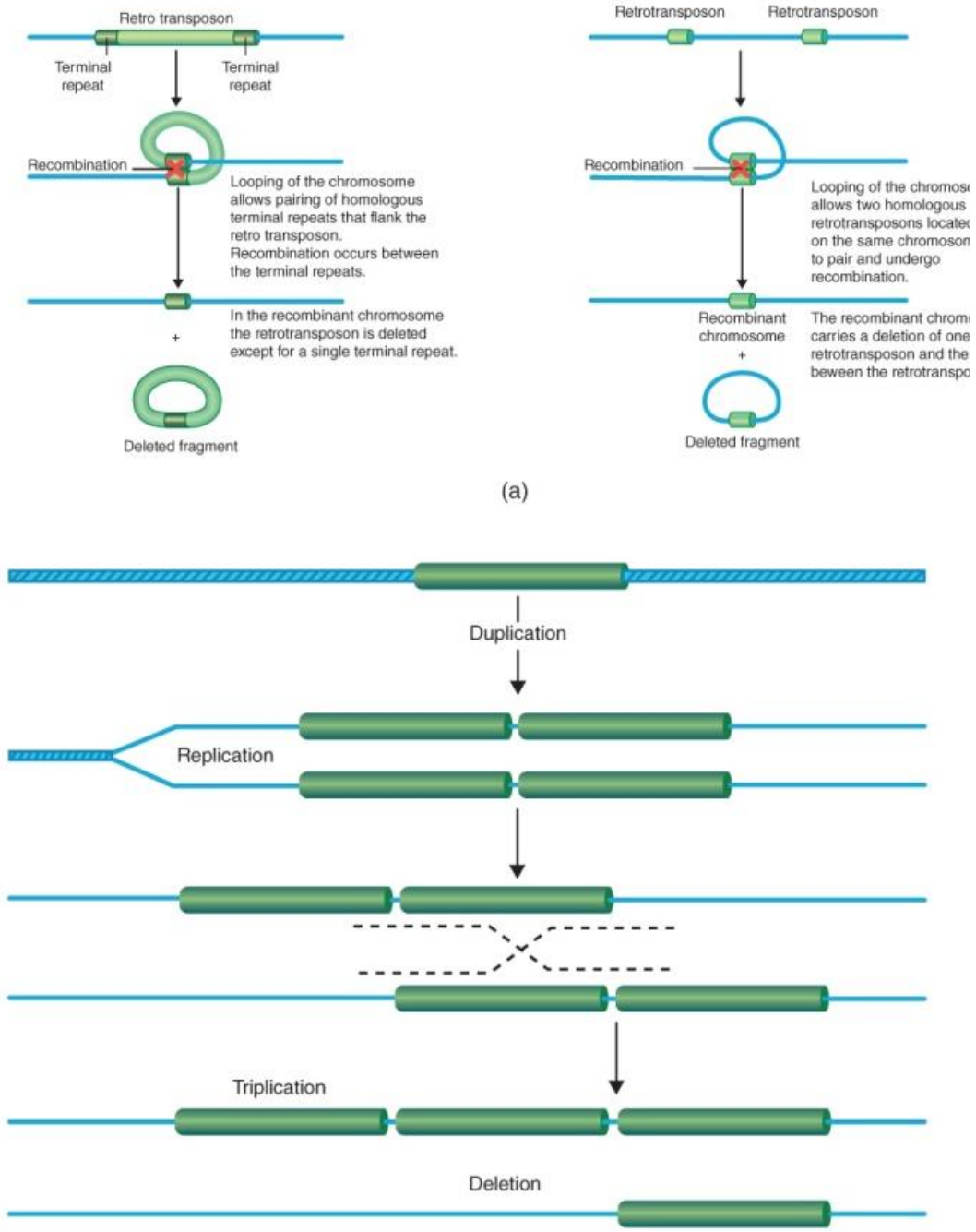



LTR tend to be integrated within older LTR (due to sequence complementarity)

- This generates tandem repeats

# LTR can both expand and contract genome size



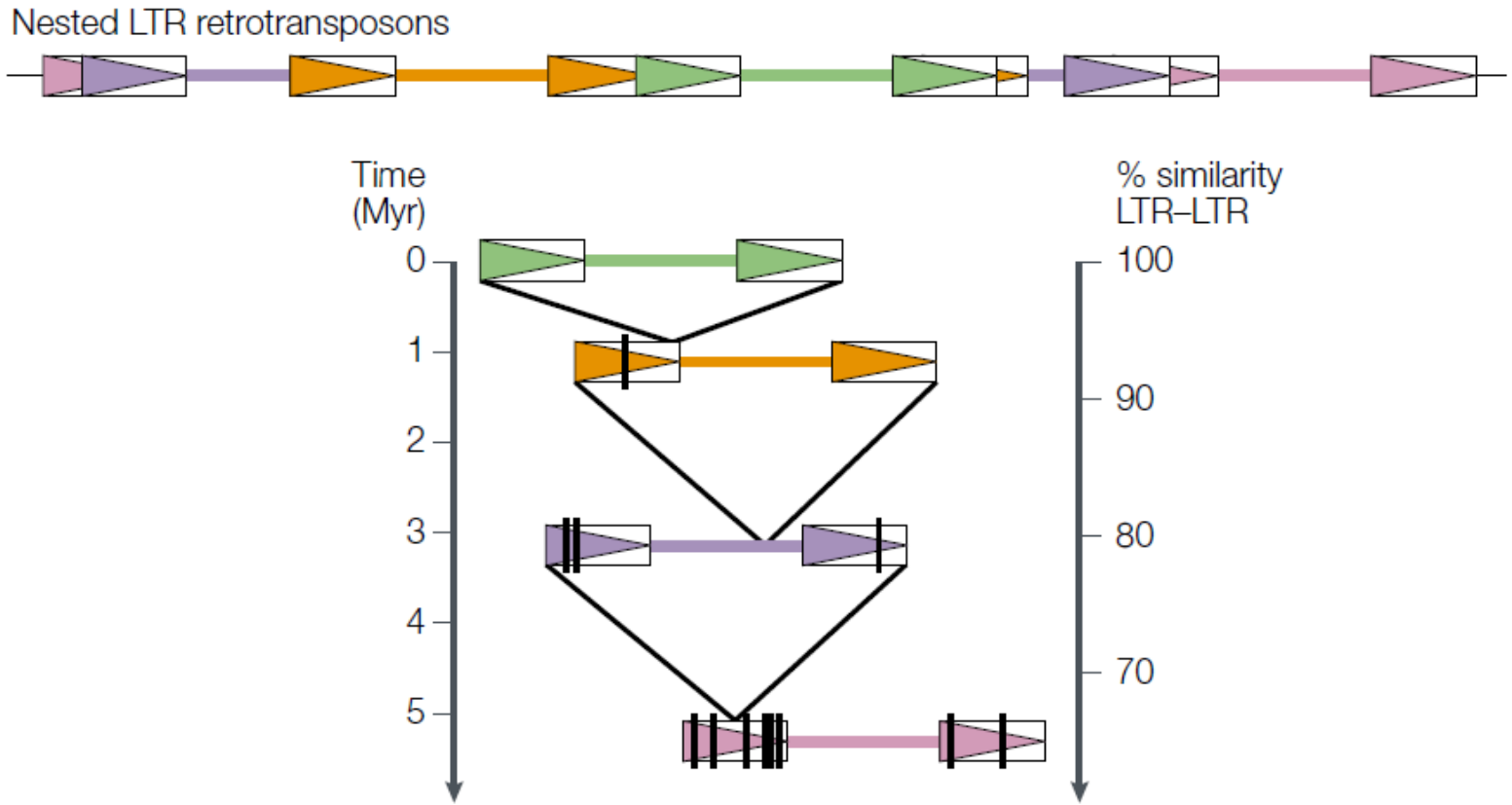



Figure 3 | **Estimating the time of retrotransposon insertion.** At the time of insertion, the long terminal repeats (LTRs) of an element are identical because both are copied from the same template during cDNA synthesis. As time passes, nucleotide changes accumulate in each LTR (represented by vertical bars in the LTRs). If the average rate of nucleotide substitution per year is known for the host organism, then sequence divergence between the LTRs provides an estimate of when insertion occurred. This method has been applied to date the insertions of LTR retrotransposons nested in the intergenic regions that surround the maize *alcohol dehydrogenase1* (*adh1*) gene[56]. See text for details. Myr, million years.

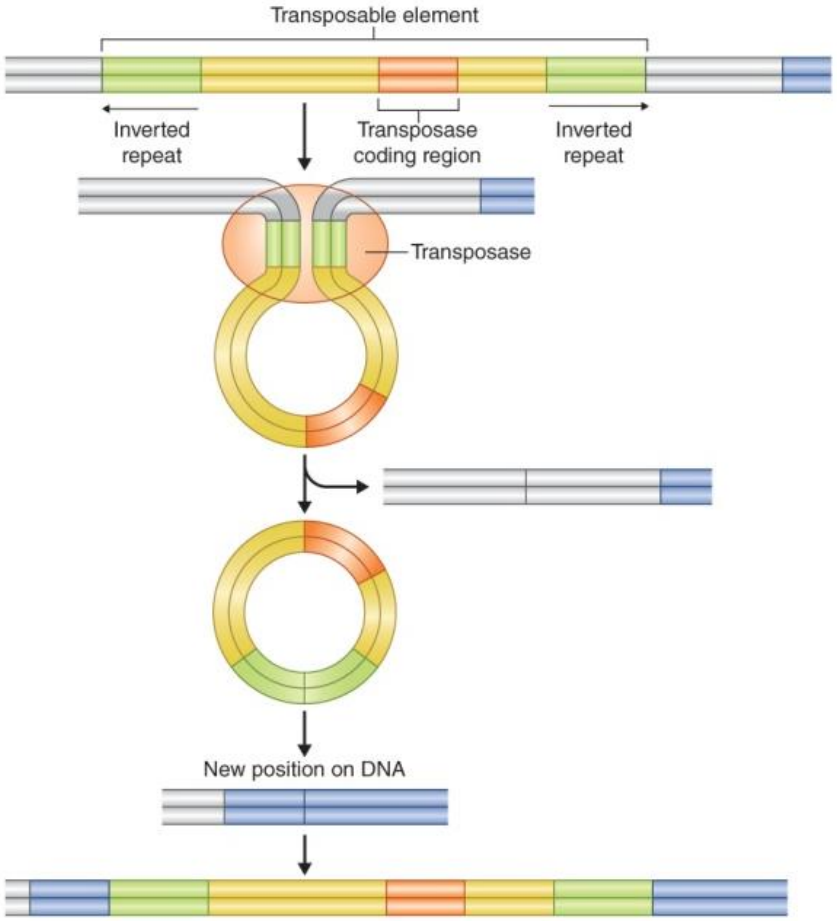
## non-LTR retrotransposons

Lack LTRs; common in vertebrates; two types

- Long interspersed nuclear elements (LINEs)
  - Contains a 5’ UTR followed by two ORFs: ORF1, ORF2 (encoding a RT)
  - Encode for an exonuclease that causes a nick in the target DNA to enable reinsertion
  - The only TE active in humans is a LINE, L-1
  - Most of LINE sequences in plant and animal genomes are truncated at the 5’ end, rendering them inactive
- Short interspersed nuclear elements (SINEs), < 500bp
  - the non-autonomous version of LINEs, as they lack a RT
  - Very variable in sequence, hard to detect
  - Similar to tRNA (are transcribed by RNA pol III)

## DNA elements

- Replicate with a cut-paste mechanism; don't affect genomes size
- All contain a transposase flanked by terminal inverted repeats (TIR) → different from LTRs, which are not inverted
- Once they are excised, they leave a target-site duplication of two or more bps (with a sequence specific of each RT)
- Target-site duplications promote mutation and have the potentiality to alter gene structure
- DNA elements are in all eukaryotes, hence very ancient

