

Plant Transposable Elements

(originally based on Bennetzen. 2000. Plant Molecular Biology 42:251; newer data)

Original transposable elements concept

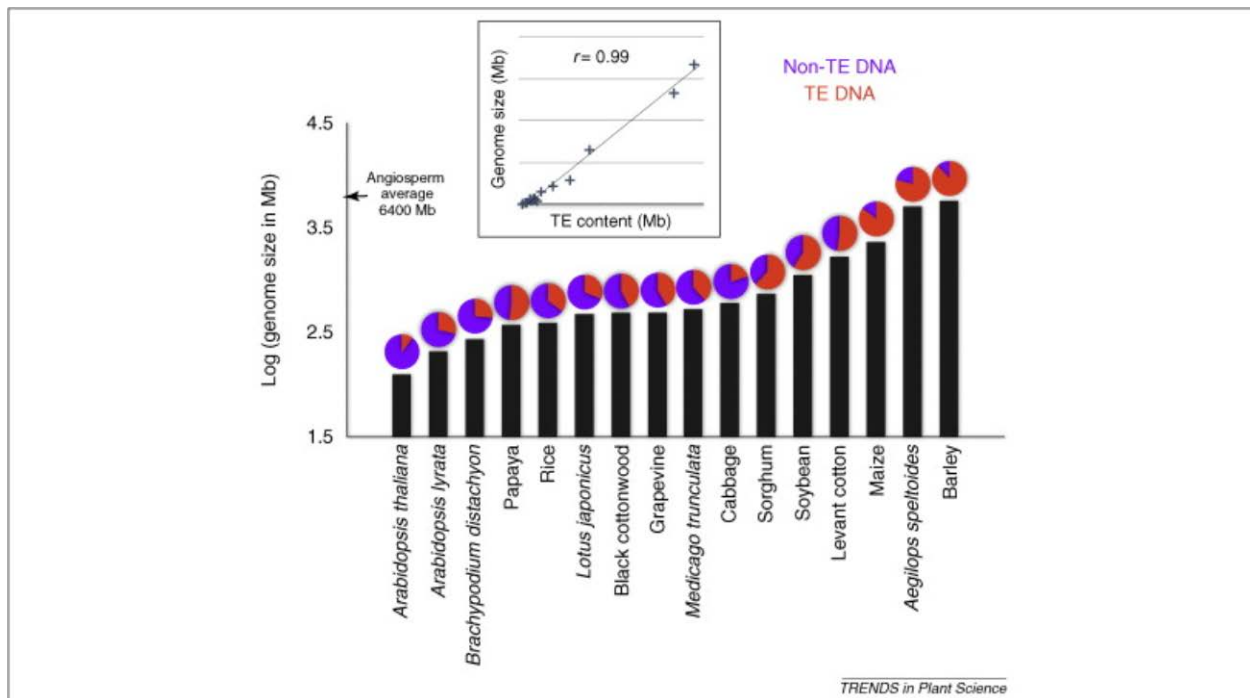
- Mutagenic agents that affected genes and the genome

