From: Wang et al. (2021) Which factors contribute most to genome size variation withing angiosperms? Ecology and Evolution 11:2660.

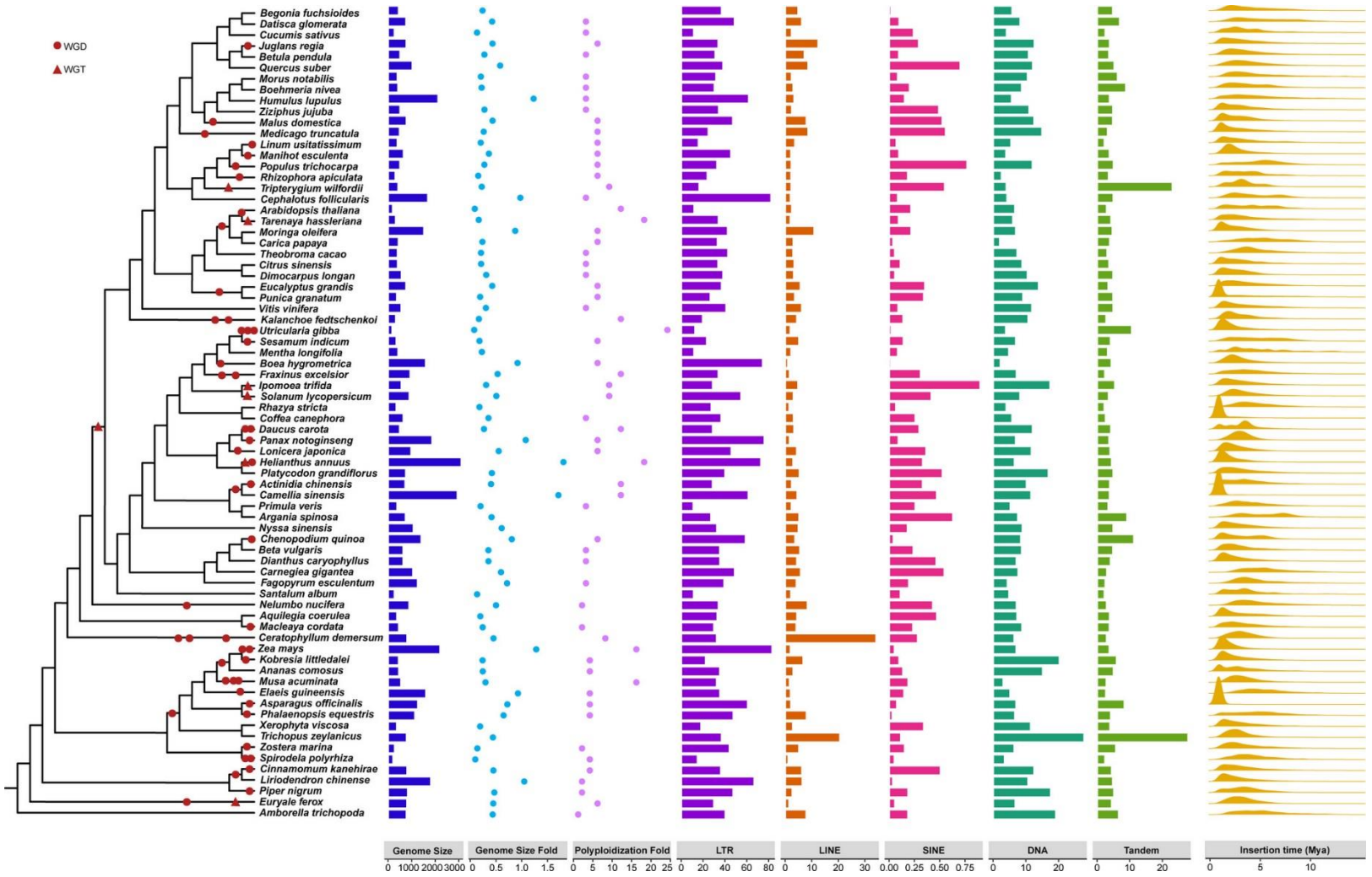


Figure 2. Representation of phylogeny and the correlation factors analyzed in 74 genomes. GF: genome size fold, PF: polyploidization fold, tandem: tandem repeats, and LTR insertion dates. GF indicates the genome size fold in plants scaled by the ancestral genome size for angiosperms and PF indicates the value of polyploidization fold which is the number of times that whole-genome duplication and the whole-genome triplication occurred. LTR-tandem indicates different proportions of corresponding repeating elements in genomes as a percentage (%). Mean insertion date indicates the estimated distribution of LTR insertion dates in plants in millions of years. The WGDs and WGTs are labeled in the branches. The topology information cited is from Li et al. (2019)

Plant Transposable Elements (TEs)

- **General Information: Bennetzen. 2000. Plant Molecular Biology 42:251**
- **Other references noted in text**

What are primary components of a genome?

- **Genes**
  - Expressed portion of the genome
  - Most abundant in the **euchromatic region of chromosomes**
- **Transposable elements**
  - Located in
    - **Heterochromatic repetitive region** of the genome
    - **Euchromatic gene-rich** regions and portion of the genome
  - Most **NOT** biologically active (**as promoters or repressors**)
- **Satellite DNA** (from Vondrak et al. 2020 Plant Journal 101:484)
  - Important in
    - Chromosome structure
    - Expression of genes
    - Chromosomal replication
  - **Located in**
    - **Pericentromeric region**
    - **Telomeres**
  - Tandem-arrayed, repeats of monomer sequences
    - Repeats <10 bp to 5 kb in size
    - Arrays of repeats 5-120 kb long
  - Repeat sequence often specific to species or genera
  - Evolved from
    - **Tandem amplification of a subset of the genome**
      - **Introns**
      - **Short repeat arrays**
      - **Retrotransposons**

Transposable Element Chromosomal Locations

- **De Assis et al. 2020. BMC Genomics 21:237**
- Fluorescent In Situ Hybridization (**FISH**)
- Long terminal repeat (LTR) retrotransposon probes

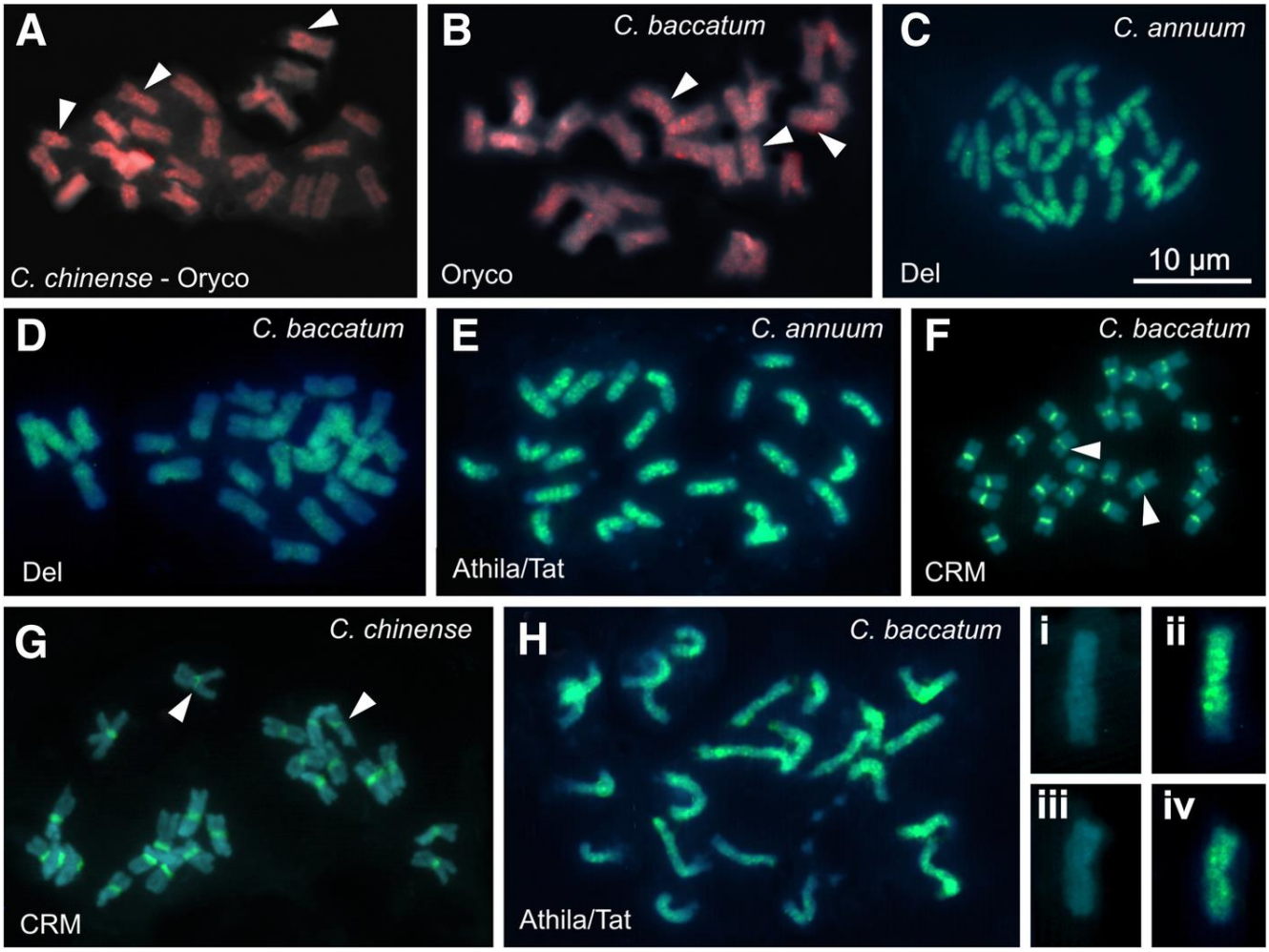
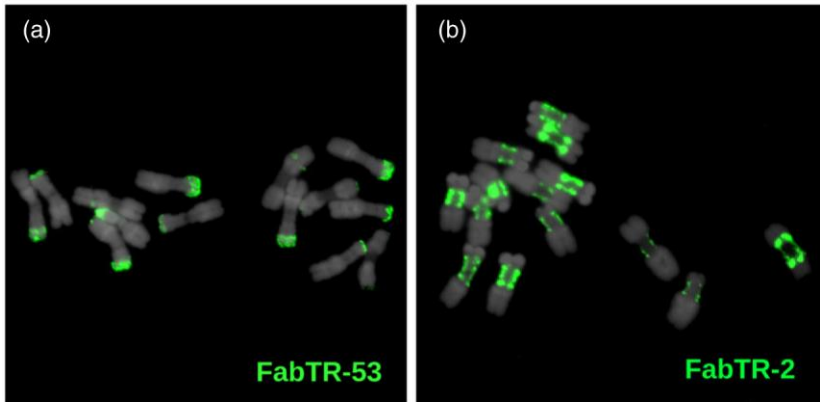


Figure 4. Genome relationships and LTR-retrotransposon diversity in three cultivated *Capsicum* L. (Solanaceae) species. The ***Copia* Ivana/Oryco** probe showed few hybridization signals scattered along chromosomes, with a low accumulated profile in both *C. chinense* (**A**) and *C. baccatum* (**B**). The ***Gypsy* Tekay/Del** probe exhibited hybridization signals dispersed along the chromosomes in the three species, but with a larger accumulation in *C. annuum* chromosomes (**C**) than the other two species, such as in *C. baccatum* (**D**). The ***Gypsy* Athila/Tat** probe showed brighter hybridization signals than **Tekay/Del**, accumulating in the pericentromeric to interstitial regions of all *C. annuum* chromosomes (**E**), differently of *C. baccatum* because some chromosomes accumulated many signals and others very few (**H**). The ***Gypsy* CRM** probe showed FISH signals accumulated in the centromeric regions, but with two pairs in each species with much less intense signals. Note the arrows in *C. baccatum* (**F**) and *C. chinense* (**G**). The boxes **i, ii, iii** and **iv** are highlighting differences in the pericentromeric and interstitial **Athila/Tat** signals in two *C. baccatum* chromosomes. The bar represents 10 µm.