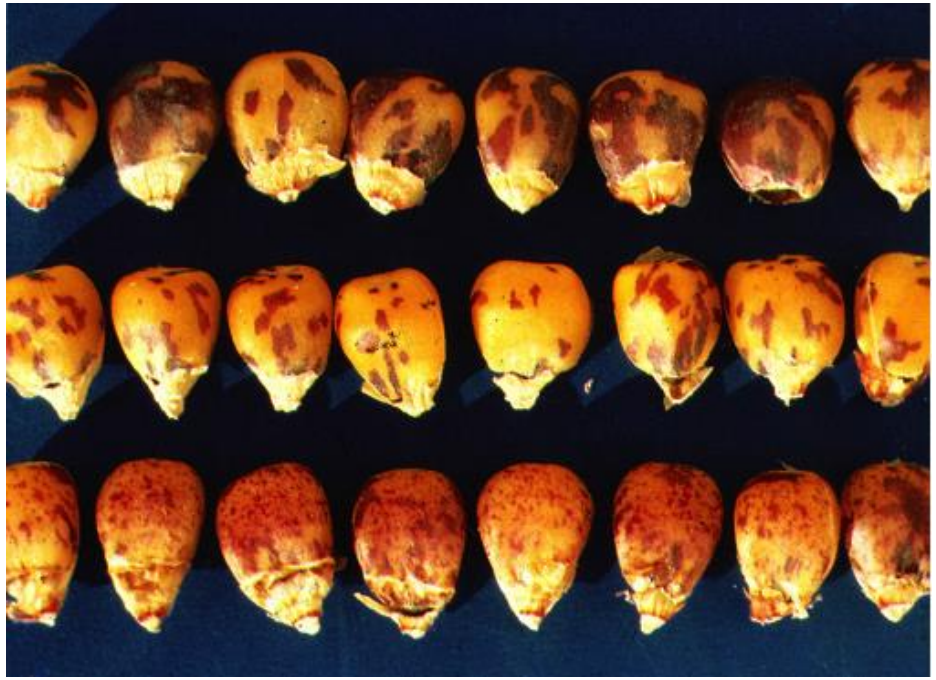


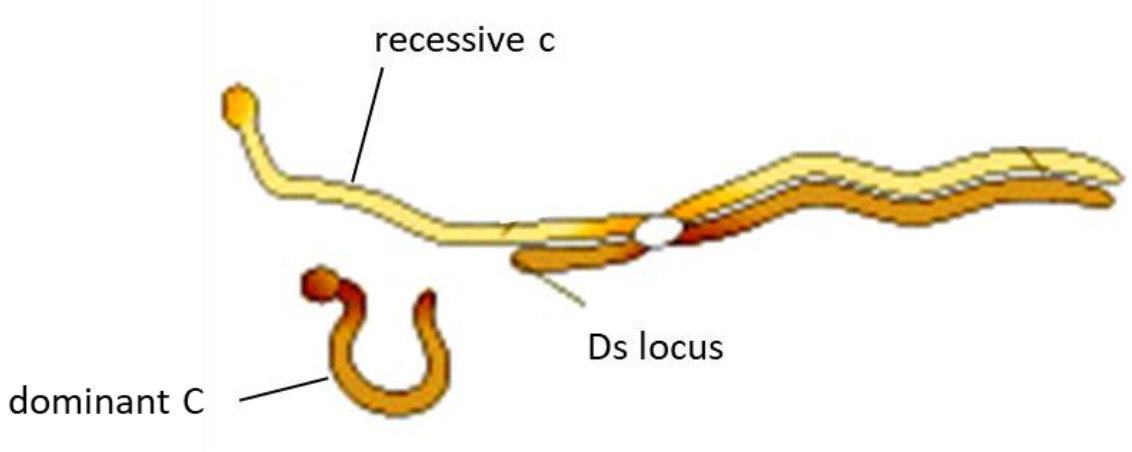

Mostly categorized according to sequence homology in the transposase: *Tc1/mariner*, *Pif/Harbinger*, *hAT*, *MULEs, …*

# DNA elements and McClintock

- McClintock was studying chromosome breackage and found a group of maize genotypes in which part of a particular chromosome consistently dissociated fron the rest of the chrosomome
- She concluded that dissociation was casued by a locus that she called ***Ds*** (**dissociator**)
- Analyzing several maize stocks she found that *Ds* was necessary but not sufficient for dissociation, and it needed another locus that she called ***Ac*** (**activator**)
- She tried to map *Ds* and *Ac* (more later on genetic maps), and she found that *Ds* and *Ac* could move around the genome!
  - Some loci can move around
  - Some could move on their own (the autonomous *Ac*), some could not (the non-autonomous *Ds*, which is derived from *Ac*)

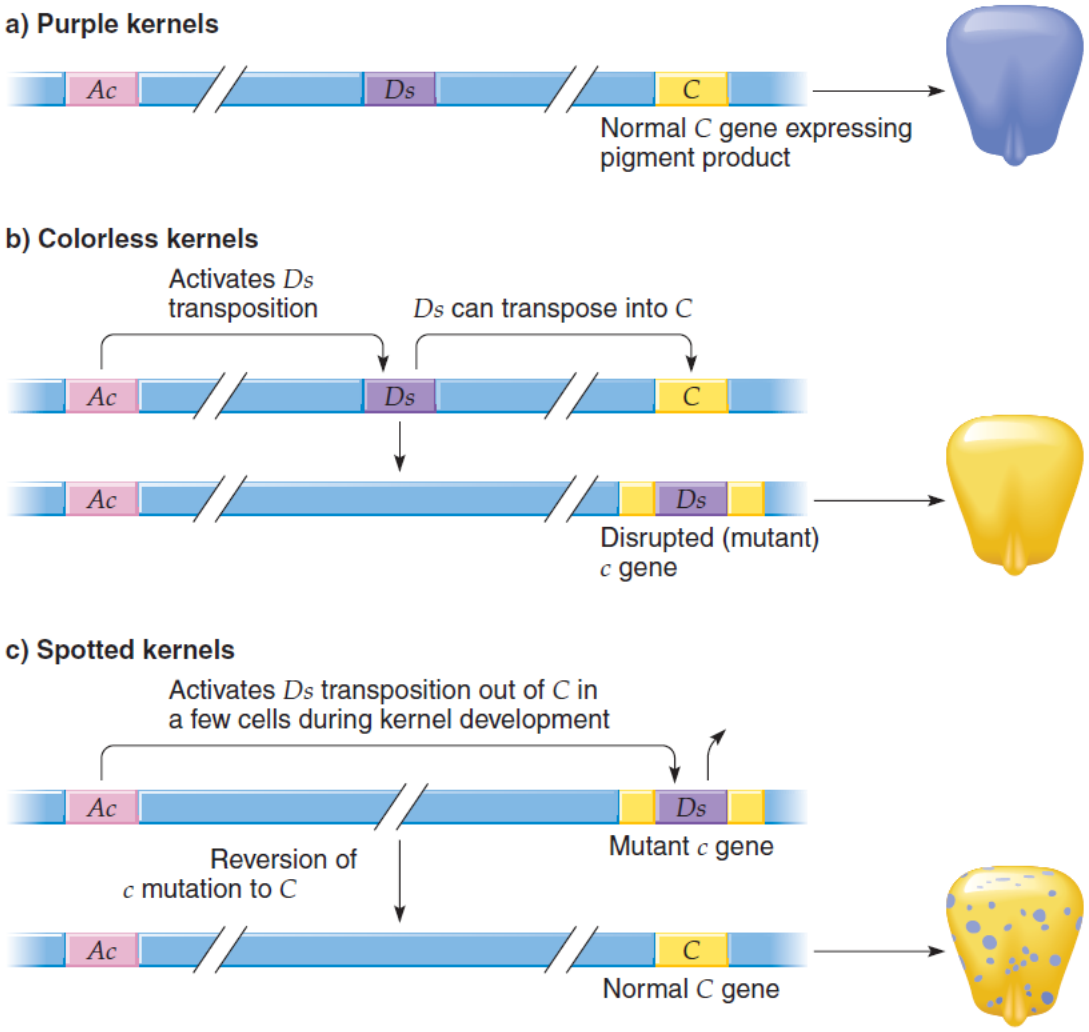
Focus on the C locus on maize chromosome 9, controlling purple/yellow phentoype

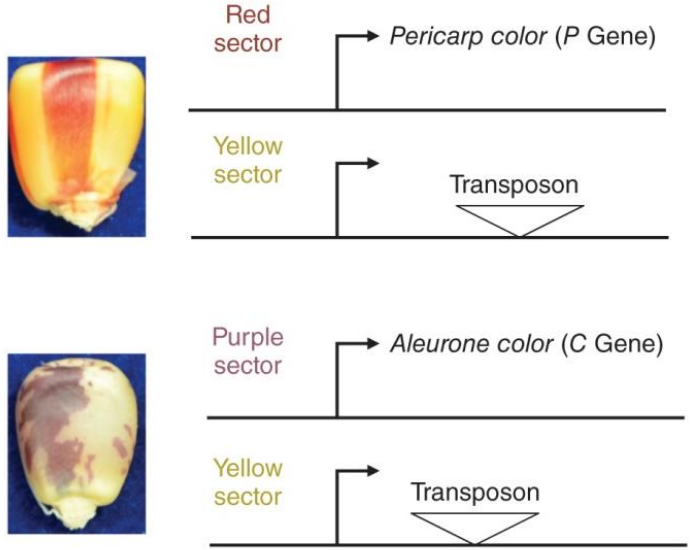





*Ds* may cause breakage of the chromosome (can be seen on the the karyotype) but only if *Ac* is present

- Loss of the dominant allele shows a phenotype
- But the phenotype can be reverted (from c to C)! In this case the culprit is not a chromosome breakage, but rather insertion/excision of *Ds* in the C locus, with a phenotype that is dose dependent of *Ac*

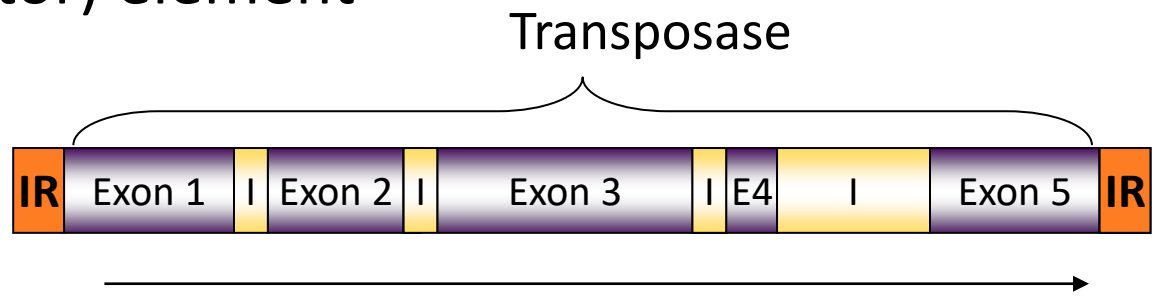


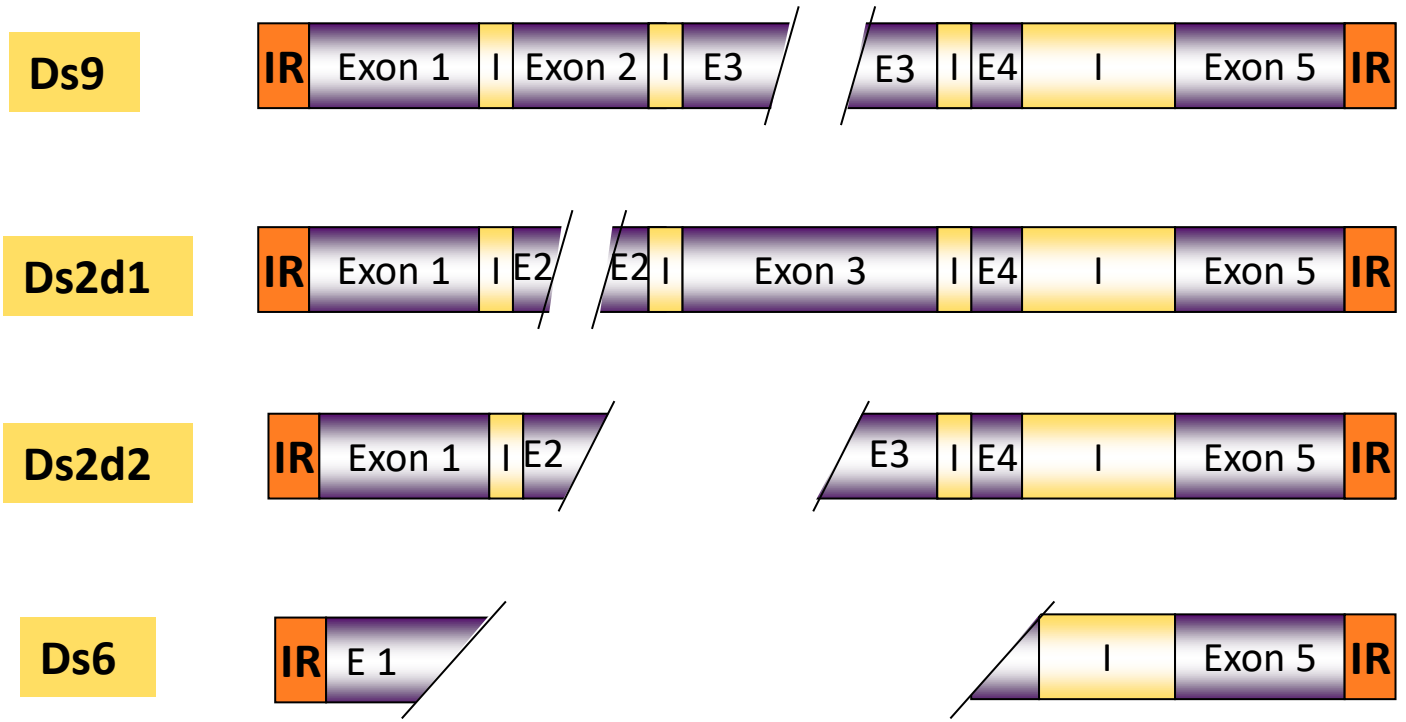

"If you know you are on the right track, if you have this inner knowledge, then nobody can turn you off... no matter what they say.."
–Barbara McClintock