Abundance of TEs

- They are found in all plant species
- Greater than 50% of some genomes

Relationship of Genome Size and TE Content

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



Result of transposition

- Change in gene structure and gene activity

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

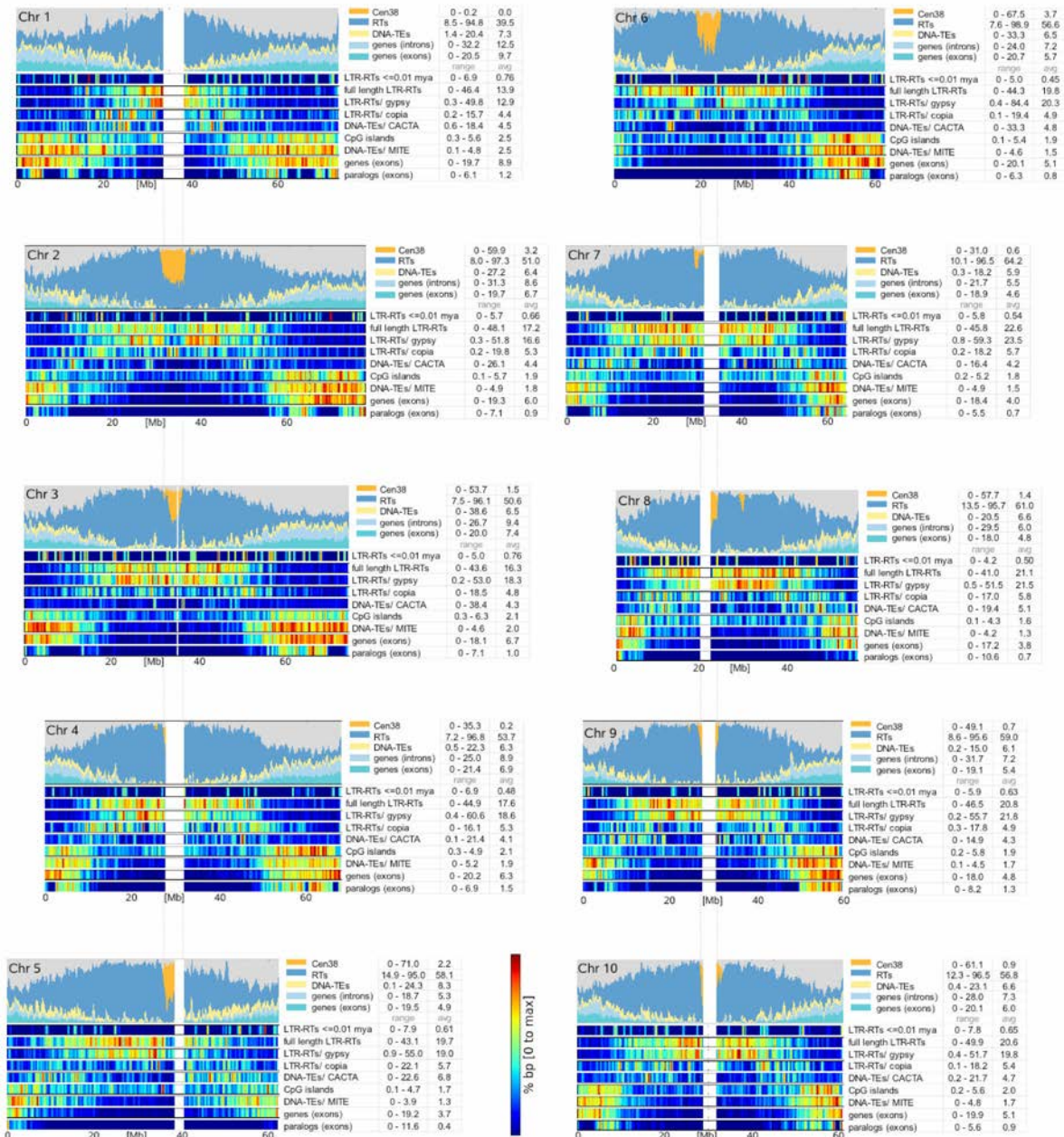
Types of TEs

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

Distributions of transposable elements in sorghum

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

- Many elements are clustered in the central heterochromatic region while others found more distally on 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

Can positively impact genomes

- Generate new variation
- Retrogenes
 - Genes derived by insertion of retroelement into a gene
 - Requires
 - loss of introns
 - recruitment of regulatory element
- **Rice example** (Plant Cell 18:1791)
 - Minimum of 1,235 primary events in rice
 - 5,734 including tandem duplications
 - *21% of the rice genome*

Genes derived from movement of retrotransposon elements

- Pea shrunken locus
- Regulators of Arabidopsis photoresponse
 - *FAR1* and *FHY3* transcription factors
 - Mediate phytochrome response