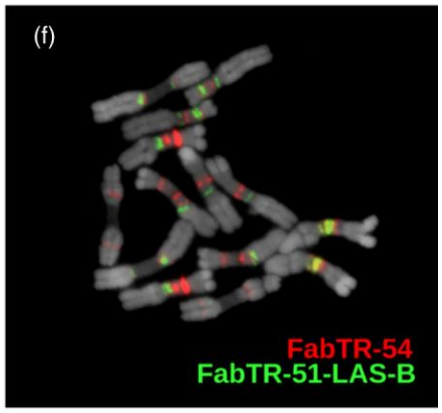
Satellite Chromosomal Locations

- **Vondrak et al 2020. Plant Journal 101:484**

Satellite Telomere and Pericentromeric Locations in Pepper



Paired Unique Satellites in Pericentromeric Region in Pepper



Concepts of TEs Changed Over Time

Original transposable elements concept

- **Mutagenic agents** that affected genes and the genome
  - TEs found in genes controlling mutant phenotypes supported the concept

Abundance of TEs

- **Found in all eukaryotic species**
  - Plants
  - Animals
- Greater than 50% for some genomes

What Are the Physical Locations of TEs in Genomes???

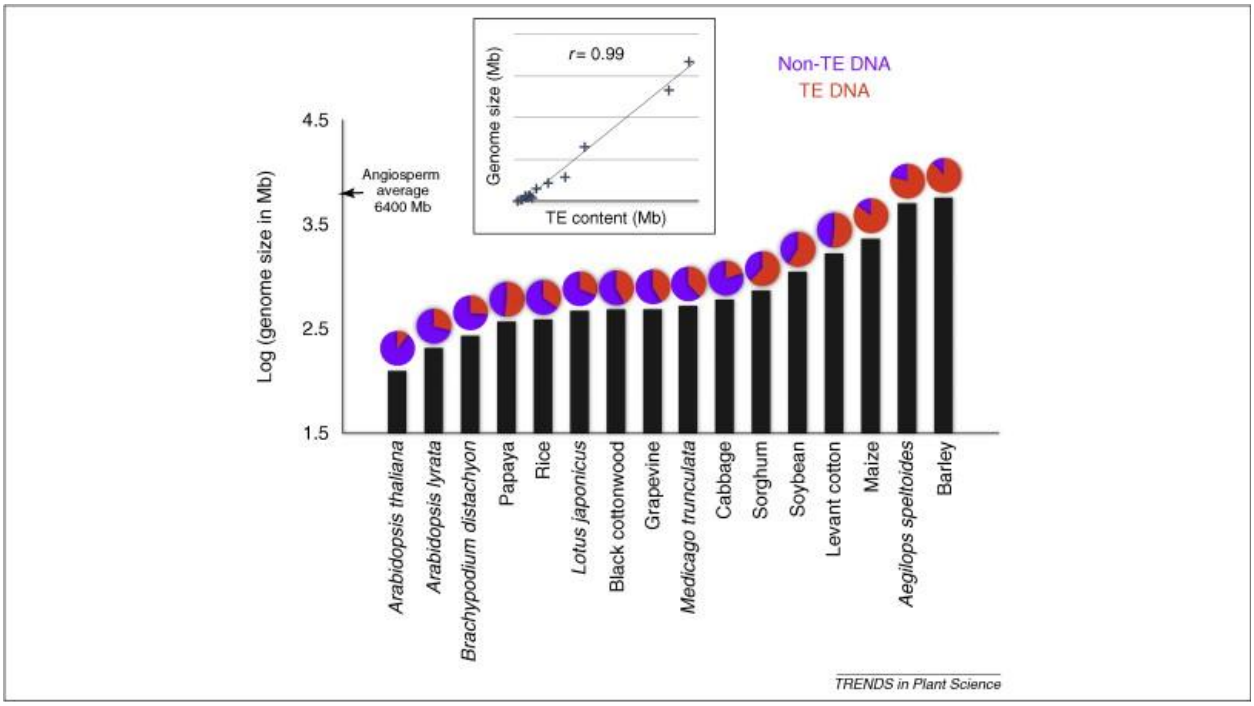
- Heterochromatic regions
- Euchromatic regions
- Promoters
- Introns of genes

Different Methylation Patterns

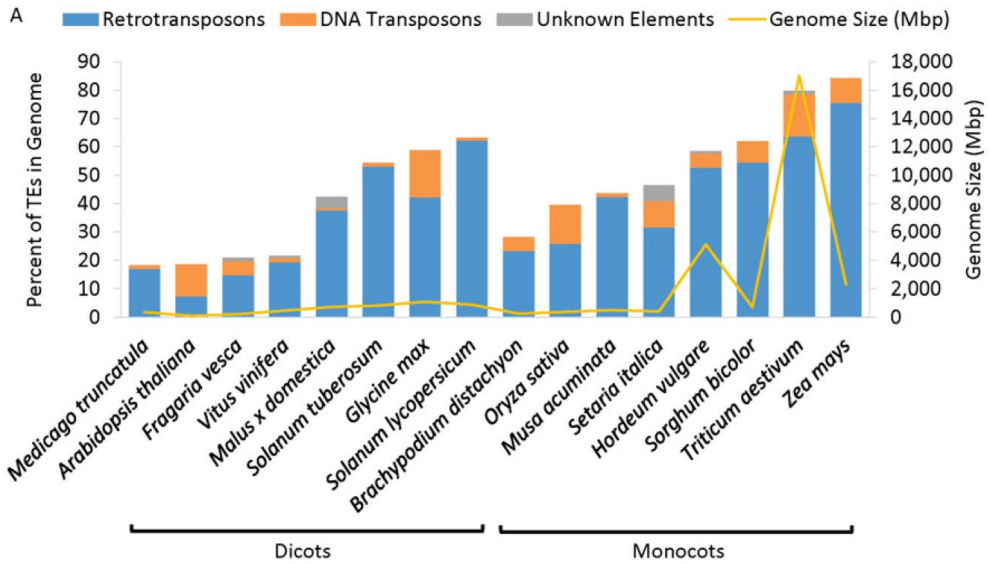
- **Genes**
  - Very low levels of CHG, CHH (H=A, T, or G) methylation
  - Moderate levels of CG methylation for moderately expressed genes
- **Introns**
  - CHG methylation in long introns
    - These introns contain TEs

Relationship of Genome Size and TE Content

- **TEs major component in all plant genomes**



- **Tenaillon et al (2010) Trends in Plant Science 15:471**



- **Hirsh and Springer 2017.** *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1860*:157.

Result of transposition

- Change in gene structure and gene activity

Evolving Concepts of Transposable Elements

McClintock concept of "controlling elements"

- Mid-late 1950s concept presented
- **Elements move from location to a new location**
  - ***Movement of elements is a source of new genetic variation for stressed populations***

Brittan and Davidson (1969)

- Gene Regulation for Higher Cells: A Theory
  - *Science* 165 (1969):349–57.
- Concept
  - Genes are a collection of modular units
    - One unit expresses the protein
    - Other units not expressed
      - ***Suggested mobile elements (transposable elements) could be a source of variation of the non-expressed region***
    - Predicted transcription factors bind to these sites

Properties of all TEs

- Move from genomic location to location
- Increase copy number

Types of TEs

- **Class I** elements in genomic terminology
  - ***Retrotransposon elements***
- **Class II** elements in genomic terminology
  - ***DNA elements***

First Detailed Analysis of Transposable Element Structure in Plants

- **San Miguel et al. 1996. Science 274:765.**
- **Corn Alcohol dehydrogenase 1 locus**
- ***280kb region around the gene***
  - Nesting of TEs within other TEs
    - ***Multiple rounds of transposition of retroelements in the genome history***