- *Ac* and *Ds* are TEs with IRs that may move around in the genome
- They are similar in sequence, although *Ds* is smaller than *Ac*
- *Ac* is a complete TE which can encode for a transposase which is necessary for transposition
- *Ds* cannot produce the transposase by itself, but has IRs that enable transposition
- When *Ac* is present, its transposase can move *Ds* and have it inactivate genes

## *Ac* (Activator) element




## *Ds* (Dissociator) element(s)

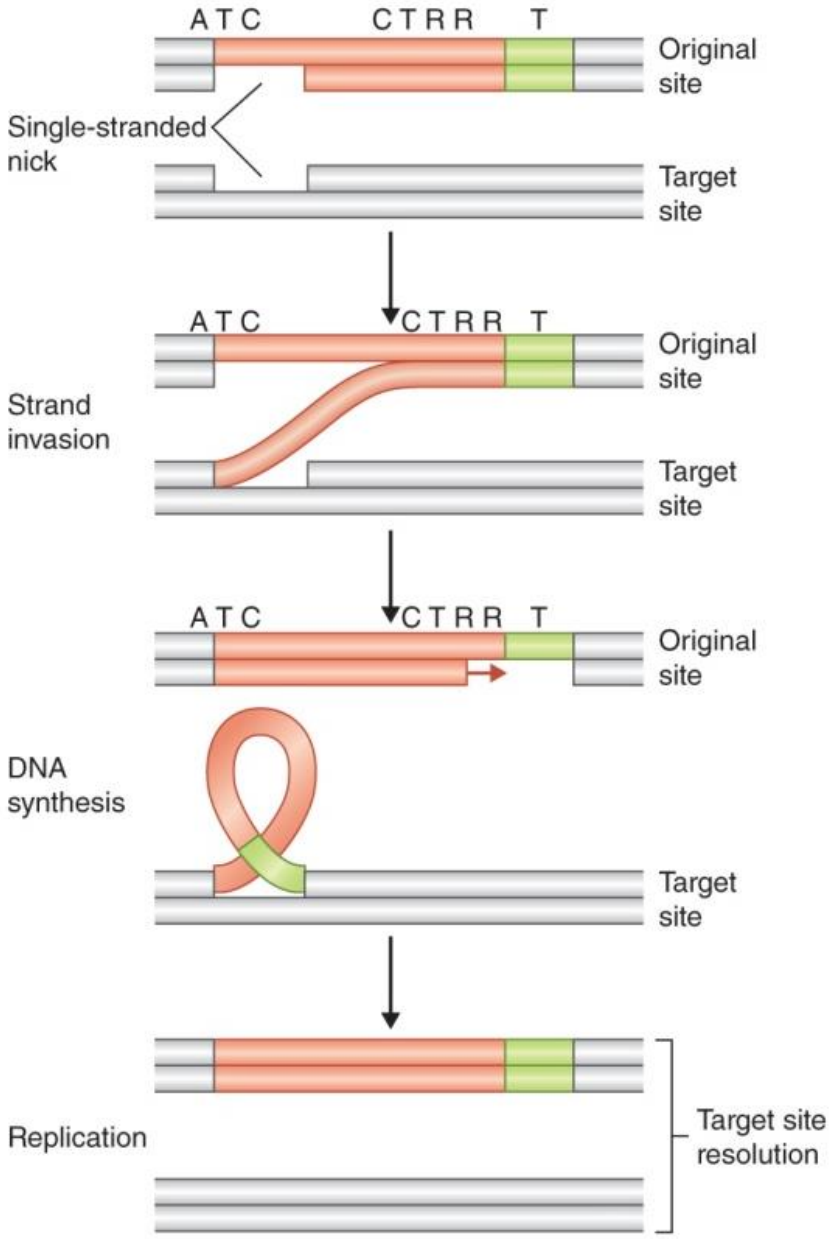


## Types of DNA elements

- hAT elements. McClintock's ones
  - Short IRs, about 12 bp
  - Mutations in IRs abolish translocation
  - Binding of the transposase to the DNA typically cannot occur if the DNA is methylated
  - Insertion happens in linked regions, tipically a few Kb apart
- Mutator (Mu) elements / MULEs
  - More copies than hATs
  - Discovered in maize genomes where they can cause a mutation rate 30x the normal
  - Can insert in sites far away; leave large indels after excision
  - IRs are about 200bp
- CACTA elements
  - IRs 13 bp, characterized by the CACTA sequence
  - Encode different proteins through alternative splicing
- Miniature inverted repeat transposable elements (MITEs)
  - 100Ks in any genome
  - Tipically <600 bp, with IRs
  - Don't encode proteins; seem to be related (by sequence homology) to other class of TEs, such as *Tc1/mariner* and *Pif/Harbinger*

## Helitrons

- Recently described, the first being identified by bioinformatics only (not yet tested!)
- Genomic copy numbers are highly variable, even among closely related species
- Structurally asymmetric, replicated with a rolling circle mechanism
- Encode a rolling-circle replication initiator (Rep) and DNA helicase (Hel) domains. The Rep/Helicase protein includes zinc finger motifs
- The only class of eukaryotic DNA transposons that do not generate duplications of target sites during transposition



# There's a whole lot of different TE families and a specific nomenclature that considers sequence similarity, gene order (and generally evolutionary origin)

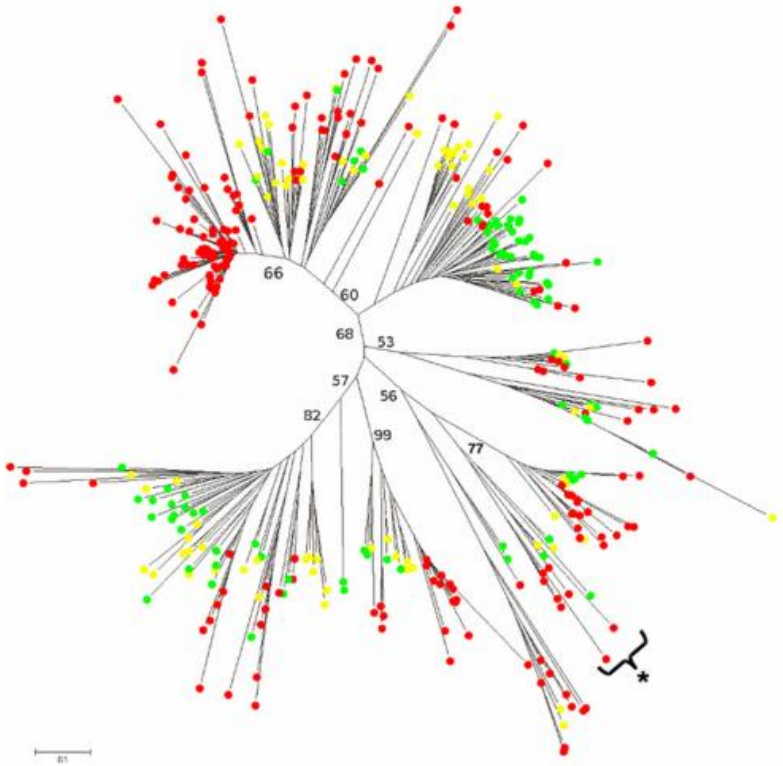



**Fig. 1** Phylogenetic analysis of Ty1-***copia*** retroelements**.** Bootstrap values were calculated for 1000 replicates; only those greater than 50 are shown. Paralogs from maize elements are marked with yellow circles; those from rice with green circles, and those from teff with red circles. "*" indicates the clade containing elements related to the rice LTR-RT family RIRE1

[Arabidopsis TE catalogue]

| | | | |
|---|---|---|---|
| **M** | | | |
| | **Family Name** | **Super Family** | **Number of Transposable Elements** |
| | META1 | LTR/Copia | 138 |
| **R** | | | |
| | **Family Name** | **Super Family** | **Number of Transposable Elements** |
| | RathE2_cons | RathE2_cons | 74 |
| | ROMANIAT5 | LTR/Copia | 49 |
| | RathE3_cons | RathE3_cons | 104 |
| | RathE1_cons | RathE1_cons | 213 |
| | RP1_AT | DNA | 87 |
| **S** | | | |
| | **Family Name** | **Super Family** | **Number of Transposable Elements** |
| | SADHU | null | 16 |
| | SIMPLEHAT2 | DNA/HAT | 73 |
| | SIMPLEHAT1 | DNA/HAT | 56 |
| | SIMPLEGUY1 | DNA/Harbinger | 116 |
| **T** | | | |
| | **Family Name** | **Super Family** | **Number of Transposable Elements** |
| | TA1-2 | LTR/Copia | 4 |
| | TA12 | LINE/L1 | 2 |
| | TA11 | LINE/L1 | 157 |
| | TSCL | LINE? | 81 |
| | TNAT1A | DNA | 162 |
| | TNAT2A | DNA | 38 |
| | TAG2 | DNA/HAT | 86 |
| | TAG3N1 | DNA/HAT | 97 |
| | TAG1 | DNA/HAT | 28 |
| | TAT1_ATH | LTR/Gypsy | 87 |
| | TA1_AT | LTR/Gypsy | 3 |

Prokaryotic genomes have some repeated sequences (<1% of the genome)

- **Insertion sequences** (**IS**); small (<2.5 kb) sequences encoding proteins implicated in transposition flanked by inverted repeats. (**DNA elements**)
- **Repetitive extragenic palindromic** (**REP**) **sequences**; 20–50 bp in length, singly or in arrays. If one or more REP sequences are located immediately downstream of a gene, then they can be transcribed as an extension of the mRNA, then folding into stem-loop structures which might play a role in regulation
- **Clustered regularly interspaced short palindromic repeats** (**CRISPRs**); 20-50 bp in tandem arrays, with repeats separated by spacers of even length (but different sequence). The spacers are the bases of “immunity”, acting as guide RNA that guide the Cas endonuclease (encoded by a sequence near the CRISPR) towards phage DNA