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

- Only element in bacteria
- Original elements found in plants
 - First described by McClintock in maize

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

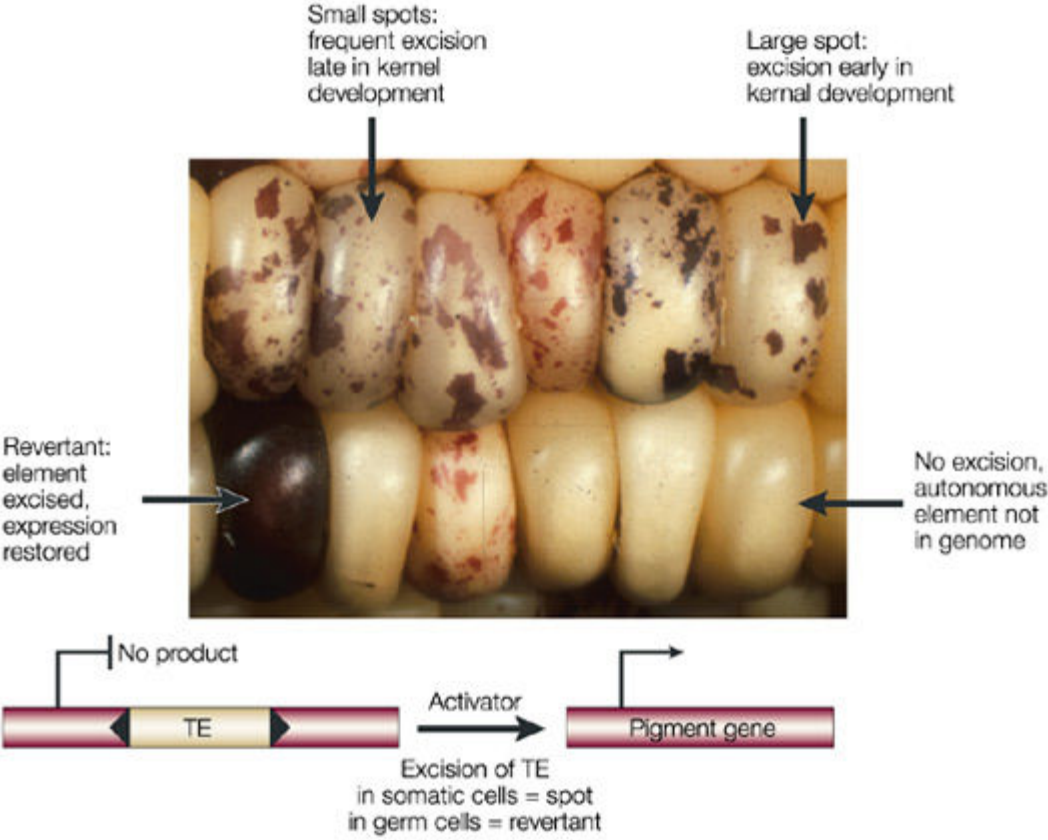
Autonomous vs. non-autonomous elements

- Autonomous element
 - Fully functional
 - Transpose independently
- 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