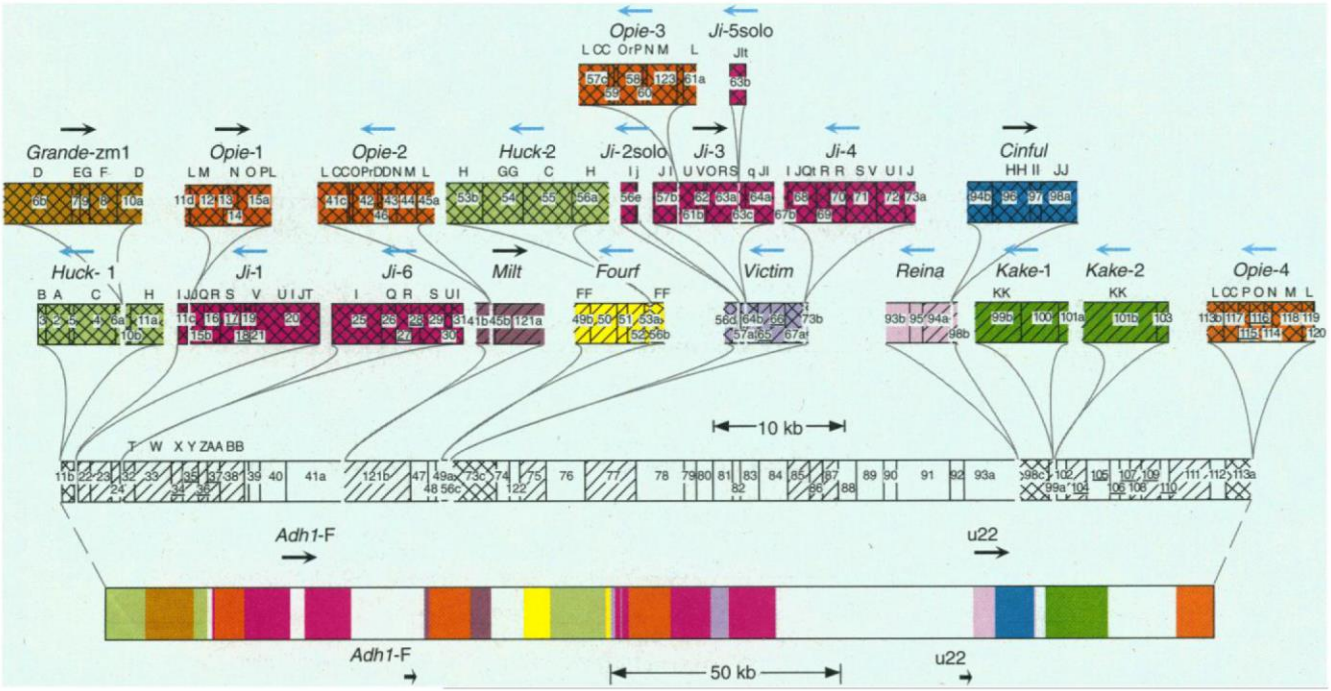
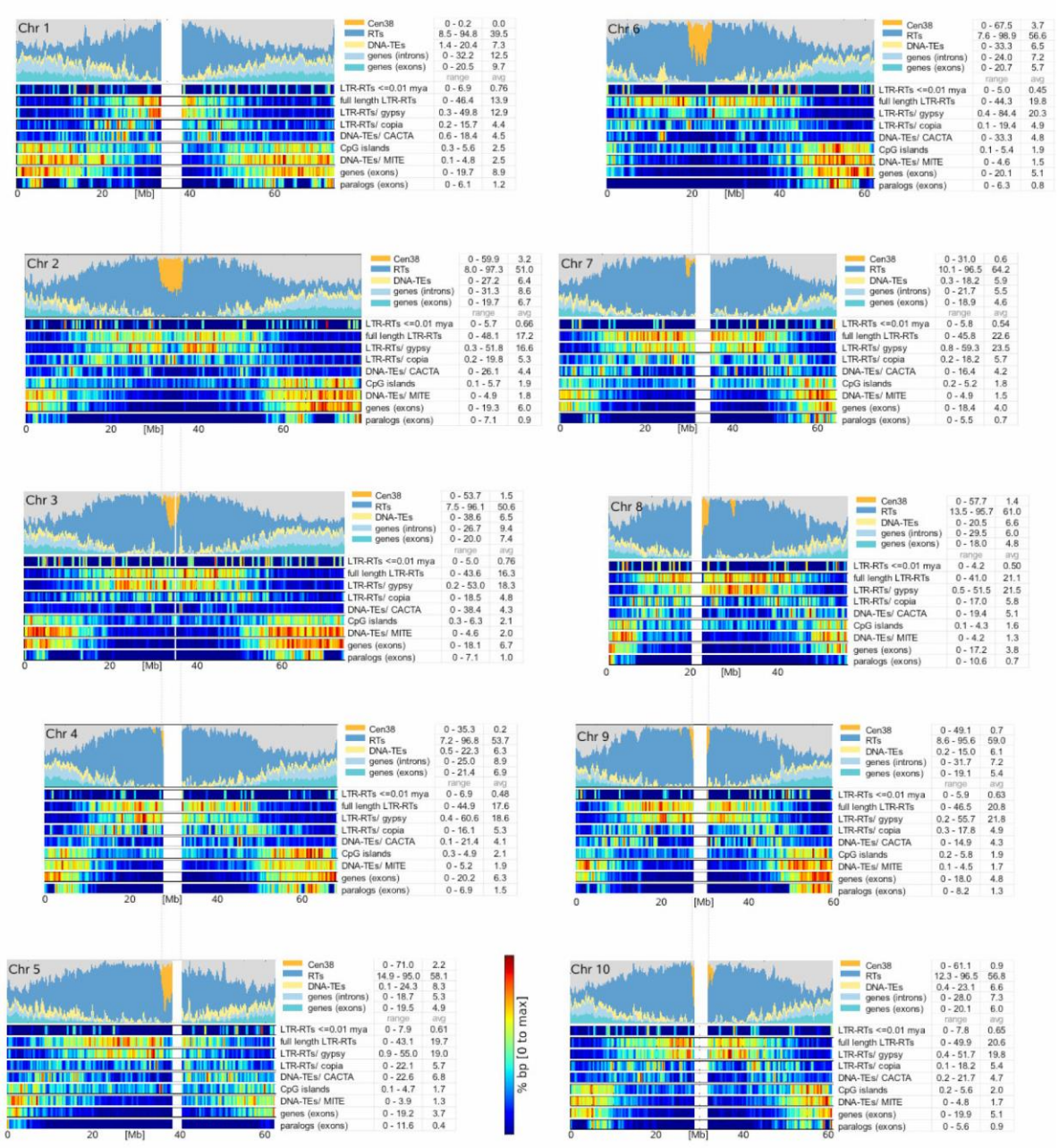


Fig. 3. Structure of the *Adh1*-F region of maize, showing identified retrotransposons. Labeling, bar patterns, and underlines are as in Fig. 1, but confirmed elements have been positioned above the DNA into which they have inserted. Curved lines below each element converge at the insertion site. The arrow above each element indicates its orientation. Fragments with lowercase a, b, or c designations are components of the fragment with the same number. GenBank accession numbers: *Fourf*, U68401; *Cinful*, U68402; *Grande*-Zm, U68403; *Huck*-2; U68404; *Ji*-3, U68405; *Kake*-1, U68406; *Milt*, U68407; *Reina*, U68409; and *Victim*, U68410.

Distributions of Transposable Elements in Sorghum

Paterson et al. 2010. Nature 457:551 (see supplement)

- In sorghum, ***Gypsy*** retrotransposons are clustered in the central heterochromatic region
- ***Copia*** retrotransposons and **DNA elements** distributed across the chromosomes
- Expressed genes found at the ends of chromosomes
    - Sorghum example



General Comments about TEs

Results of transposition

- **Change in gene structure and gene activity**
- **Source of new genetic variation for stressed populations**
  - Based on McClintock's concept of "controlling elements"

Properties of all TEs

- **Move from genomic location to location**
- **Increase copy number**

Types of TEs

- ***Retrotransposon* elements**
  - **Class I elements** in genomic terminology
- ***DNA* elements**
  - **Class II elements** in genomic terminology

Transposable Element Classification System

- **Wicker et al. 2007. Nat. Rev. Genetics 8:973**

Classification		Structure	TSD	Code	Occurrence
Order	**Superfamily**				
Class I (retrotransposons)					
LTR	*Copia*	GAG AP INT RT RH	4–6	RLC	P, M, F, O
	Gypsy	GAG AP RT RH INT	4–6	RLG	P, M, F, O
	Bel–Pao	GAG AP RT RH INT	4–6	RLB	M
	Retrovirus	GAG AP RT RH INT ENV	4–6	RLR	M
	ERV	GAG AP RT RH INT ENV	4–6	RLE	M
DIRS	*DIRS*	GAG AP RT RH YR	0	RYD	P, M, F, O
	Ngaro	GAG AP RT RH YR	0	RYN	M, F
	VIPER	GAG AP RT RH YR	0	RYV	O
PLE	*Penelope*	RT EN	Variable	RPP	P, M, F, O
LINE	*R2*	RT EN	Variable	RIR	M
	RTE	APE RT	Variable	RIT	M
	Jockey	ORF1 APE RT	Variable	RIJ	M
	L1	ORF1 APE RT	Variable	RIL	P, M, F, O
	I	ORF1 APE RT RH	Variable	RII	P, M, F
SINE	*tRNA*		Variable	RST	P, M, F
	7SL		Variable	RSL	P, M, F
	5S		Variable	RSS	M, O
Class II (DNA transposons) - Subclass 1					
TIR	*Tc1–Mariner*	Tase*	TA	DTT	P, M, F, O
	hAT	Tase*	8	DTA	P, M, F, O
	Mutator	Tase*	9–11	DTM	P, M, F, O
	Merlin	Tase*	8–9	DTE	M, O
	Transib	Tase*	5	DTR	M, F
	P	Tase	8	DTP	P, M
	PiggyBac	Tase	TTAA	DTB	M, O
	PIF–Harbinger	Tase* ORF2	3	DTH	P, M, F, O
	CACTA	Tase ORF2	2–3	DTC	P, M, F
Crypton	*Crypton*	YR	0	DYC	F
Class II (DNA transposons) - Subclass 2					
Helitron	*Helitron*	RPA Y2 HEL	0	DHH	P, M, F
Maverick	*Maverick*	C-INT ATP CYP POL B	6	DMM	M, F, O

Structural features

Long terminal repeats; Terminal inverted repeats; Coding region; Non-coding region; Diagnostic feature in non-coding region; Region that can contain one or more additional ORFs

Protein coding domains

AP, Aspartic proteinase; APE, Apurinic endonuclease; ATP, Packaging ATPase; C-INT, C-integrase; CYP, Cysteine protease; EN, Endonuclease; ENV, Envelope protein; GAG, Capsid protein; HEL, Helicase; INT, Integrase; ORF, Open reading frame of unknown function; POL B, DNA polymerase B; RH, RNase H; RPA, Replication protein A (found only in plants); RT, Reverse transcriptase; Tase, Transposase (* with DDE motif); YR, Tyrosine recombinase; Y2, YR with YY motif

Species groups

P, Plants; M, Metazoans; F, Fungi; O, Others

Figure 1 | **Proposed classification system for transposable elements (TEs).** The classification is hierarchical and divides TEs into two main classes on the basis of the presence or absence of RNA as a transposition intermediate. They are further subdivided into subclasses, orders and superfamilies. The size of the target site duplication (TSD), which is characteristic for most superfamilies, can be used as a diagnostic feature. To facilitate identification, we propose a three-letter code that describes all major groups and that is added to the family name of each TE. DIRS, *Dictyostelium* intermediate repeat sequence; LINE, long interspersed nuclear element; LTR, long terminal repeat; PLE, *Penelope*-like elements; SINE, short interspersed nuclear element; TIR, terminal inverted repeat.

DNA elements (Class II)

Found in all species

- **Original elements found in plants**
  - First described by McClintock in maize
  - ***(1940s-1950s)***
- ***Only element in bacteria***

Structure

- All have TIR (terminal inverted repeats)
  - Size: 11-100s nt
  - TIR sequence defines each DNA element class
  - Transposition factors recognize TIR in a specific manner

Autonomous vs. non-autonomous elements

- **Autonomous element**
  - ***Fully functional***
  - ***Transposes (moves) independent*** of other elements
- **Non-autonomous element**
  - ***Requires autonomous*** element transposase activity

The *Ac/Ds* System in Action
Maize pigment gene example
Feschotte et al. 2002. Nature Review Genetics 3:329

Mobile Element Movement in an Organism

Bronze locus of maize

- Encodes a protein required for seed color development
- If TE element resides in the gene
  - ***Gene is not functional***
    - ***The seed is colorless***
- If the element moves out of gene
  - **Color expressed in cells with element in gene**

During kernel development

- Movement can occur
  - Early development = ***full color development***
  - Mid-development = ***larger colored spots***
  - Late development = ***small colored spots***

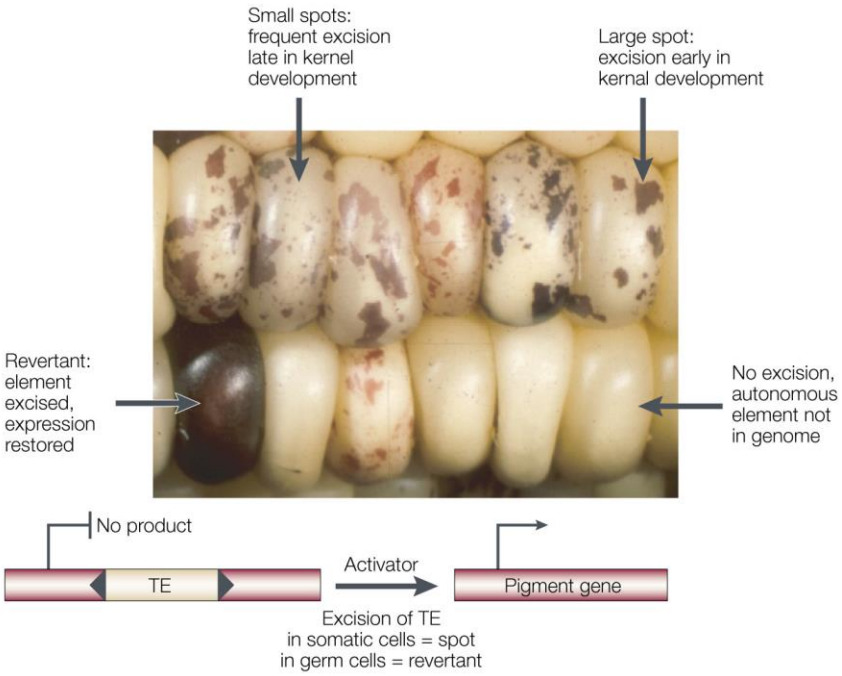


Figure 1 | **Using kernel phenotypes to study transposon behaviour.** Kernels on a maize ear show unstable phenotypes due to the interplay between a transposable element (TE) and a gene that encodes an enzyme in the anthocyanin (pigment) biosynthetic pathway. Sectors of revertant (pigmented) aleurone 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 (see main text for details).