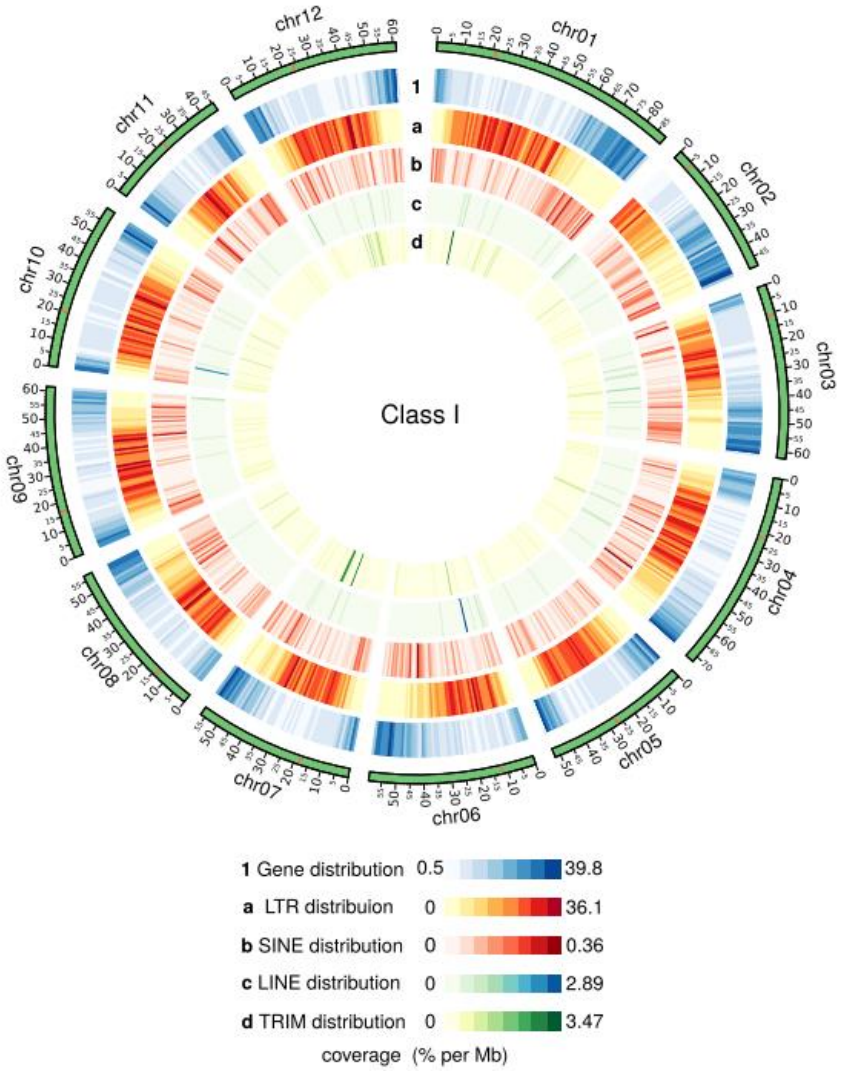
Original Article | Published: 20 May 2020

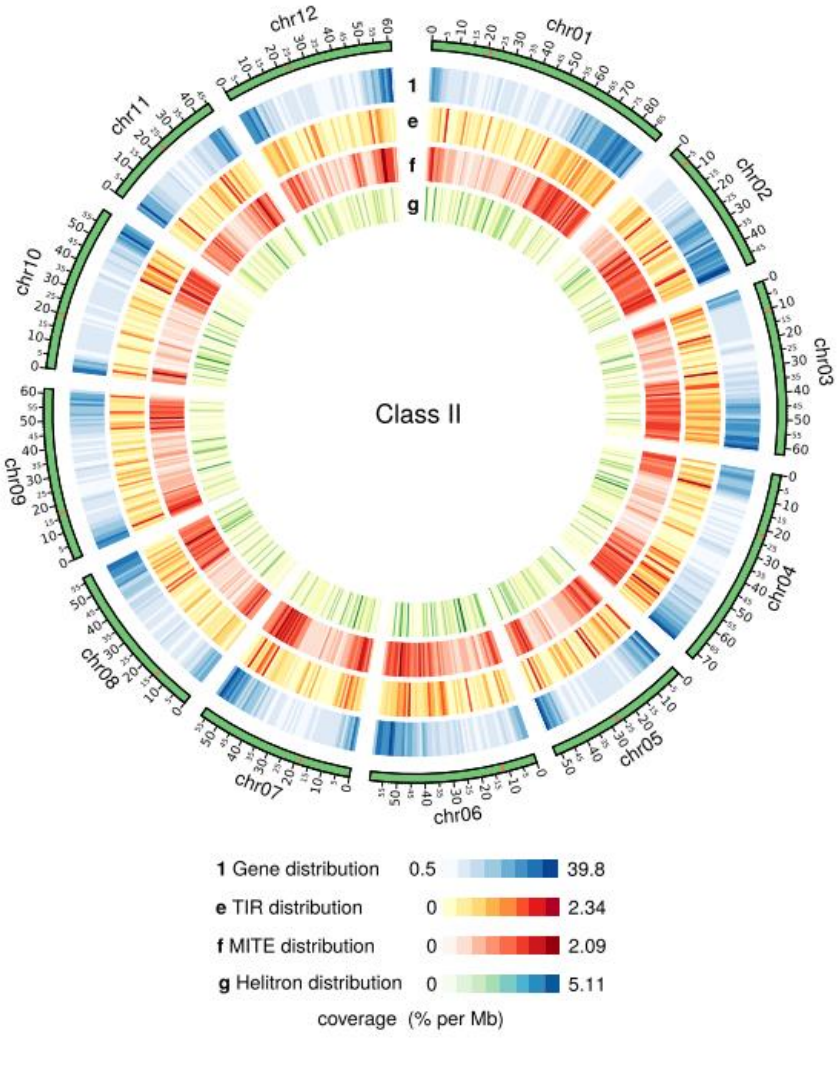
# Genomic re-assessment of the transposable element landscape of the potato genome

Diego Zavallo, Juan Manuel Crescente, Magdalena Gantuz, Melisa Leone, Leonardo Sebastian Vanzetti, Ricardo Williams Masuelli & Sebastian Asurmendi

*Plant Cell Reports* **39**, 1161–1174 (2020) | Cite this article

Genetic elements and TEs are differentially distributed along the genome



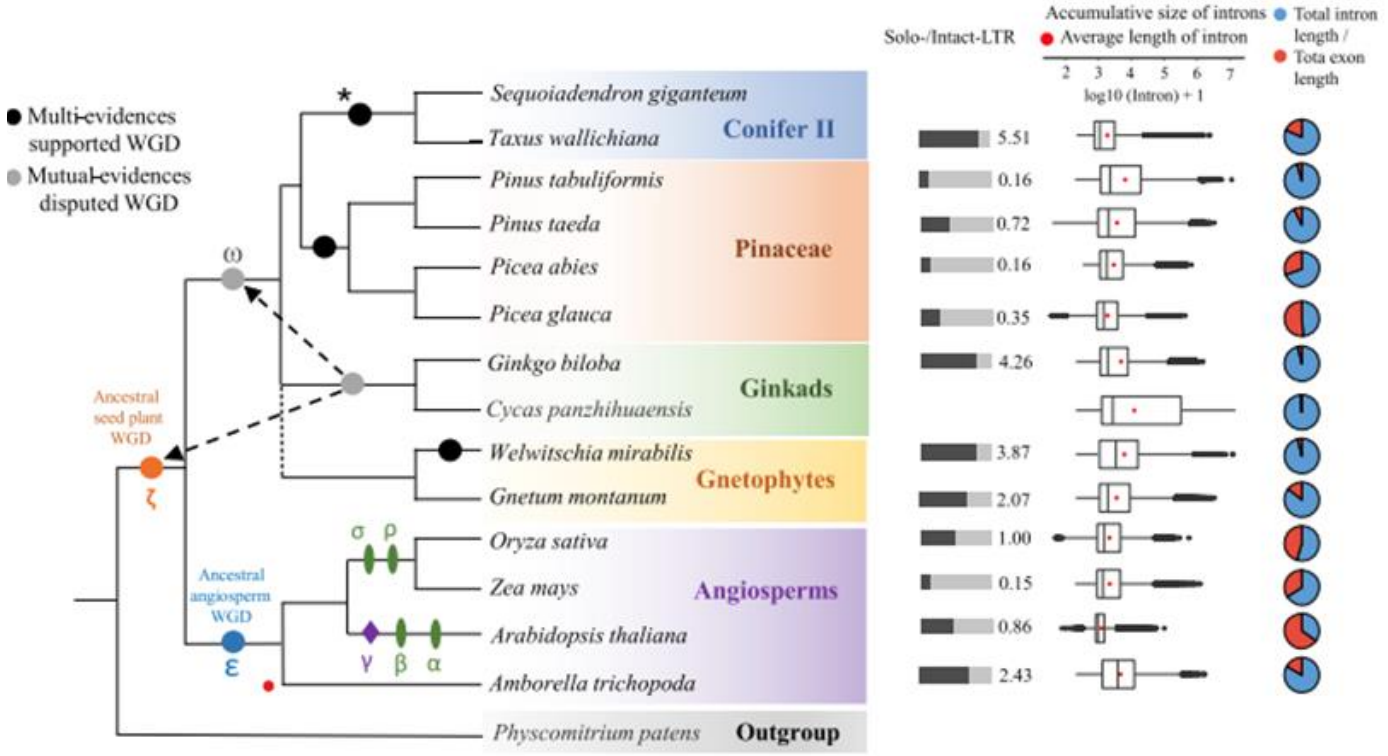
OXFORD (GIGA)n SCIENCE

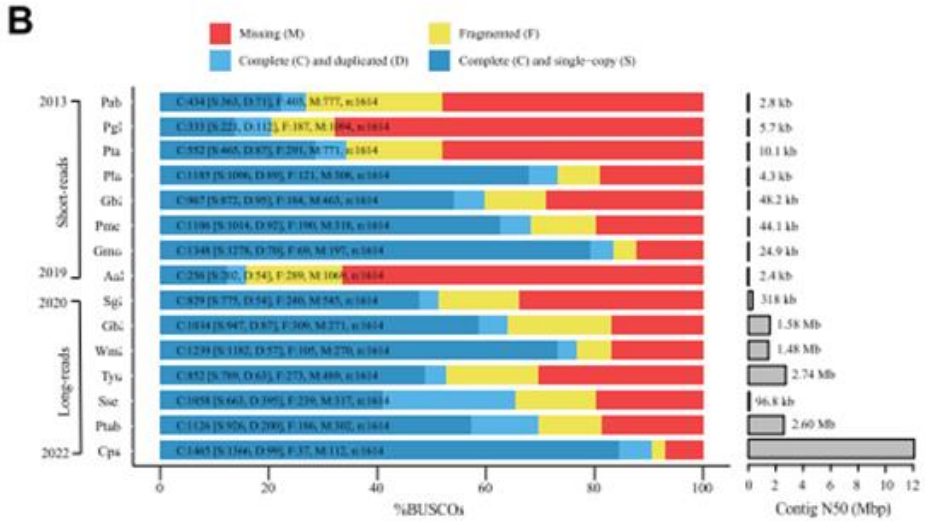
GigaScience, 2022, 11, 1–10
DOI: 10.1093/gigascience/giac078
Review Article

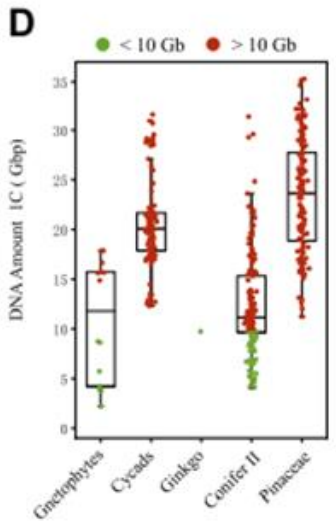
# Evolution of complex genome architecture in gymnosperms