Ac/Ds Transposable Element System of Maize

Ac element

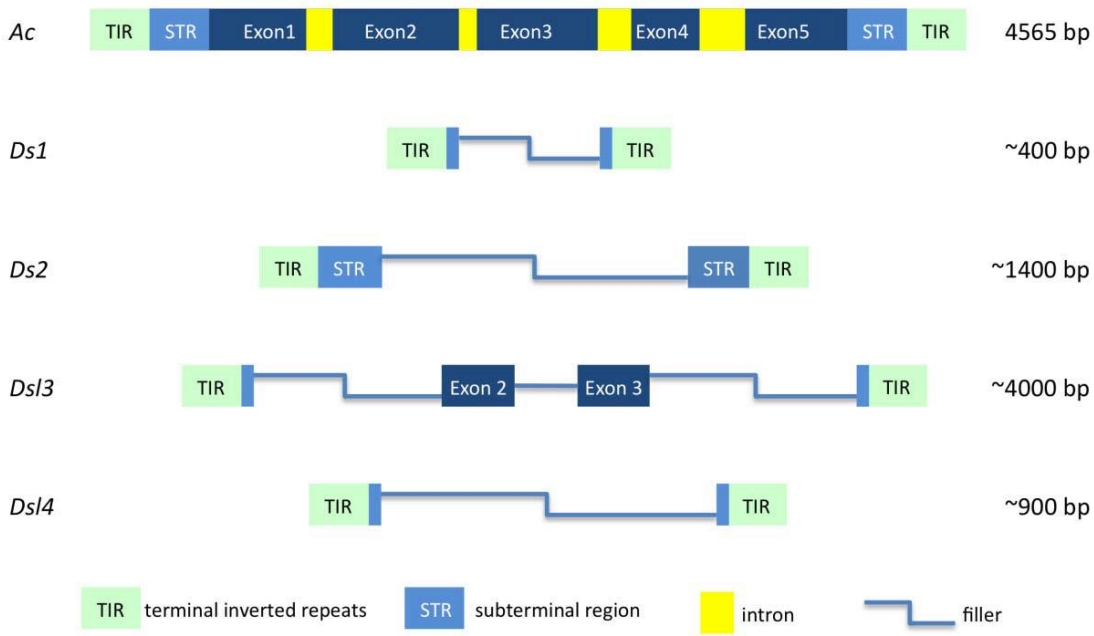
- ***Fully autonomous element of Ac/Ds system in maize***
  - 11 bp terminal inverted repeats (TIR)
  - subterminal repeats (STR)
  - ***5-exon, 807 amino acid transposase enzme***
    - ***Enzyme controls element movement***
  - Multiple hexameric repeats within 200 bp of each end
    - Site where transpose binds
- ***Causes 8bp direct repeat when inserted in new location***
  - Repeat sequence used as ***<u>evidence of element activity</u>***

Ds element

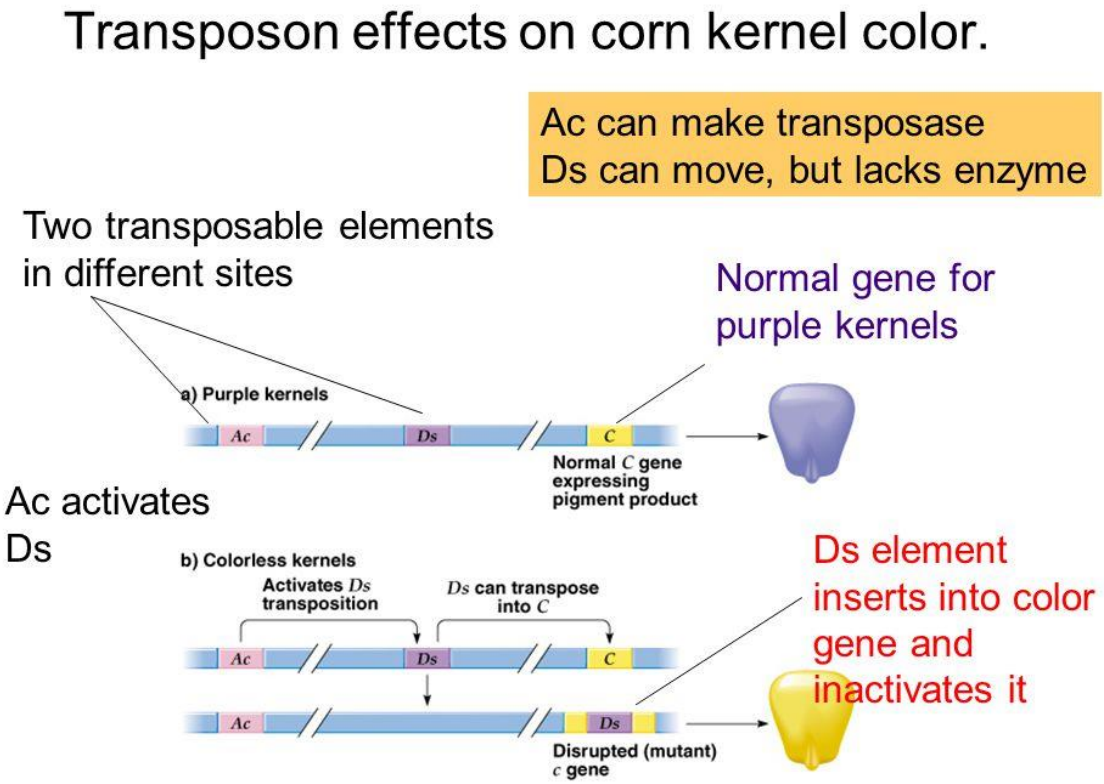
- ***Non-autonomous element of Ac/Ds system of maize***
  - ***<u>Truncated version of Ac</u>***
  - **Requires active *Ac* element to move**
  - **Multiple versions of *Ds* exist**
    - Each version has different components of the full *Ac* element

Structure of maize Ac/Ds elements

- **From Du et al. 2011, BMC Genomics 12:588**



Close-Up of Phenotypic Effects of Transposon Movement



From: https://slideplayer.com/slide/4735619/

- **Result of Ac or Ds movement during development**

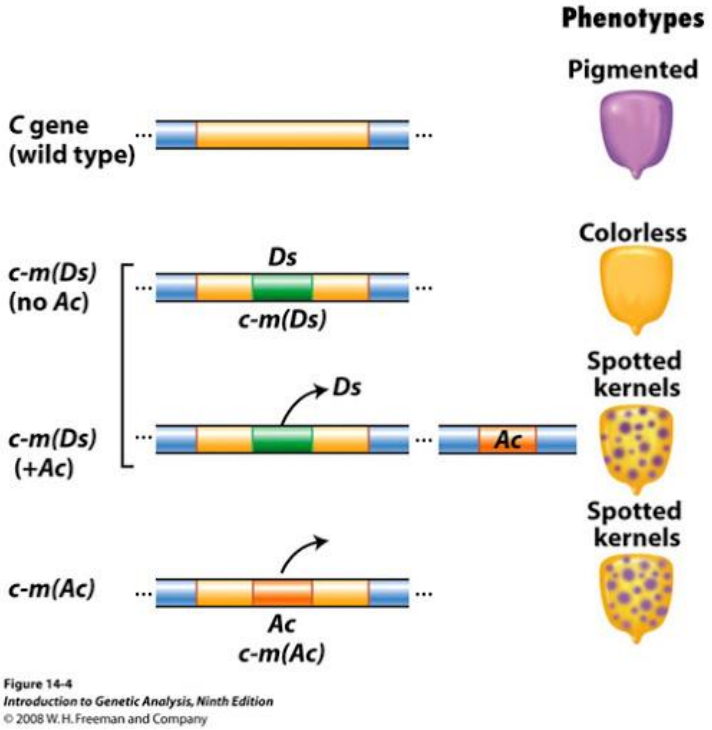


Figure 14-4
Introduction to Genetic Analysis, Ninth Edition
© 2008 W. H. Freeman and Company

Copy number in genetic stocks

- **Maize *Ac/Ds* system**
  - Autonomous element
    - Most lines do not contain an *Ac* element
    - ***Ac* active lines contain only one element**
  - Dissociation element
    - Maize B73 stock
      - 903 *Ds* elements

Transposition moves element from location A to B

- **Element moves from one location to another**
  - Element reconstituted at donor site by gene conversion
    - Or
  - Donor site ligated with lose of element

Transposition destination

- **Unique or low copy regions of the genome**

Element amplification

- Donor element replicated
- Element moves to unreplicated receptor site
- Receptor site replicated
  - Two elements become three
  - One donor site plus two receptor site
  - Donor site can be maintained

MITES: Subclass of DNA Elements

- ***M***inature ***I***nverted-repeat ***T***ransposable ***E***lements
  - Small, minimal **DNA elements**
    - Truncated versions of autonomous DNA elements
  - Structure
    - Nearly identical sequence
      - 400 bp
    - Contain terminal inverted repeats
      - **5' GGCCAGTCACAATGG…..**
        400nt…….CCATTGTGACTGGCC 3'
    - Direct repeats flank insertion sites
      - ***No open reading frames***
  - Location
    - Found in the arms of chromosomes
    - ***Associated with genes***
  - **Associated with genes**
    - 58% (23,623) of rice genes associated with MITES
      - ***MITES located in***
        - **Introns** or
        - **500 bp upstream/downstream of gene**
  - Relationship to small RNAs
    - **24% of rice small RNAs derived from MITE sequences**
  - Multiple families in rice
    - Defined by
      - TIR (terminal inverted repeat sequence)
      - TSD (target site duplication sequence)
  - **Copy number**
    - 178,533 copies in rice
      - **6% of the genome**

The Association of MITE and Rice Genes

- **Lu et al. 2012. Mol Biol Evol 29:1005**

MITE Superfamily	Total elements	Associated with genes	Expressed	Expressed with genes	Small RNAs
Tc1/Mariner	50,207	14,830	2,042	983	33,917
PIF/Harbinger	59,407	14,101	2,298	974	70,257
CACTA	3,859	739	134	58	7,380
hAT	15,299	4,341	737	280	15,395
Mutator	49,126	15,252	2,665	1,162	56,646
Micron	655	138	11	6	242
Total	**178,533**	**49,401**	**7,887**	**3,463**	**183,837**

MITES role in gene regulation

- MITES are a source of double-stranded RNA (dsRNA)
  - ***dsRNA template for small interfering RNA***
- siRNA biogenesis
  - ***siRNAs a component of RNA interference gene regulation***

Evolution of MITEs

- **Feschotte et al. 2002. Nature Reviews Genetics 3:329.**

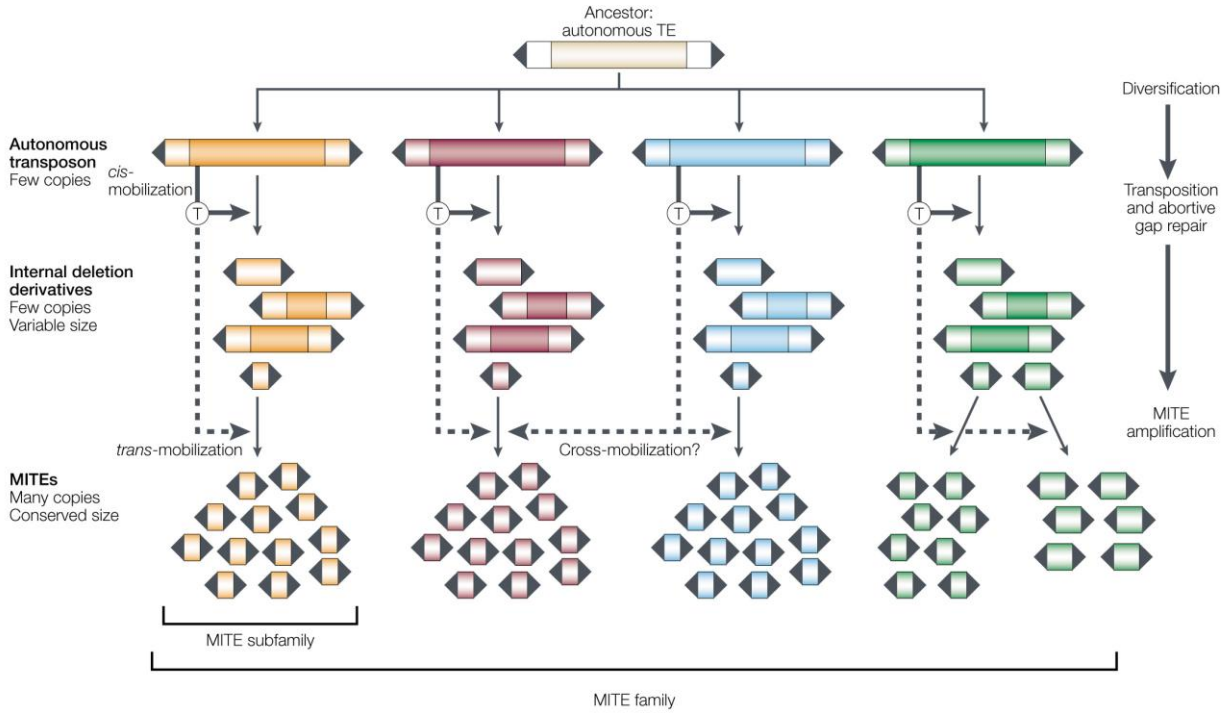


Figure 2 | **Model for the origin and amplification of MITEs.** In this model, the accumulation of miniature inverted-repeat transposable elements (MITEs) within a genome is explained by the activity of numerous related, but distinct, autonomous elements (shown in different colours). Related autonomous elements arise from a single ancestral element but have diversified to the point at which they only share sequence similarity in their terminal inverted repeats (TIRs; black triangles) and transposase gene (boxes, in darker colour). The activity of each element, mediated by its transposase (circled T), is proposed to form non-autonomous derivatives through mechanisms such as ABORTIVE GAP REPAIR43,44. The subsequent amplification of one or a few deletion derivatives gives rise to a group of homogeneous non-autonomous elements (that is, a MITE subfamily). This step is likely to be mediated by the same transposase or one that is produced by a close relative (‘*trans*-’ or ‘cross-mobilization’, respectively). See text and REF. 36 for further discussion of the model. TE, transposable element.

Retroelements (Class I elements)

General Features

- **Abundant in eukaryotes**
- ***All transpose via an RNA intermediate***
- **Major component of TEs in plants**
  - 70% of maize nuclear DNA
- Abundant in species with large genome sizes
- Related to LINEs

LINES: the terminal core of retroelements

- **L**ong **I**nterspersed **N**uclear **E**lements
  - **Ancient retrotransposons**
    - **DO NOT** contain a LTR (long terminal repeats)
- **Genes**
  - ***pol***
    - **Reverse transcriptase**
      - **Creates the RNA sequences that moves location**
        - **Only conserved gene among element types**
  - Other genes
    - Unique to different subclass
    - ***Int***
      - Most conversed of the other genes
      - **Involved in integrating the newly produced element elsewhere in the genome**

LTR Retrotransposons

- Contain flanking Long Terminal Repeat
  - **LTR – Long Terminal Repeat**
    - Varies in size
      - 100s of nt to 5 kilobases
- Two open reading frames
  - ***gag***
    - RNA packaging
  - ***pol***
    - **Polyprotein processed into**
      - **RT**
        - Reverse transcriptase
      - **INT**
        - Integrase (integrates circular RT product
      - **RNase H**
        - Removes RNA:DNA duplex following reverse transcription
      - **AP**
        - Aspartic proteinase
  - Other genes in ***some*** but not all elements
    - ***env***
      - Genome integration
    - ***prot***
      - Protease that cleaves the polyprotein

LTR Retrotransposon Classes

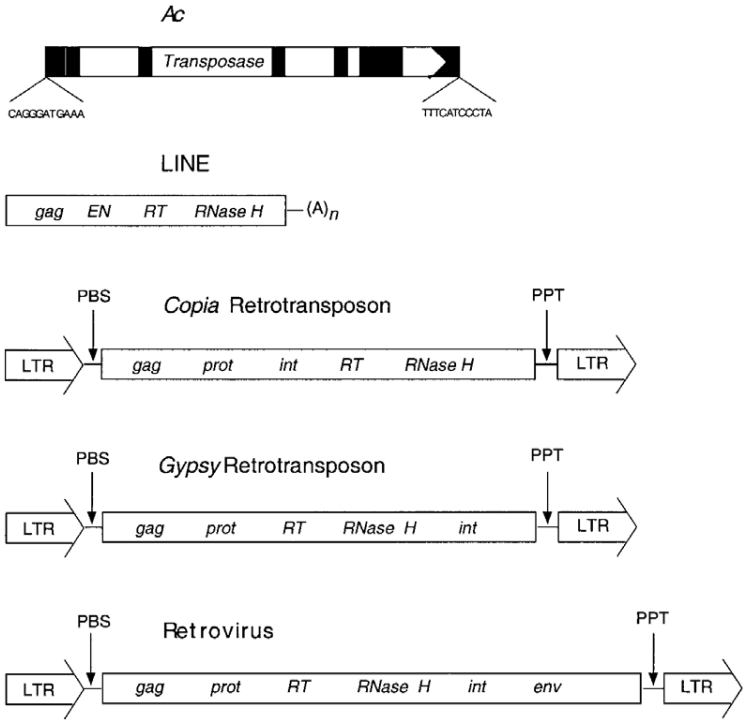
- **Defined by gene order**
  - **Ty1/Copia**
    - Gene order
      - **gag/int/RT/RNase H**
    - General location **but variation exists**
      - Gene rich regions
  - **Ty3/Gypsy**
    - Gene order
      - **gag/ RT/RNase H/int**
    - General location **but variation exists**
      - Pericentromeric and heterochromatic regions

Retroviruses

- Related to plant retrotransposons
- Same structure as Gypsy
- **Contain extra gene**
  - ***Env***
    - Envelope packaging of retrovirus particle

Structures of Retroelement Related Sequences

- **Bennetzen 2000. Plant Molecular Biology 40:251**



Sites required for transposition

- PBS
  - Primer binding site
- PPT

○ PolyPurine Tract

General comments about Class I elements

- **Not easily defined specifically**
  - Defined by sequence similarity to other elements
- **Ancient classes found**
  - Some ubiquitous in grass species
  - Ancient origin suggested
    - Recent elements resulted from expansion
- Maize: 2-6 mya
  - **Transpositon and amplification**
    1. **Element transcribed**
    2. **Pol makes DNA copy**
    3. **New DNA copy is integrated in new position**
  - **Copy number per class**
    - 100s to 10,000s
  - **Many elements inactivated by insertion of other retroelements**
    - May reduce potential mutational load of large numbers of active elements

SINEs

- Short Interspersed Nuclear Elements
  - Reduced in size from LINE elements
  - Lack LTR region
  - Rare in plants
  - Pol III-derived
    - Require trans-acting Pol and Int functions
- tRNA derivatives