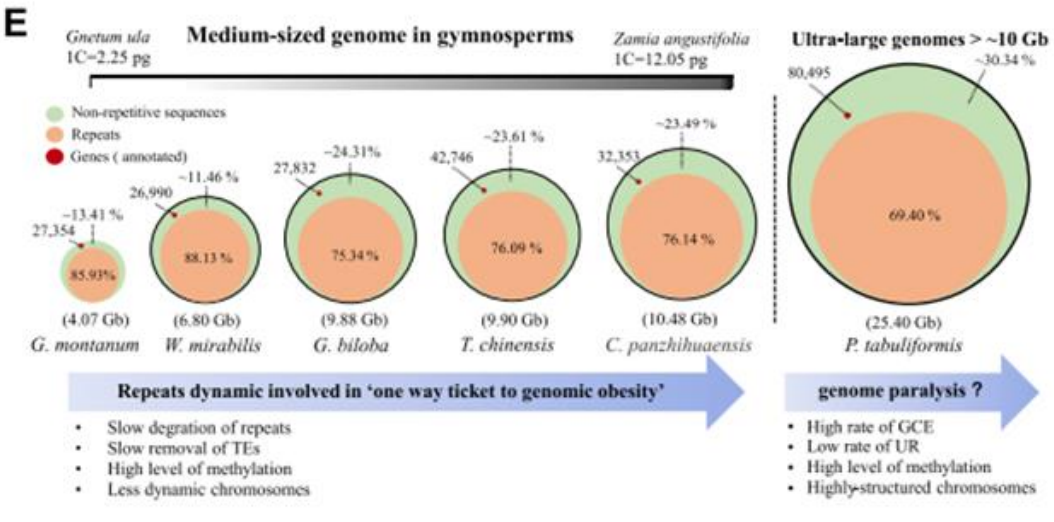
Tao Wan[1,2,3], Yanbing Gong[4,5], Zhiming Liu[3], YaDong Zhou[6], Can Dai[7] and Qingfeng Wang[1,2,*]

Ancient but continuous amplification of TEs within a range of 5 to 50 Ma explains much of coniferales genome size



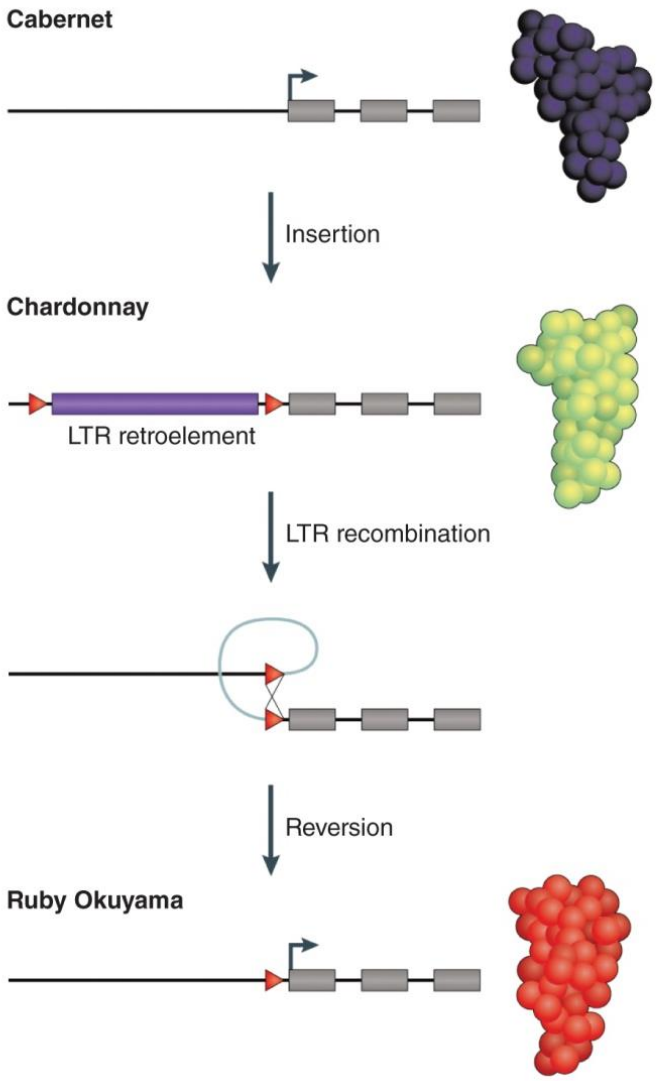



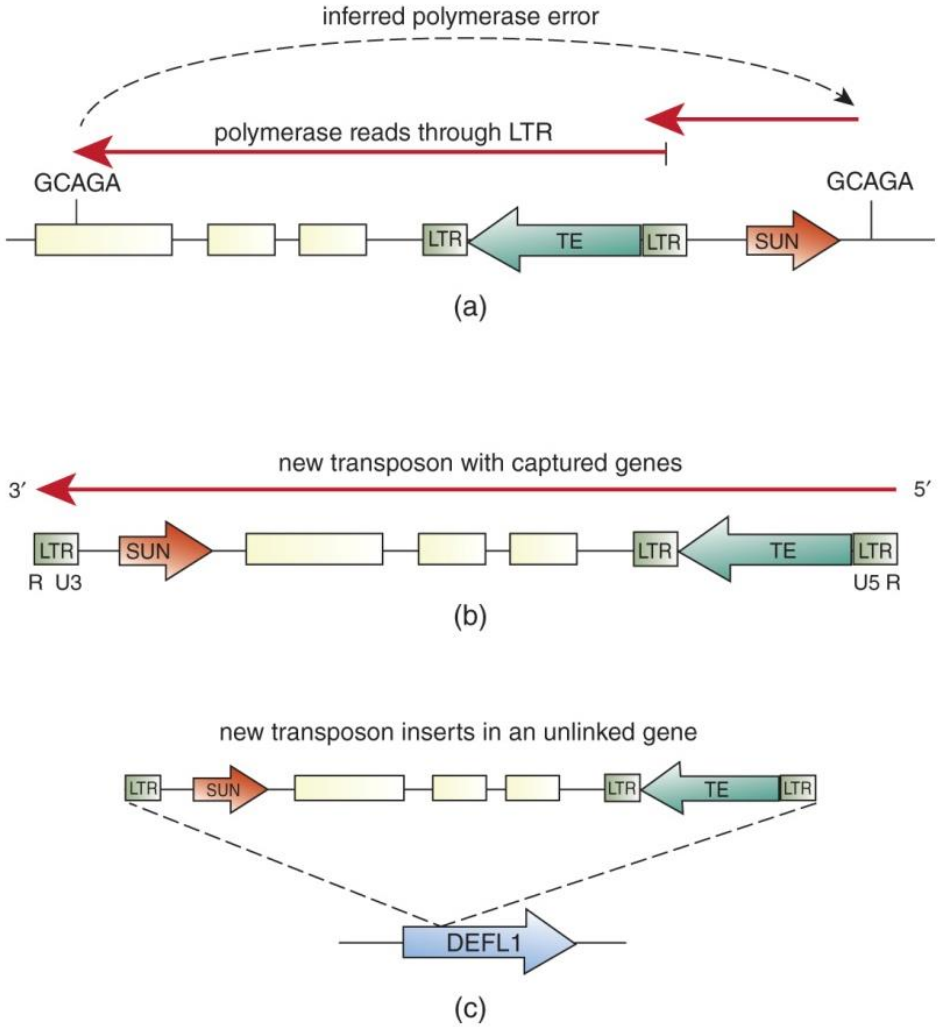
## Transposable elements can perturb gene expression

- Depending on where TEs end up being, they may perturb gene function or expression
- *e.g.* in grapes, normally dark (cabernet);
  - When an LTR retroelement landed right upstream the pigment genes, their expression was suppressed
  - If the LTR retroelement recombines leaving the LTR in, the expression of pigments is in between
- A similar mechanism is the cause of Mendel's wrinkled peas, as well as of sticky rice
- Note that TE can also enhance gene transcription: the LTR also act as promoters of the transposon genes, but may be coopted by nearbly genes (*e.g.* wheat tolerance to Al+ through TE-driven overexpression of citrate channel)



## Transposable elements can move genes/gene fragments

- When a DNA element is excised, may promote reshuffling of gene fragments
- Tipically, most of gene fragments don't work; but some do
- In rice,22% of the genes carried by MULEs is transcribed
- About 60% of maize helitrons carry gene sequences and could easily contribute to their movement and reshuffling with different promoters




*SUN* locus in tomato causes elongated fruit. Transcription begins from 5' LTR, continues through 3' UTR and continues in a downstream gene with homologous sequence GCAGA. The new TE with captured genes gets inserted in DEFL1 so that is subjected to DEFL1 promoter

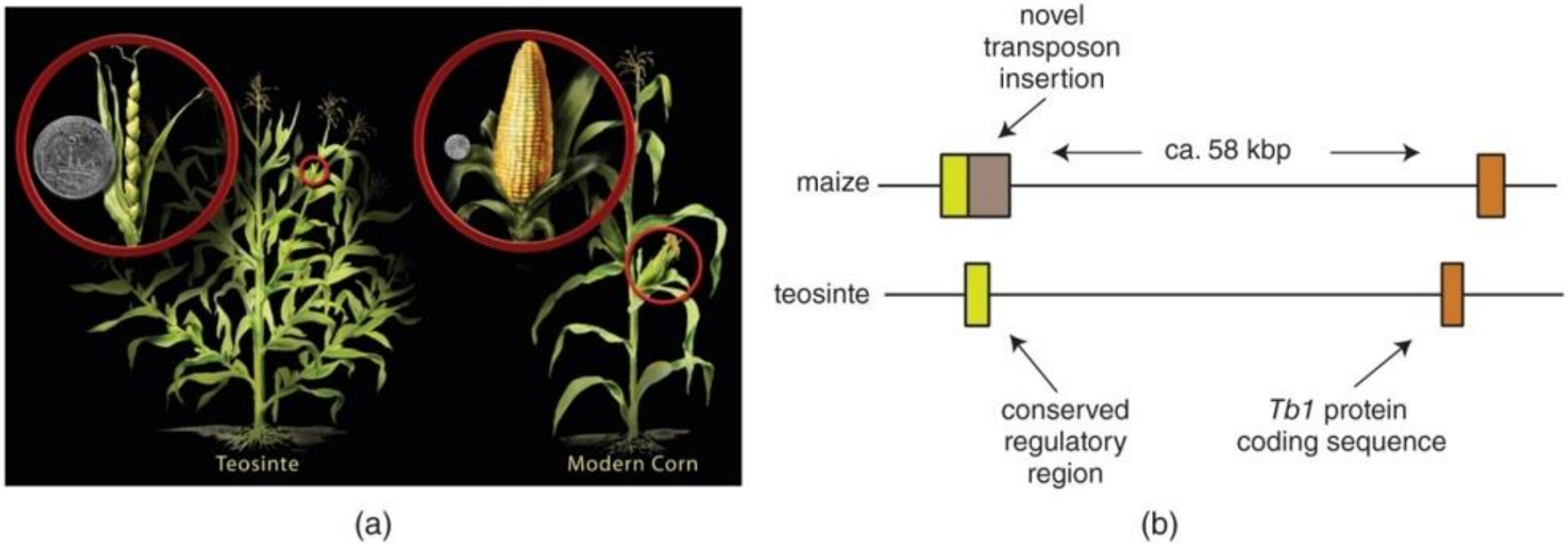



**Figure 2.4** Maize vs. teosinte. (a) Overall phenotype of the plants, showing many more branches in teosinte; figure from http://nsf.gov/news/mmg/media/images/maize1_f.jpg. Photo Nicolle Rager Fuller, National Science Foundation. (b) Cartoon of the *tb1* locus in maize and teosinte. Orange boxes, *Tb1* coding region; yellow boxes, upstream regulatory regions; brown box, novel insertion of transposon in cultivated maize. Drawing approximately to scale

RESEARCH ARTICLE Open Access

# Domestication of rice has reduced the occurrence of transposable elements within gene coding regions

Xukai Li[1,2,3], Kai Guo[1,2,4], Xiaobo Zhu[1,2,3], Peng Chen[1,2,3], Ying Li[1,2,3], Guosheng Xie[1,2,3], Lingqiang Wang[1,2,3], Yanting Wang[1,2,3], Staffan Persson[1,3,5*] and Liangcai Peng[1,2,3*]