Localization of TEs

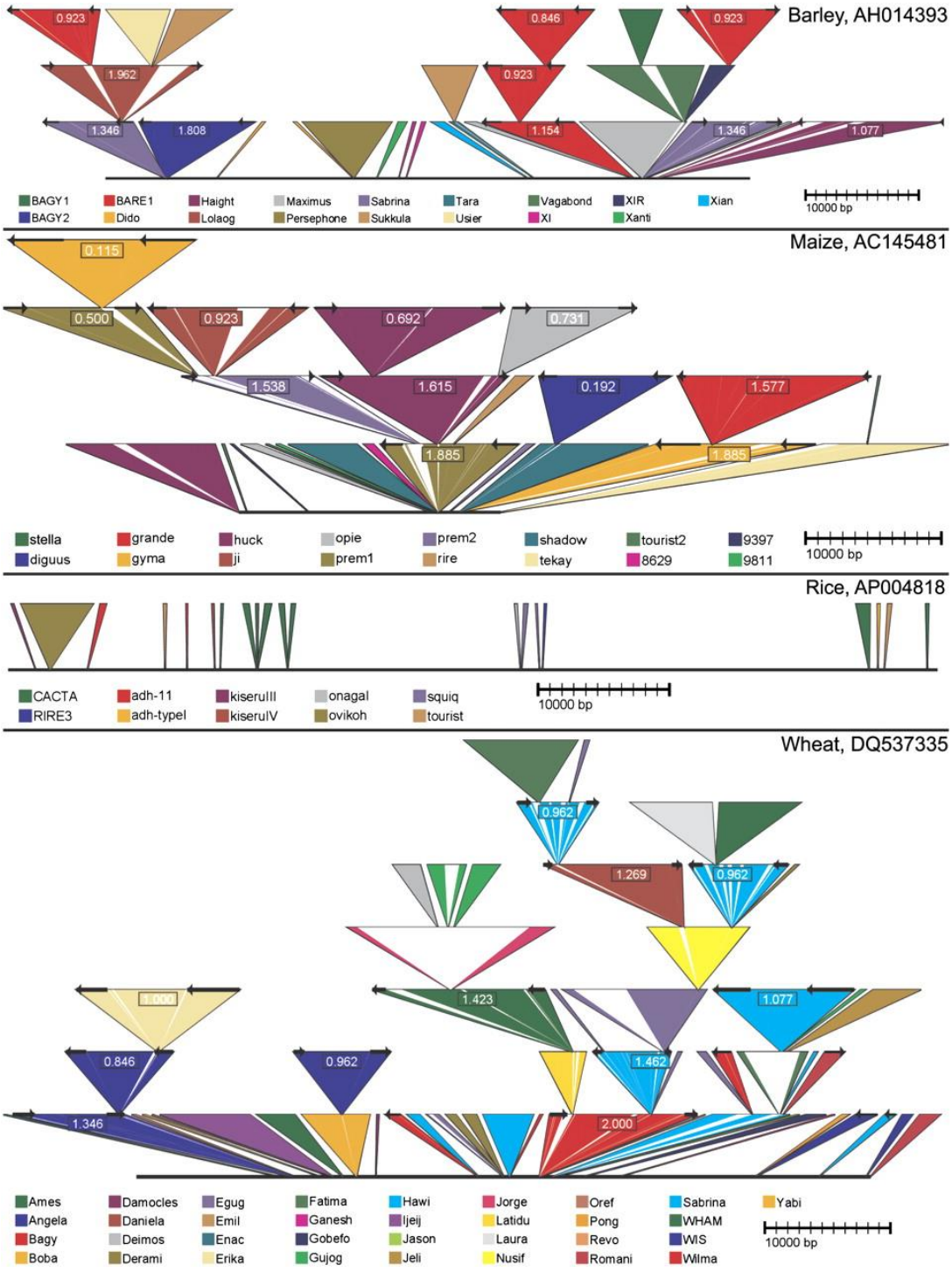
- **DNA elements**
  - **Genetically active genomic regions**
  - **MITES**
    - **5' and 3' regions of genes**
    - Near matrix attachment regions (MARs)
      - MARs
        - Insulate genes from neighboring DNA elements
        - Prevent spread of unmethylated region into genetically active region
        - Prevent genetic interactions with neighboring regions
- **Retroelements**
  - **Genetically inactive regions**
    - **IRP regions**
      - IRP: intergenic retrotransposon
        - Maize retroelements concentrated in IRP regions
    - **Centromeric heterochromatin**

Retrotransposons and maize genome evolution

- **Genome size increased due to whole genome duplication**
  - **Massive increase in retrotransposons in last 3 million years**
  - **Recent expansions**
    - 1.5 million to 500,000 years ago
  - 50% of maize genes have elements somewhere in gene
    - Promoter
    - Exonic region
    - Intron elements are very rare
  - **Historical footprints**
    - Excision events
      - Leave small repeats that can modify gene activity
  - Selection will maintain useful modifications

Nesting of TE in grass genomes

- **Kronmiller and Wise Plant Physiol. 2008;146:45-59**
  - **Elements integrate into other elements**



Unique Mobile Element Structure Among Eight Maize Lines

- **Wang and Dooner. 2006. PNAS 103:17644**

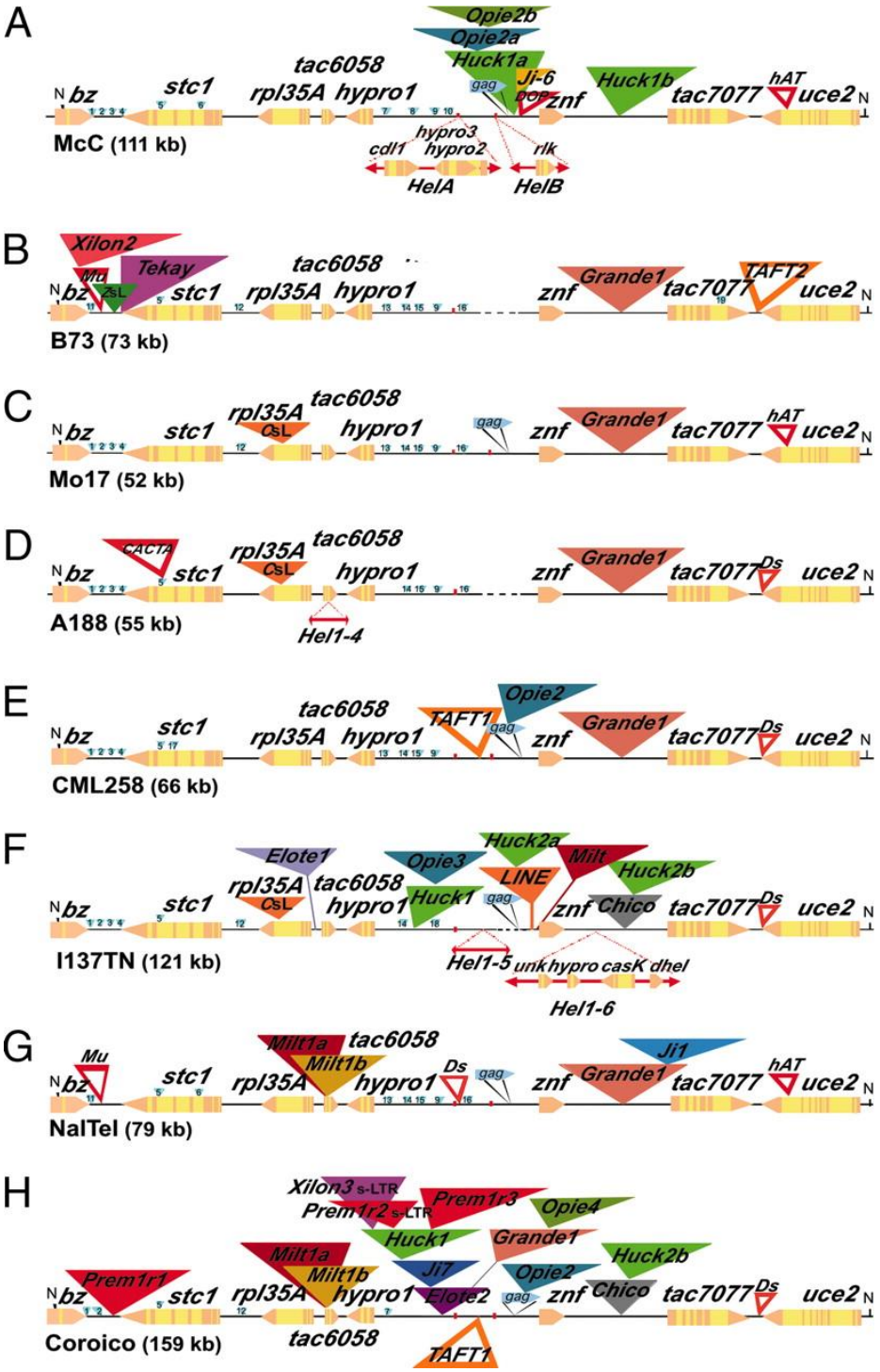


Fig. 1. Organization of eight *bz* haplotypes. Each haplotype is identified by the name of the genetic line, followed by the size of the cloned *Not*I fragment,in parentheses. The locations of the *Not*I sites at the proximal and distal ends are marked by Ns on the left and right, respectively. Genes are shown as pentagons pointing in the direction of transcription; exons are in bronze and introns in yellow. There are eight genes in the region: *bz*, *stc1*, *rpl35A*, *tac6058*, *hypro1*, *znf*, *tac7077*, and *uce2* (21). The same symbols are used for gene fragments carried by helitrons (Hels), which are represented as bidirectional arrows below the line for each haplotype. The vacant sites for HelA and HelBin each haplotype are provided as reference points and marked with short vertical strokes. Dashed lines represent deletions. **RETROTRANSPOSONS** are indicated by **SOLID TRIANGLES OF DIFFERENT COLORS**. **DNA TRANSPOSONS AND TAFTS**, which are probably also DNA transposons, are **INDICATED BY OPEN TRIANGLES IN RED AND ORANGE**, respectively**. SMALL INSERTIONS** are indicated in **LIGHT BLUE** and are numbered as indicated in Table 3. Only the genes have been drawn to scal

Maize Transposable Element Summary

- **LTR retrotransposons (Class I)**
  - ~1 million copies
    - **75% of total genomic DNA**
  - Located in both genic and non-genic regions
  - Elements nested inside other elements
  - 80% of elements are *Copia* or *Gypsy* elements
- **DNA elements (Class II)**
  - **8.6% of total genomic DNA**
    - Preferentially located in genic regions
- Comparison with other monocot genomes

Comparison of Repetitive Element Distribution Among Monocot Genomes

Species (size)		Class I (Retroelements)	Class II (DNA transposons)	Protein coding genes
Brachypodium (271 Mb)	Copies	50,419	29,630	25,532
	% genome	23 %	5%	37%
Rice (420 Mb)	Copies	61,900	163,800	29,717
	% genome	19 %	13%	29%
Sorghum (739 Mb)	Copies	216,519	76,883	27,640
	% genome	54%	8%	15%
Maize (2160 Mb)	Copies	1,139,990	142,800	32,540
	% genome	76%	9%	6%

Distribution of Repetitive Elements in *Phaseolus vulgaris* L.