TEs interfering with genes are a major force of evolution

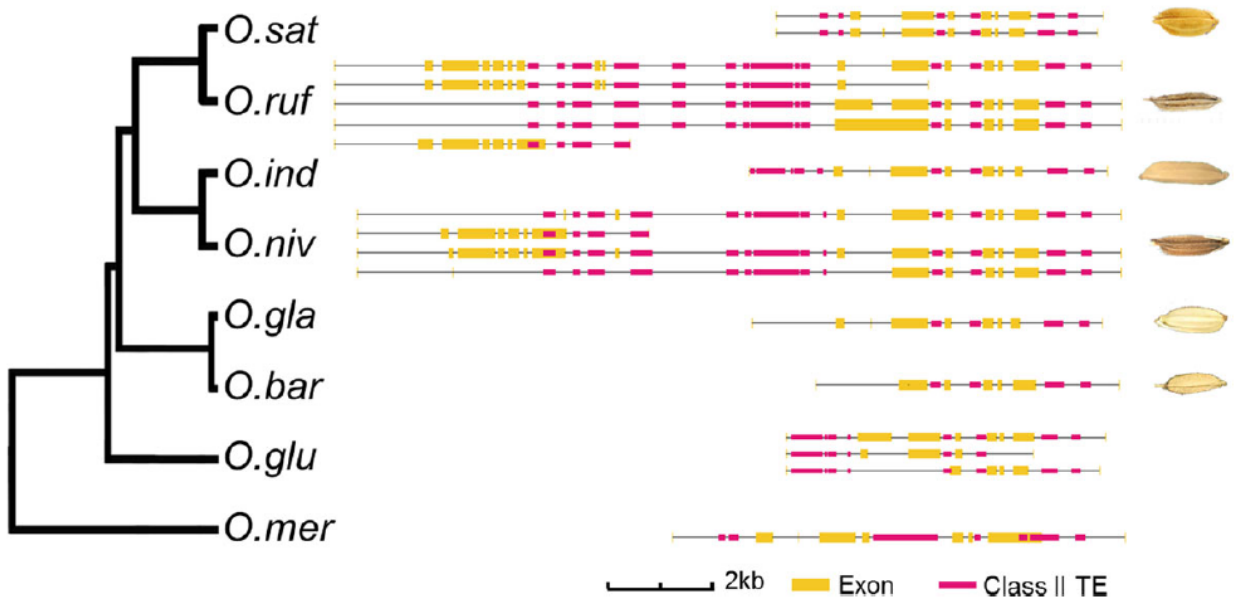
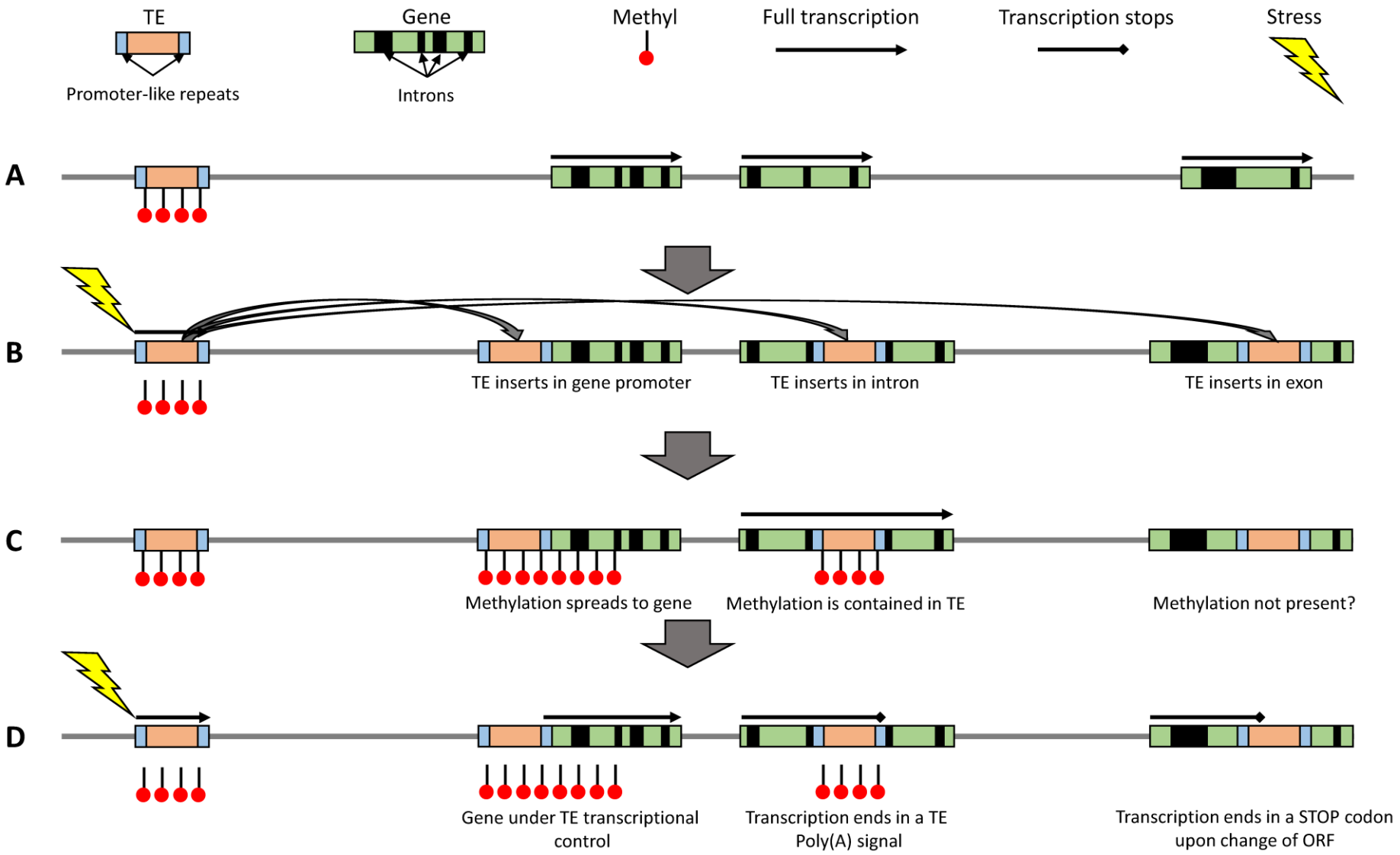



**Fig. 6** Comparisons of gene structure and TE locations of *GIF1* gene critical for grain filling in the eight rice species. Organization of exons, introns and TEs of *GIF1* (*GRAIN INCOMPLETE FILLING 1*; LOC_Os04g33740) gene in gene body and 2-kbp flanking sequences of the gene. Seed images of ancestral wild rice and cultivated rice are shown to the *right*

# How TEs are controlled?



# DNA methylation enables transposable element-driven genome expansion

Wanding Zhou[a,b,1], Gangning Liang[c], Peter L. Molloy[d], and Peter A. Jones[e,1]

[a]Center for Computational and Genomic Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA 19104; [b]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104; [c]Department of Urology, USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90089; [d]Nutrition and Health Program, Health and Biosecurity, Commonwealth Scientific and Industrial Research Organisation, North Ryde, NSW 2113, Australia; and [e]Center for Epigenetics, Van Andel Institute, Grand Rapids, MI 49503

Contributed by Peter A. Jones, June 19, 2020 (sent for review December 12, 2019; reviewed by Haig Kazazian, Matthew C. Lorincz, and Dustin E. Schones)

**Multicellular eukaryotic genomes show enormous differences in size. A substantial part of this variation is due to the presence of transposable elements (TEs). They contribute significantly to a cell's mass of DNA and have the potential to become involved in host gene control. We argue that the suppression of their activities by methylation of the C–phosphate–G (CpG) dinucleotide in DNA is essential for their long-term accommodation in the host genome and, therefore, to its expansion. An inevitable consequence of cytosine methylation is an increase in C-to-T transition mutations via deamination, which causes CpG loss. Cytosine deamination is often needed for TEs to take on regulatory functions in the host genome. Our study of the whole-genome sequences of 53 organisms showed a positive correlation between the size of a genome and the percentage of TEs it contains, as well as a negative correlation between size and the CpG observed/expected (O/E) ratio in both TEs and the host DNA. TEs are seldom found at promoters and transcription start sites, but they are found more at enhancers, particularly after they have accumulated C-to-T and other mutations. Therefore, the methylation of TE DNA allows for genome expansion and also leads to new opportunities for gene control by TE-based regulatory sites.**

- Eykariotic genomes display a 64,000-fold variation in their sizes, mostly due to transposable elements and methylation
- A long-term outcome of methylation is an increase in C-to-T transition mutations both in the TEs and host DNA
- This can be observed as a decreased proportion of CpG dinucleotides over evolutionary time
- Surviving TE provide additional DNA for evolvability of the host organism

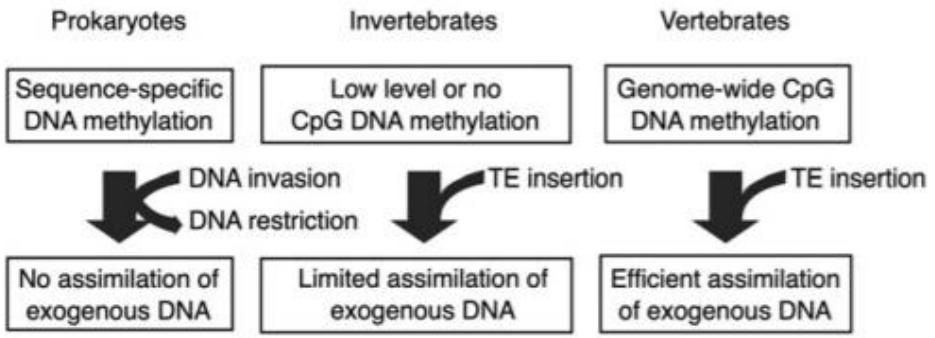
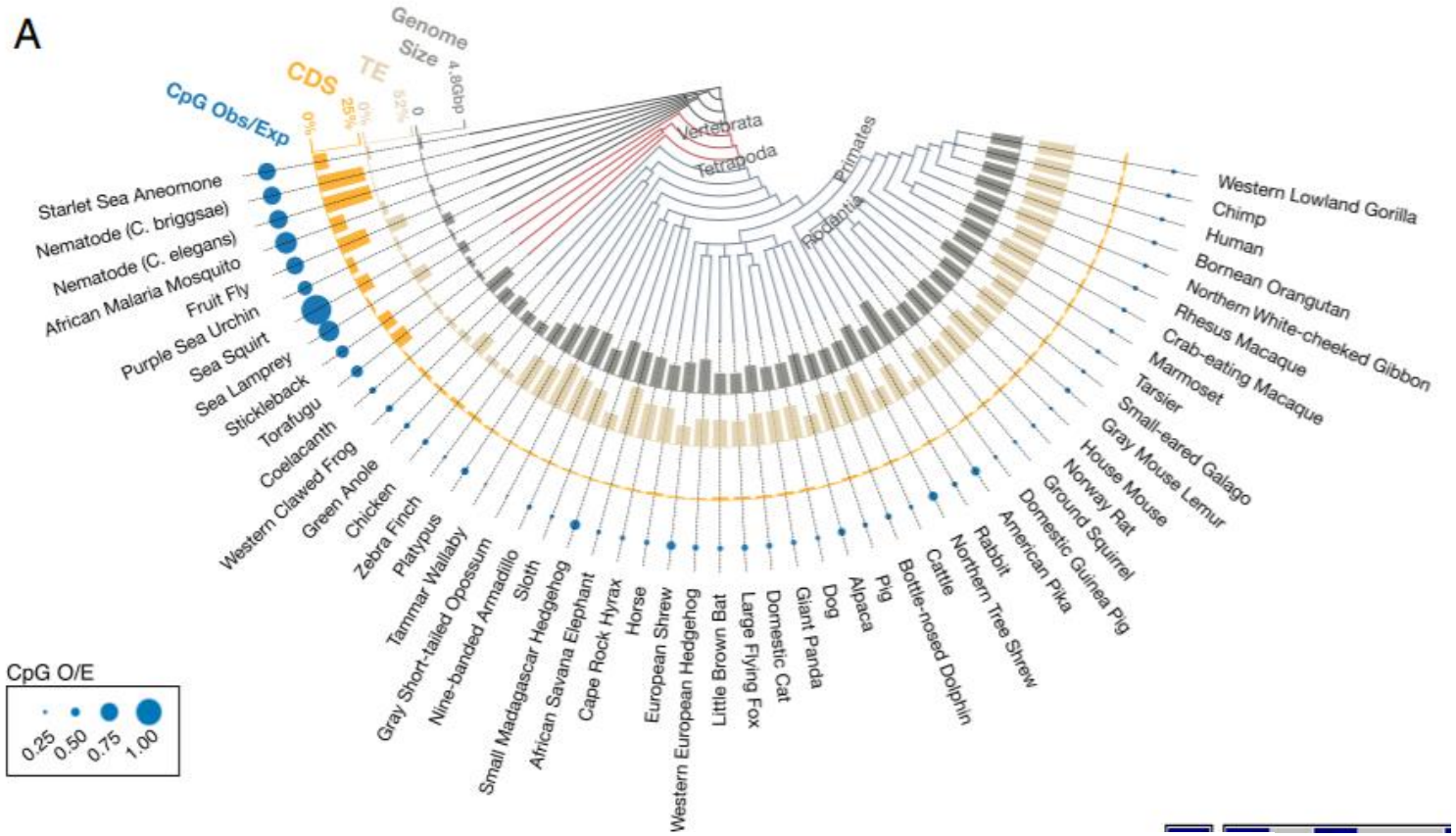
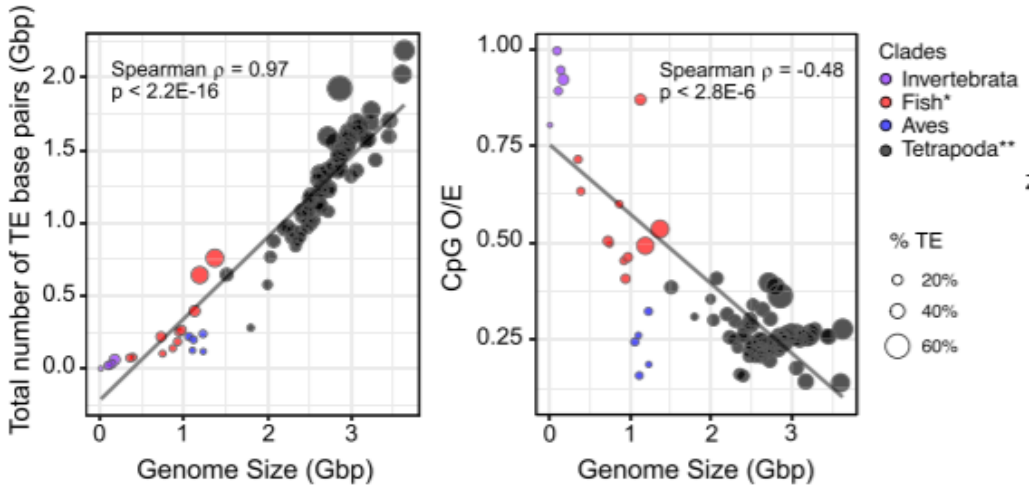



**Fig. 1.** Model illustrating differing roles for DNA methylation in handling exogenous DNA. DNA methylation in prokaryotes is part of their restriction/modification system of host defense. Invertebrates can accommodate TE DNA to a limited extent due to low prevalence of DNA CpG methylation. Vertebrates, especially mammals, have extensive CpG methylation on a genomic scale and can tolerate high levels of TEs.



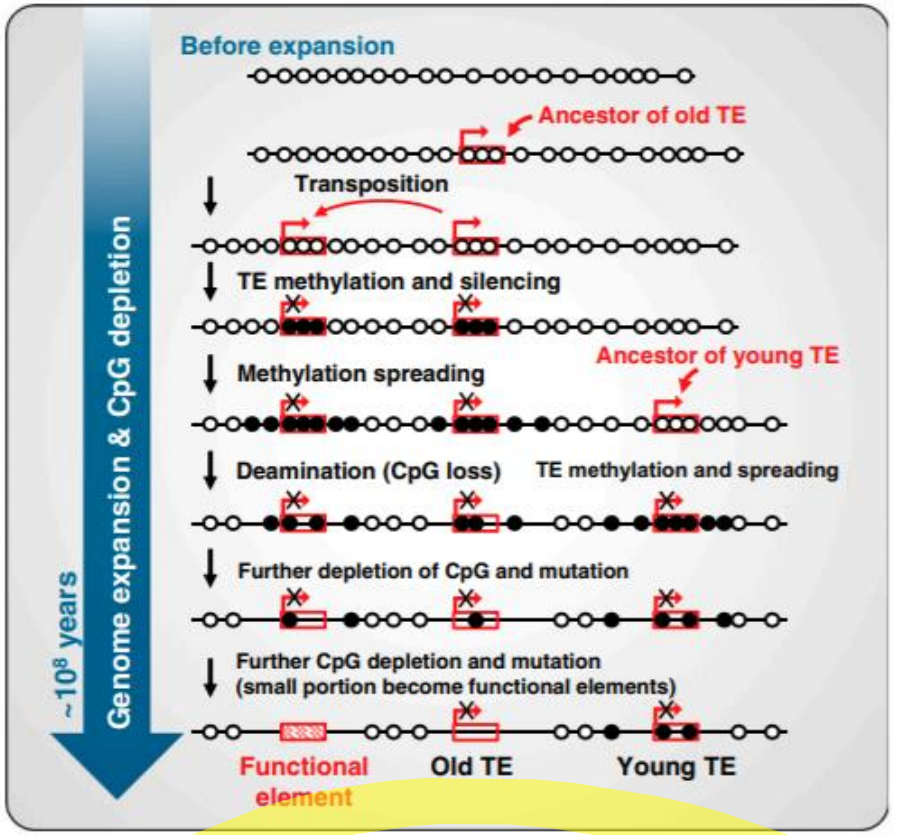




**Fig. 6.** CpG methylation contributes to TE-mediated genome expansion and ultimately to CpG depletion by deamination and neofunctionalization of TEs in the expanded genome. The model depicts an early genome with no TEs and the unmethylated CpG sites shown as open circles and methylated CpGs as solid black circles. At this stage, the CpG O/E ratio is about 1. Insertion and transposition of a TE lead to its de novo methylation (shown as black circles) and silencing of the TE. Methylation can then spread into the flanking host DNA. Methylated CpGs have an enhanced mutation frequency relative to unmethylated CpGs and a half-life of about 35 million y in the primate germline (10). Over evolutionary time, this leads to an overall depletion of CpGs in the entire genome with the exception of CpG islands (11) and ultimately to the creation of new functional elements such as enhancers, depicted by the decreasing number of methylation sites and a decrease in CpG O/E ratio.

There’s a delicate equilibrium b/w TE density and genome «evolvability»

- The host genome tries to suppress their movement, but not too much
- Tipically, different classes of TEs will follow different fates

Why more TEs near centromeres?

- With time, accumulated TEs can get lost by (illegitimate) recombination and deletion
- In regions with low recombination, TEs are lost with less efficiency
- Selection may also purge TEs close to genes; their methylation will expand to nearby loci and inactivate genes reducing individual fitness

- Indeed, repetitive DNA explains much of the **C paradox**
- Bursts of TE expansions may explain different sizes also in close taxa
- A balance exists between new insertions and DNA loss (via, *e.g.* unequal recombination)

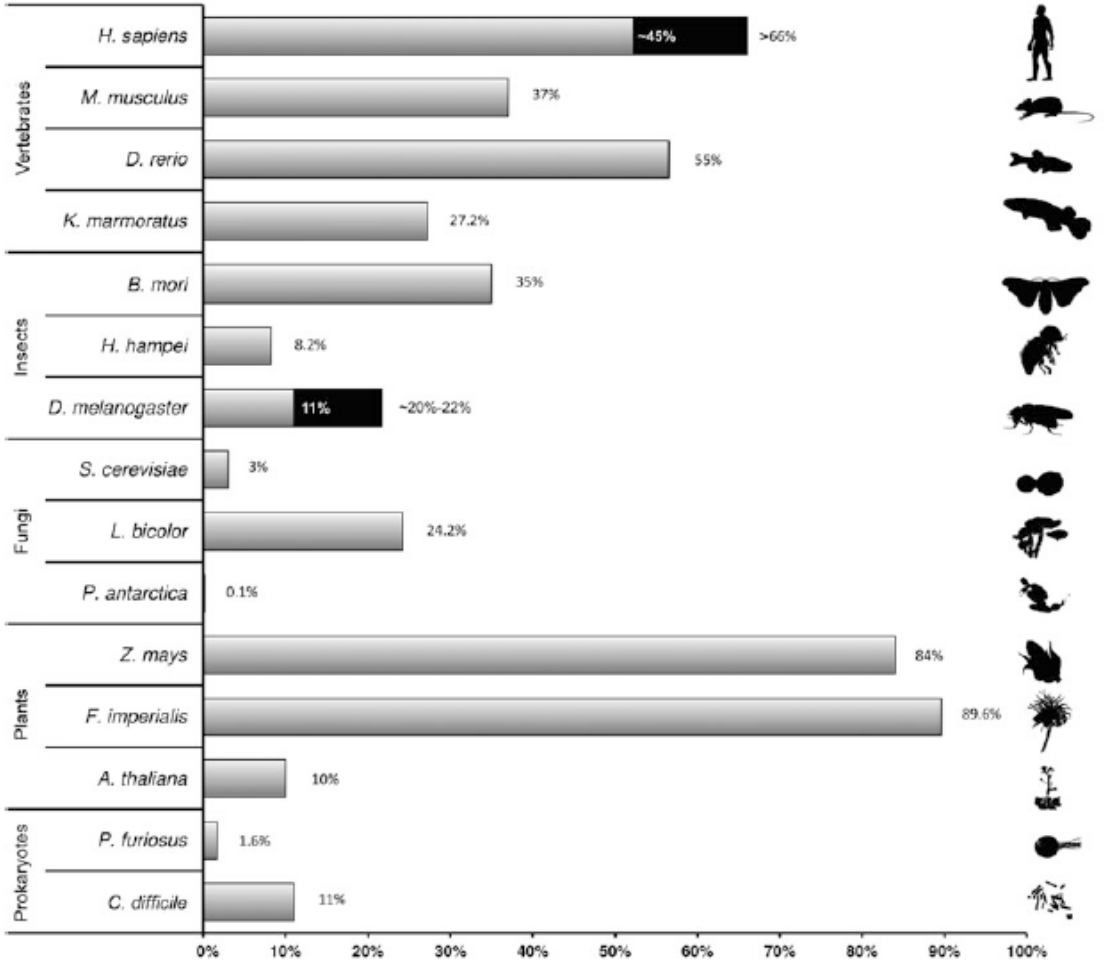
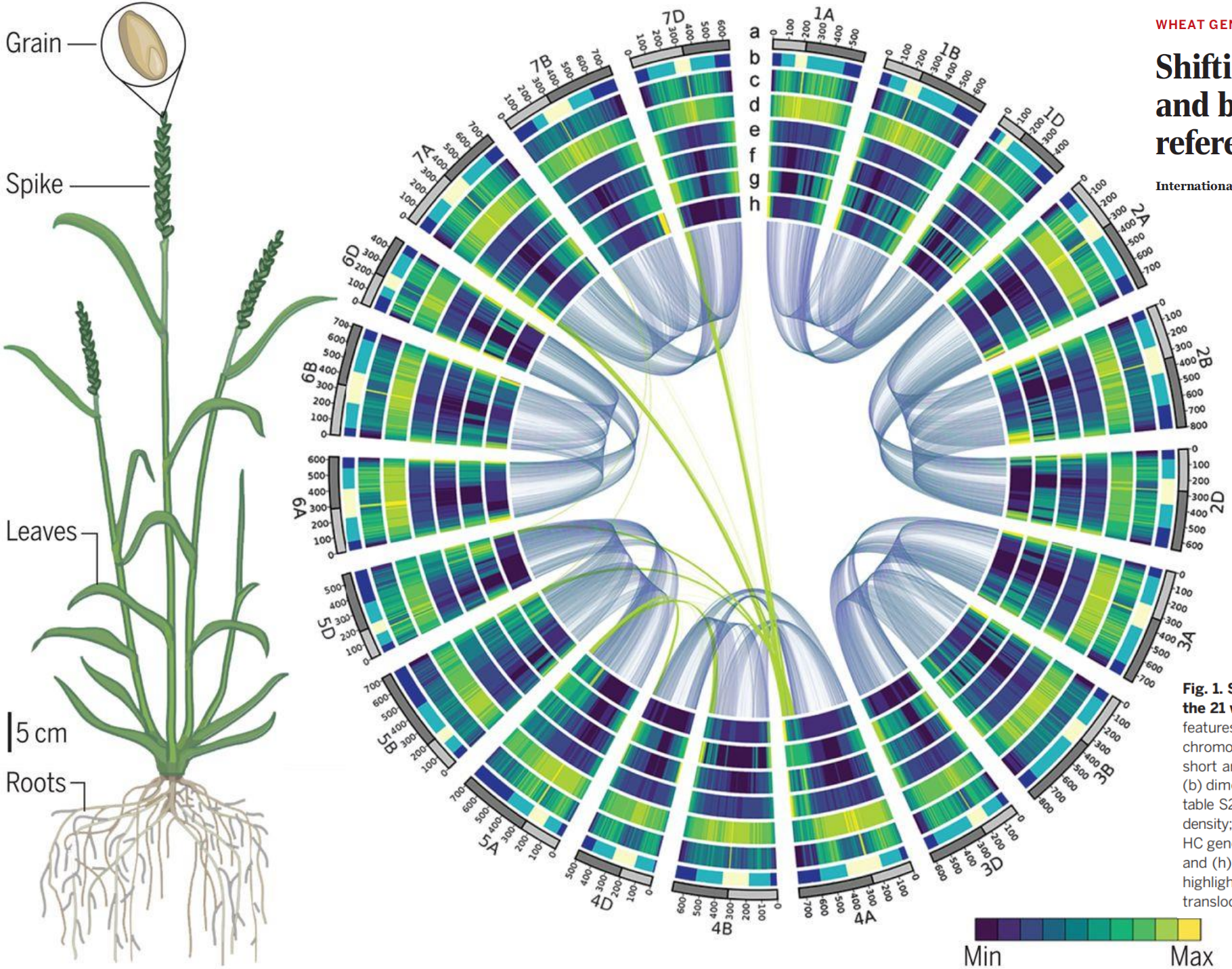