Super families of TEs	Number of TEs ($X10^3$)	Coverage of TEs (bp)	Fraction of genome (%)
CLASS I	**281.3**	**185,960,175**	**39.36**
LTR retrotransposon	242.9	173,201,891	36.66
Ty3-gypsy	145.1	118,698,650	25.12
Ty1-copia	61.2	44,242,298	9.37
others	36.6	10,260,943	2.18
LINEs	37.5	12,599,869	2.67
SINEs	1.0	158,415	0.03
CLASS II	**87.1**	**25,979,571**	**5.50**
CACTA	43.9	12,726,168	2.69
Harbinger/PIF	0.5	264,755	0.06
hAT	3.9	1,028,733	0.22
Helitron	18.2	5,037,722	1.07
MULE	20.6	6,922,193	1.46
Unclassified TEs	**14.7**	**2,680,413**	**0.57**
Total	**383.2**	**21,4620,159**	**45.43%**

How Transposable Elements Can Modify Genes

Notes from

- **Vicient and Casacuberta (2017) Annals of Botany 120:195.**

Effects of Transposition

General Effects

- **Insertion of TE can change**
  - ***Protein sequence***
  - ***Expression pattern***
  - ***Generate new splicing variant***
- **TE insertion can introduce**
  - ***New promoter***
    - TEs have their own promoters
  - ***Enhancers***
- **TEs can move transcription factor binding sequences**
  - ***Create new regulatory network***
  - ***Mobilize a gene into a new network***

TEs affect neighboring genes via epigenetic effects

- **TEs are generally silenced by methylation**
  - Are restricted to site of insertion
  - Epigenetic mark (methylation) abundant in heterochromatic regions
- **TE insertion will introduce the epigenetic mark**
  - Neighboring genes will be silenced
  - Negative correlation in Arabidopsis between methylation and expression of neighboring genes
- **TEs are the source of other regulatory elements**
  - miRNAs in rice

Relationship Between TEs and Stress

- TE have stress-inducible promoters
- **Stress activates the movement of the TE**
  - **Insertion of TE near a gene will**
    - **Convert the gene to a stress-inducible gene**
      - *mPing* TE of rice
        - Creates stress inducible gene
      - *ONSEN* TE of Arabidopsis
        - Creates an ABA insensitive gene
  - ***Athila* TE of Arabidopsis**
    - Induces siRNAs (small interfering RNA)
      - siRNA regulates an RNA-binding protein

Examples of Transposable Element Insertional Mutants

- **Pea R shrunken seed locus**
  - Ac/Ds type transposon
    - **Starch branching enzyme**
- **Maize waxy gne**
  - Ac/Ds DNA transposon
    - **Starch granule-bound glucosyl transferase**
- **Rice glutinous kernel**
  - LTR retrotransposon
    - **Granular bound starch synthase**
- **Sorghum color gene Y**
  - CACTA DNA transposon
    - **MYB transcription factor**
- **Arabidopsis FAR1/FHY3 regulated phytochrome A response**
  - MULE DNA transposon
    - ***FAR1/FHY3* transcription factors**

Promoter Transposable Element Effects

- **Rice blast disease resistance**
  - LTR retrotransposon
    - **Element present = resistance; Element missing = susceptibility**
- **Grape branching**
  - hAT DNA transposon
    - **Activated enhancer = greater inflorescence branching**
- **Orange slice color**
  - *Copia*-like retrotransposon
    - **Cold-induced expression = darker slice color**

Maize Domestication and Transposable Elements

- **Studer et al. 2011. Nature Genetics 43:1160**

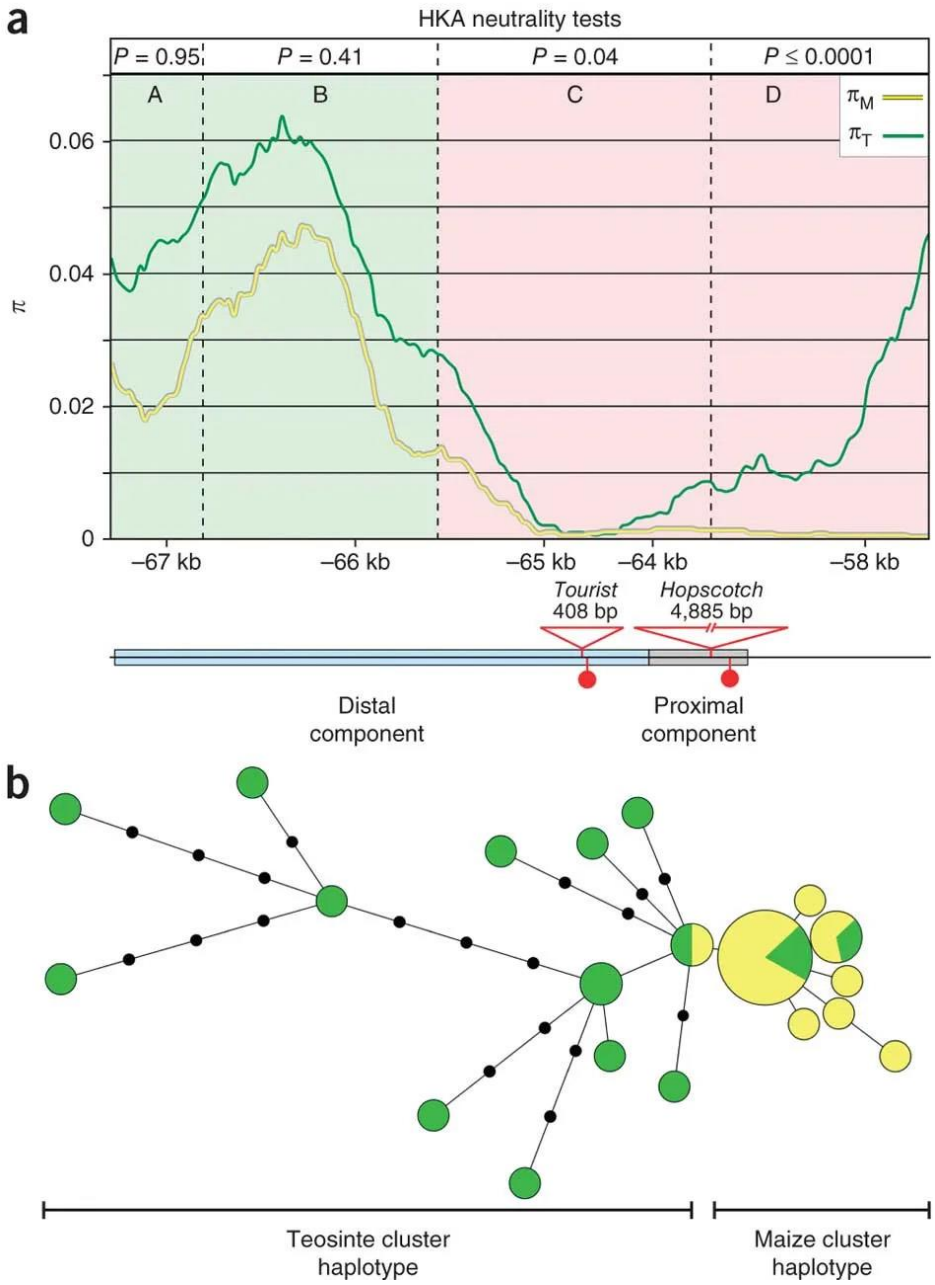


Figure 3: Sequence diversity in maize and teosinte across the control region. (**a**) Nucleotide diversity across the *tb1* upstream control region. Base-pair positions are relative to AGPv2 position 265,745,977 of the maize reference genome sequence. *P* values correspond to HKA neutrality tests for regions A–D, as defined by the dotted lines. Green shading signifies evidence of neutrality, and pink shading signifies regions of non-neutral evolution. Nucleotide diversity (π) for maize (yellow line) and teosinte (green line) were calculated using a 500-bp sliding window with a 25-bp step. The distal and proximal components of the control region with four fixed sequence differences between the most common maize haplotype and teosinte haplotype are shown below. (**b**) A minimum spanning tree for the control region with 16 diverse maize and 17 diverse teosinte sequences. Size of the circles for each haplotype group (yellow, maize; green, teosinte) is proportional to the number of individuals within that haplotype.

Summary of TE Element Effects in Plants

- **Wei and Cao 2016. Science China 59:24**

Table 1 Effect of TEs on phenotypic variation in plants[a)]

Regulatory mechanism	TE classification	Regulated gene	Plant phenotypes	References
Insertional mutagenesis	*Class II, *Ac/Ds*	*C*	Variation in pigmentation pattern in maize kernels	(McClintock, 1950)
	*Class I, LTR, *Dasheng*	*OsCHI*	Rice gold hull and internode (*gh*) mutants	(Hong et al., 2012)
	*Class II, *CACTA* superfamily, *Cs1*	*Y*	Variegated pericarp in sorghum grain	(Chopra et al., 1999)
	*Helitron	*BrTT8*	Yellow seed coat in *Brassica rapa*	(Li et al., 2012)
	*Class I and Class II	*CHS-D*	Flower color variation in morning glory	(Clegg and Durbin, 2000, 2003)
	*Class II, Mutator	*bHLH2*	Pale flowers and ivory seeds in *Ipomoea purpurea*	(Park et al., 2007)
	*Class I, *Gypsy*-type LTR, *Gret1*	*Vvmby1A*	Changes in grape skin color	(Kobayashi et al., 2004)
	Class II, MITE	*F3'5'H*	Changes in potato tuber skin color	(Momose et al., 2010)
	Class I, LTR, *dem1*	*MdPI*	Parthenocarpic production of apple fruit	(Yao et al., 2001)
	*Class II, *Ac/Ds*	*Wx*	Waxy kernels in maize	(Wessler et al., 1986)
	*Class II, *Ac/Ds*	*SBEI*	Wrinkled-seed character in peas	(Bhattacharyya et al., 1990)
	*Class I and Class II	*GBSS1*	Waxy and low-amylase types of foxtail millet	(Kawase et al., 2005)
	*Class I, LTR, *Dasheng*	*GBSS*	Glutinous rice seed in *Oragamochi*	(Hori et al., 2007)
	Class II, *hAT* family, *dTok0*	*FON1*	Multiple floral organs and numerous seeds in rice	(Moon et al., 2006)
	Class II, MuDR	*ZmGE2*	Increased embryo to endosperm ratio in maize	(Zhang et al., 2012)
Regulatory elements	*Class II, *Mu1*	*Hcf106*	White sectors on maize leaves	(Martienssen et al., 1990)
	Class II, *Harbinger*	*Pr*	Purple cauliflower	(Chiu et al., 2010)
	Class I, retrotransposon	*b1*	Maize seed color	(Selinger and Chandler, 2001)
	*Class I, LTR, *Renovator*	*Pit*	Blast resistance in rice	(Hayashi and Yoshida, 2009)
	*Class I, *Copia* family, *Hopscotch*	*tb1*	Increased apical dominance in maize	(Studer et al., 2011)
	*Class II, *hAT* family, *Hatvine1-rrm*	*VvTFL1A*	Somatic variation in cluster shape in grapevine	(Fernandez et al., 2010)
	Class II, MITE	*Vgt1*	Flowering time in maize	(Salvi et al., 2007)
	Class II, MITE	*AltSB*	Aluminum tolerance in sorghum	(Magalhaes et al., 2007)
	*Class II, MITE, *mPing*	Os01g0299700, Os02g0135500, Os02g0582900	Response to stress in rice	(Naito et al., 2009)
	*Class I, *Copia*-like	*Ruby*	The accumulation of anthocyanins in blood oranges	(Butelli et al., 2012)
	*Class I, LTR	*Wx*	Waxy kernels in maize	(Varagona et al., 1992)
	*Class I, *Copia*-like, *COPIA-R7*	*RPP7*	Pathogen responses	(Tsuchiya and Eulgem, 2013)
	*Class II, *CACTA* family	*F3H*	Flower color and seed coat in soybean	(Zabala and Vodkin, 2007; Zabala and Vodkin, 2005)
Rearrangement of gene structures	*Class II, *Ac*	*P-oo*	Orange pericarp and cob in maize	(Zhang et al., 2006)
	*Class I, *Copia*-like, *Rider*	*SUN*	Morphological variation of tomato fruit	(Xiao et al., 2008)
	*Class II, *hAT* family,*Tam3*	*nivea (niv)*	Petal color in *Antirrhinum*	(Coen et al., 1986; Uchiyama et al., 2013)
Domesticated transposase genes	*Class II, MULE	*FHY3*, *FAR1*	Response to light signaling in *Arabidopsis*	(Lin et al., 2007)
	*Class II, MULE	*MUSTANG*	Severe developmental defects in *Arabidopsis*	(Cowan et al., 2005; Joly-Lopez et al., 2012)
	*Class II, *hAT*-like TE	*DAYSLEEPER*	Essential for plant development in *Arabidopsis*	(Bundock and Hooykaas, 2005)
Epigenetic regulation	Class I, SINE	*FWA*	Late-flowering in *Arabidopsis*	(Fujimoto et al., 2008; Kinoshita et al., 2007)
	*Class II, MULE	*FLC*	Late-flowering in *Arabidopsis*	(Liu et al., 2004)
	*Class I, LINE	*BONSAI*	Severe dwarfing in *Arabidopsis*	(Saze and Kakutani, 2007; Saze et al., 2008)
	Class II, *hAT* family, *nDart1*	*OsClpP5*	Pale-yellow variegated leaves in rice seedlings	(Tsugane et al., 2006)
	*Class II, *hAT* family, *Gyno-hAT*	*CmWIP1*	Sex determination in melon	(Martin et al., 2009)
	*Class I, *Copia*-like, *SORE-1*	*GmphyA2*	Photoperiod insensitivity in soybean.	(Kanazawa et al., 2009)
	*Class II, *CACTA*	*ZmCCT*	Attenuated photoperiod sensitivity in maize	(Yang et al., 2013)
	*Class I, SINE	*VTE3*(1)	Vitamin E accumulation in tomato fruits	(Quadrana et al., 2014)
	*Class I, *Copia*-like, Sal-*T1*	*FAE1*	Erucic contents in yellow mustard (*Sinapis alba*)	(Zeng and Cheng, 2014)
Epigenetic regulation	*Class I, *Karma*	Eg*DEF1*	Mantled fruits in oil palm.	(Ong-Abdullah et al., 2015)
	*Class I, LTR, *Athila* family	*UBP1b*	Stress-sensitivity in *Arabidopsis*	(McCue et al., 2012)
	*Class II, MITE	*MAIF1*	ABA signaling and abiotic stress responses in rice	(Yan et al., 2011)
	*Class II, MITE, En/Spm-like	*CYP76M7, OsKSL7, CYP99A3, OsCPS4, EUI*	Plant height in rice	(Wei et al., 2014)
	*Class II, MITE	*OsGSR1*, *OsBR6ox*	Leaf angle in rice	(Wei et al., 2014)
	*Class II, MITE	*ZmNAC111*	Natural variation in maize drought tolerance	(Mao et al., 2015)
	*Class II, MITE	*RAV6*	Leaf Angle and Seed Size in Rice	(Zhang et al., 2015b)

TEs and Plant Genome Structure

Location

- **Heterochromatin**
  - ***Gypsy*-like retrotransposon**
- **Euchromatic**
  - ***Copia* retrotransposon**
  - **DNA (Class II) elements**

Why this pattern of element distribution?

- **Selection against deleterious mutations**
  - Results in gene-poor heterochromatin regions
- **Recombination**
  - Low rate in heterochromatic region
    - TE elements maintained in this region
- **Epigenetic silencing**
  - Maintains the heterochromatic function of this region
  - Functions
    - Centromeres can resist microtubule tension during cell division
    - Contribute to evolution of centromeres
    - Maintenance of replication origins

Recombination and Heterochromatic Regions

- Heterochromatic regions vary in size between species
    - Genes shuffled between heterochromatic and euchromatic regions during evolution of new species as size of TE clusters change
    - Recombination reduced in these regions
- Physically larger heterochromatic regions
    - Less recombination
        - Constrains evolution of genes in this region by recombination
- Ancestral history and gene distribution
    - **Species-specific genes**
        - Located in heterochromatic regions
    - **Older ancestral genes common among through lineage**
        - Found in euchromatic region

Retrogenes and Genome Evolution

- **Another mechanism of the expansion of the gene space**
- Requires action of **retrotranspons** on existing genes
  - Process called
    - **Retrotransposition**

Retrotransposition impacts genome evolution

- **Generate new variation in the species**
- ***Retrogenes = result of retrotransposition***
  - **Evolution**
    - **Cellular mRNA Duplicate generated by reverse transcription process of Class I retroelements**
    - **Copy integrated into the genome**
  - **Result**
    - **Original gene duplicated**
  - **Unique Signatures of new retrogene**
    1. Loss of introns
    2. Poly-A tail
    3. Direct repeats flanking the gene
- **Rice example** (Wang et al. 2006. Plant Cell 18:1791)
  - Minimum of 1,235 primary events in rice
  - 5,734 including tandem duplications
    - ***21% of the rice genome***
    - ***38% of duplicated genes derived by movement and duplication***
- **Retrogenes a major factor in the development of the gene space in rice and other grasses**

Example of New Gene Developed by Retrotransposition

- **Elrouby N, Bureau TE (2001) J Biol Chem276:41963**

New Maize Gene = Bz

- **Consists of:**
  - Three maize genes:
    - Bg, Xe, and Pma
  - Two LTR genes
    - gag, env
  - LTR direct repeats

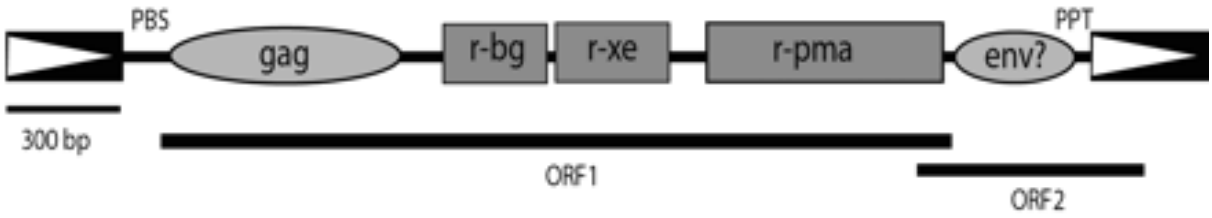


Figure 1. Schematic representation of the different domains. in *Bs1.* *Rectangles* containing *arrowheads* represent the LTRs LTRs. *PBS* and*PPT* are the characteristic retroviral primer-binding site and polypurine tract, respectively. The domains that make up most of the internal sequence are *gag*, a region that may correspond to an *env*-like domain, and the transduced sequences r-*bg*, r-*xe*, and r-*pma*. The *thick lines below* the *Bs1* structure indicate the position of open reading frames.

How TEs Can Drive Evolution of Function

- **Vicient and Casacuberta. 2017. Annals of Botany 120:195.**

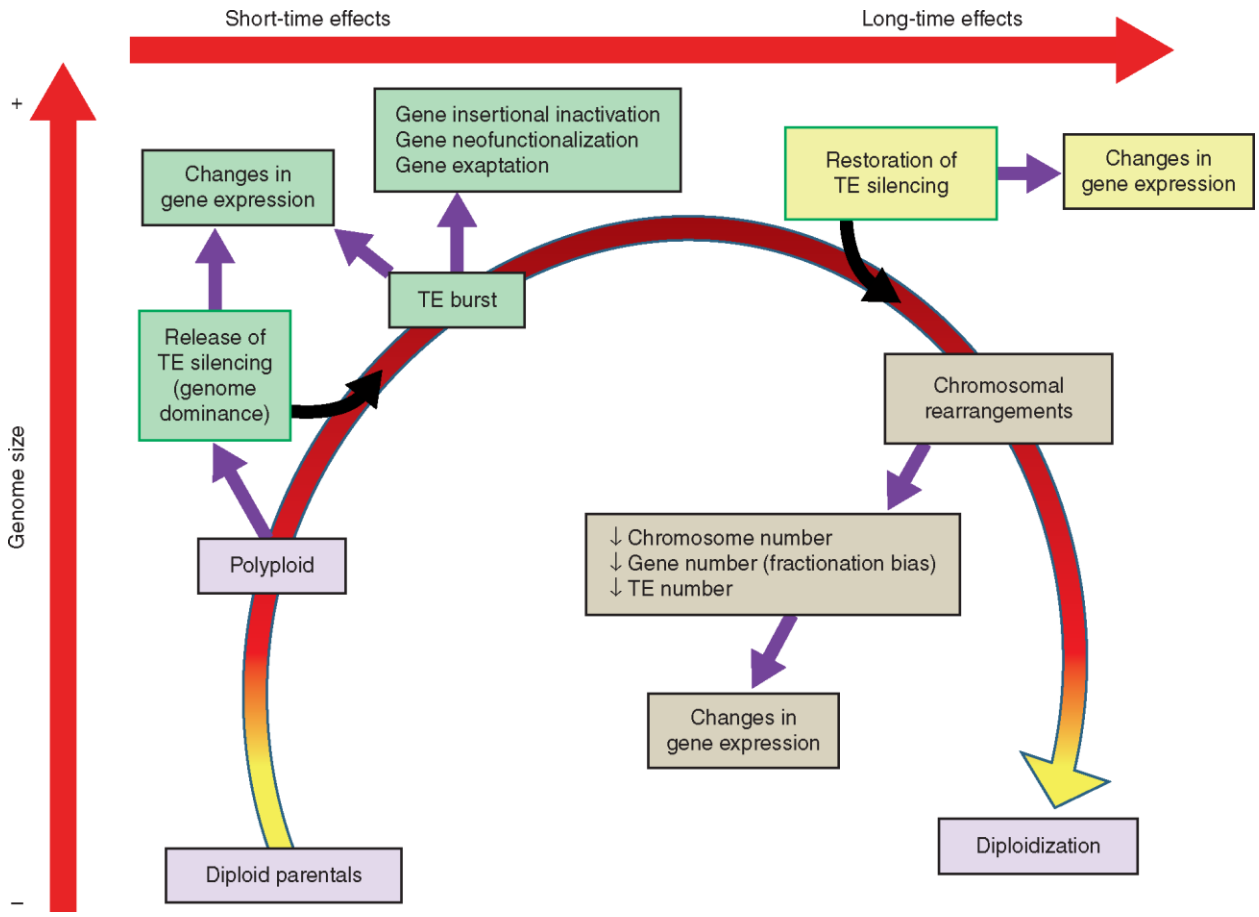


Figure 1. The close connections of polyploidization and TE dynamics. **POLYPLOIDIZATION** is accompanied by a **release of TE silencing**, which may be different for parentally or maternally inherited TEs. This release, in addition to **ACTIVATING TE MOBILIZATION**, may **induce changes in the regulation of genes located near TEs**. The **BURST OF TEs** will produce **new TE insertions that can modify the coding capacity of genes or their regulation**. The **RELEASE OF TE SILENCING** is **reversed after few generations**, and **TE SEQUENCES AGAIN BECOME** the **target of epigenetic silencing mechanisms**. The **SILENCING OF TEs**, including the new insertions resulting from the TE burst, will **influence the expression of genes located nearby**. This may result in **CHANGES OF GENE EXPRESSION** with respect to the **early phases of polyploidy** but also with respect to the diploid parents. **TEs WILL ALSO BE IMPORTANT FOR THE DIPLOIDIZATION** of the polyploid genome, as the different **TE copies may provide sequence homology for recombination**, leading to deletions and chromosome rearrangements.