**Fig. 1** TE content in the genome of different organisms expressed as percentage of the genome: *Homo sapiens* (~45% [12], >66% [13]) *Mus musculus* [143], *Saccharomyces cerevisiae* [144], *Arabidopsis thaliana* [145], *Pyrococcus furiosus* [146], *Clostridium difficile* [147], *Danio rerio* [133], *Kryptolebias marmoratus* [148], *Bombyx mori* [149], *Hypothenemus hampei* [150], *Drosophila melanogaster* (11%, [68], ~20% [69]), *Pseudozyma antarctica*, and *Laccaria bicolor* [151]. *Zea mays* [152] and *Fritillaria imperialis* [8]. All estimates were obtained with homology-based methods except [13] that uses P-cloud and [69] that uses de novo approaches




WHEAT GENOME

# Shifting the limits in wheat research and breeding using a fully annotated reference genome

International Wheat Genome Sequencing Consortium (IWGSC)*

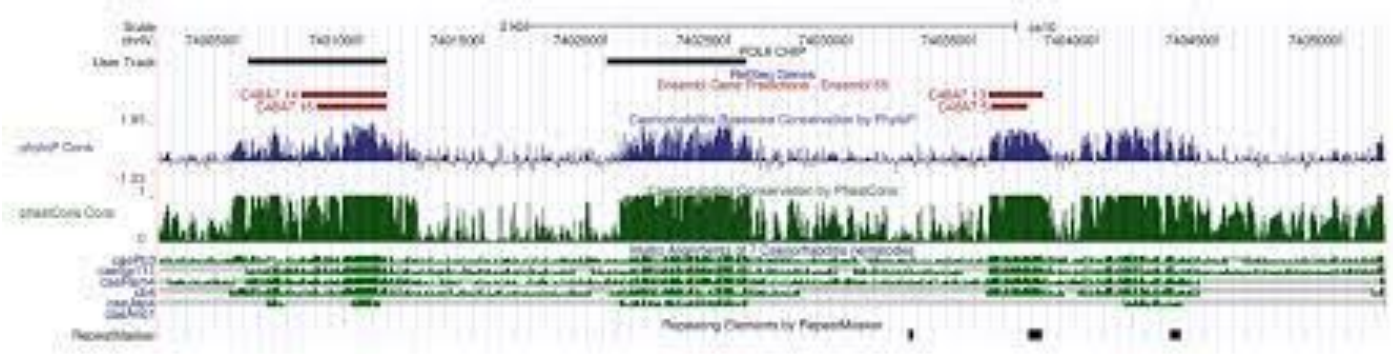
**Fig. 1. Structural, functional, and conserved synteny landscape of the 21 wheat chromosomes.** (**A**) Circular diagram showing genomic features of wheat. The tracks toward the center of the circle display (a) chromosome name and size (100-Mb tick size; light gray bar indicates the short arm and dark gray indicates the long arm of the chromosome); (b) dimension of chromosomal segments R1, R2a, C, R2b, and R3 [(*18*) and table S29]; (c) *K*-mer 20-frequencies distribution; (d) LTR-retrotransposons density; (e) pseudogenes density (0 to 130 genes per Mb); (f) density of HC gene models (0 to 32 genes per Mb); (g) density of recombination rate; and (h) SNP density. Connecting lines in the center of the diagram highlight homeologous relationships of chromosomes (blue lines) and translocated regions (green lines). (**B**) Distribution of Pfam domain

# Observe data in genome browsers

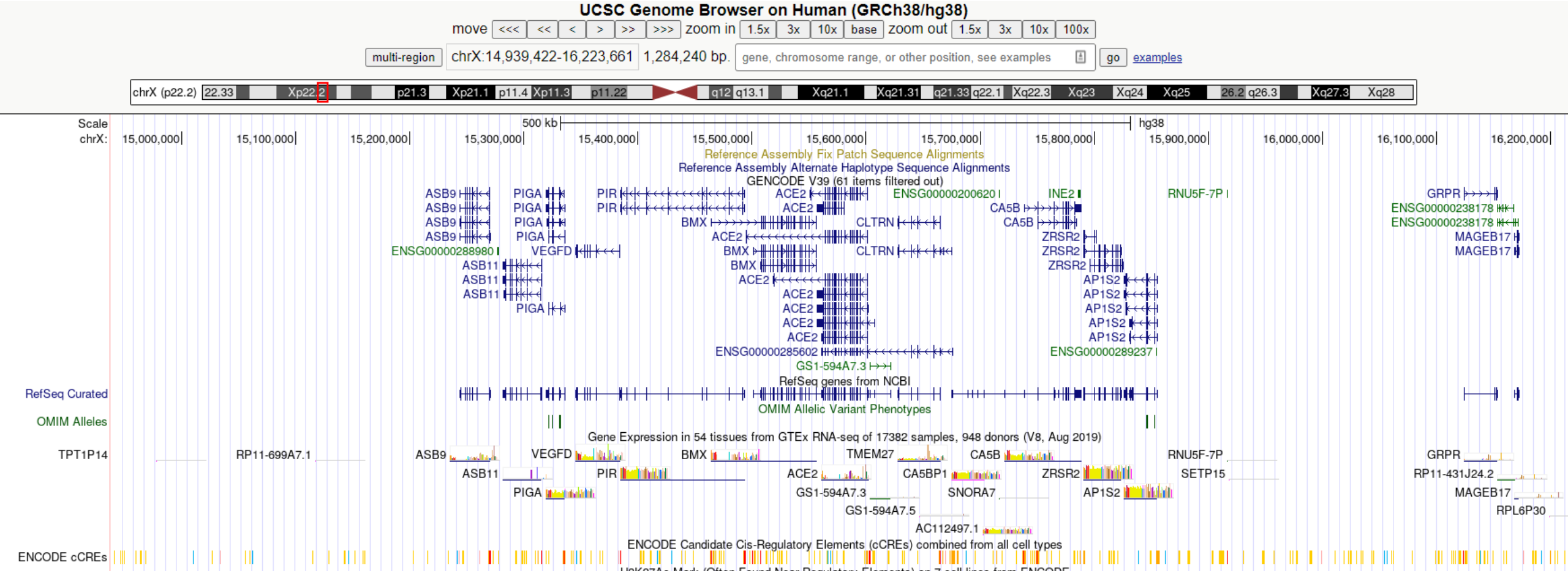
- Genome browsers are visual interfaces to genomic data
- They exist in many flavours, for different species
- Bring together different datasets in the same framework: the reference genome

Quite some differences depending on organism:

- https://genome-euro.ucsc.edu/ (human)
- https://www.gramene.org/ (plants)



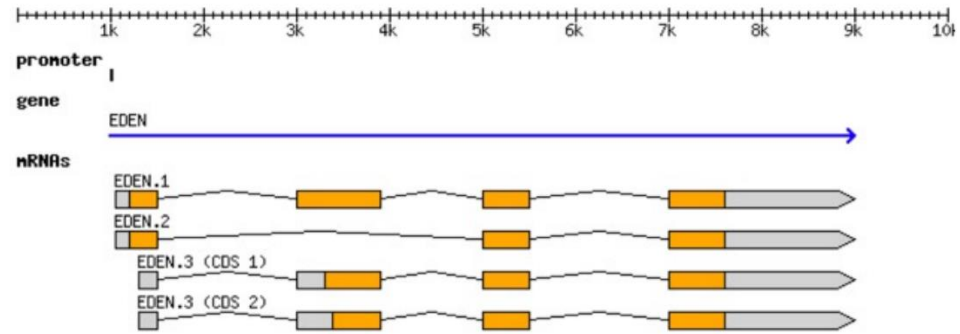
# Local organization of genes



Generic feature format (GFF) records genomic features

nine-column, tab-delimited, plain text files

1. seqid: The ID of the sequence
2. source: Algorithm or database that generated this feature
3. type: gene/exon/CDS/etc…
4. start: 1-based coordinate
5. end: 1-based coordinate
6. score: E-values/p-values/index/colors/…
7. strand: “+’ for positive “-” for minus, “.” not stranded
8. phase: For "CDS", where the feature begins with reference to the reading frame (0,1,2)
9. attributes: A list of tag=value features Parent: Indicates the parent of the feature (group exons into transcripts, transcripts into genes, …)ts in gene models




```
 0  ##gff-version 3.1.26
 1  ##sequence-region ctg123 1 1497228
 2  ctg123 . gene            1000  9000  .  +  .  ID=gene00001;Name=EDEN
 3  ctg123 . TF_binding_site 1000  1012  .  +  .  ID=tfbs00001;Parent=gene00001
 4  ctg123 . mRNA            1050  9000  .  +  .  ID=mRNA00001;Parent=gene00001;Name=EDEN.1
 5  ctg123 . mRNA            1050  9000  .  +  .  ID=mRNA00002;Parent=gene00001;Name=EDEN.2
 6  ctg123 . mRNA            1300  9000  .  +  .  ID=mRNA00003;Parent=gene00001;Name=EDEN.3
 7  ctg123 . exon            1300  1500  .  +  .  ID=exon00001;Parent=mRNA00003
 8  ctg123 . exon            1050  1500  .  +  .  ID=exon00002;Parent=mRNA00001,mRNA00002
 9  ctg123 . exon            3000  3902  .  +  .  ID=exon00003;Parent=mRNA00001,mRNA00003
10  ctg123 . exon            5000  5500  .  +  .  ID=exon00004;Parent=mRNA00001,mRNA00002,mRNA00003
11  ctg123 . exon            7000  9000  .  +  .  ID=exon00005;Parent=mRNA00001,mRNA00002,mRNA00003
12  ctg123 . CDS             1201  1500  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
13  ctg123 . CDS             3000  3902  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
14  ctg123 . CDS             5000  5500  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
15  ctg123 . CDS             7000  7600  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
16  ctg123 . CDS             1201  1500  .  +  0  ID=cds00002;Parent=mRNA00002;Name=edenprotein.2
17  ctg123 . CDS             5000  5500  .  +  0  ID=cds00002;Parent=mRNA00002;Name=edenprotein.2
18  ctg123 . CDS             7000  7600  .  +  0  ID=cds00002;Parent=mRNA00002;Name=edenprotein.2
19  ctg123 . CDS             3301  3902  .  +  0  ID=cds00003;Parent=mRNA00003;Name=edenprotein.3
20  ctg123 . CDS             5000  5500  .  +  1  ID=cds00003;Parent=mRNA00003;Name=edenprotein.3
21  ctg123 . CDS             7000  7600  .  +  1  ID=cds00003;Parent=mRNA00003;Name=edenprotein.3
22  ctg123 . CDS             3391  3902  .  +  0  ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
23  ctg123 . CDS             5000  5500  .  +  1  ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
24  ctg123 . CDS             7000  7600  .  +  1  ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
```