Ac/Ds maize transposable element system

Ac element

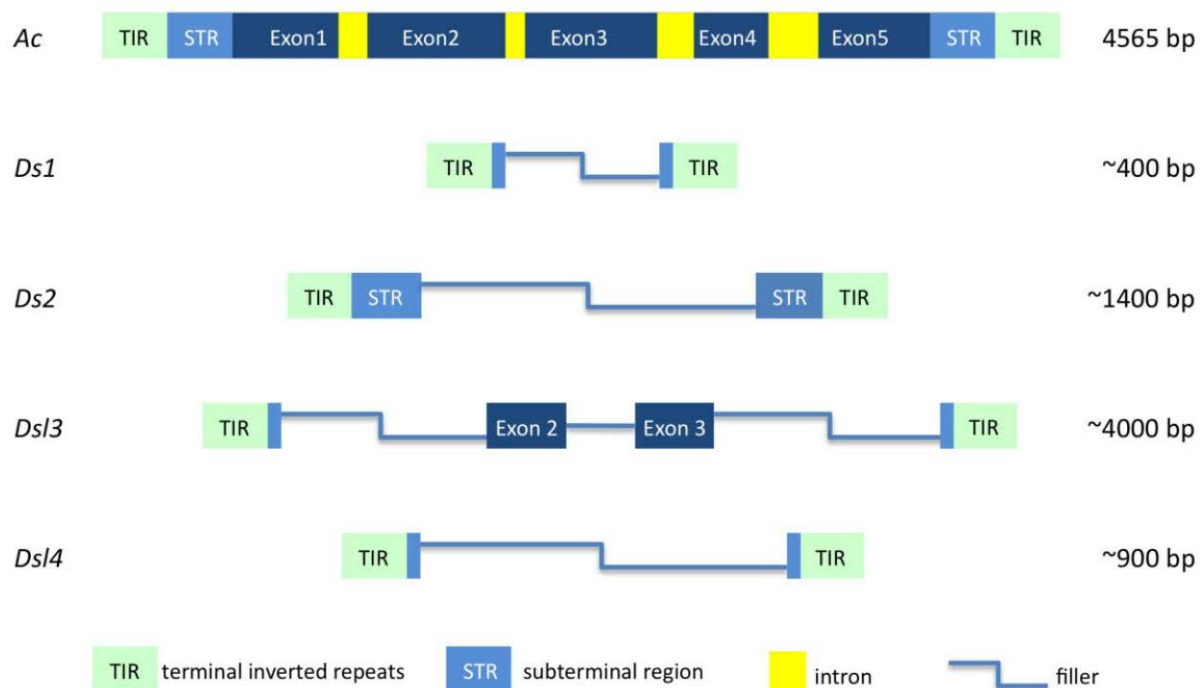
- fully autonomous element of Ac/Ds system in maize
 - 11 bp terminal inverted repeats (TIR)
 - subterminal repeats (STR)
 - 5-exon, 807 amino acid transposase
 - Multiple hexameric repeats within 200 bp of each end
 - Site where transposase binds
- causes 8bp direct repeat when inserted in new location
 - repeat sequence used as evidence of element activity

Ds element

- Truncated version of Ac
- Require active Ac element to move
- Multiple versions of Ds exist
 - Each version has different components of the full Ac element
- Located in repetitive regions of the genome

Structure of maize Ac/Ds elements

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



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

Cloning of TE elements

- Used heterologous probe for target gene
- Hybridized to stock mutated in target gene
- Sequenced the clone
 - Inserted DNA was the DNA element

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 maintain

MITES

- **Minature Inverted-repeat Transposable Elements**
 - 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 frame**
 - 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

(from: 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

Retroelements (Class I elements)

General Features

- Abundant in eukaryotes
- *All transpose via a 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

- Long Interspersed Nuclear Elements
 - Ancient retrotransposons
 - **DO NOT** contain a LTR (long terminal repeats)
- **Genes**
 - Pol
 - Reverse transcriptase
 - Only conserved gene
 - Other genes
 - Unique to different subclass
 - Int
 - Most conserved of the other genes
 - Involved in integrating the newly produced element elsewhere in the genome

LTR Retrotransposons

- Contain flanking Long Terminal Repeat
 - LTR
 - Varies in size
 - 100s of nt to 5 kilobases
- Two open reading frames
 - GAG
 - RNA packaging
 - POL
 - Protein 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
 - En
 - Genome integration
 - prot
 - Protease that cleaves the polyprotein

LTR Retrotransposon Classes

- Ty1/Copia
 - Gene order
 - gag/int/RT/RNase H
 - Location
 - Gene rich regions
- Ty3/Gypsy

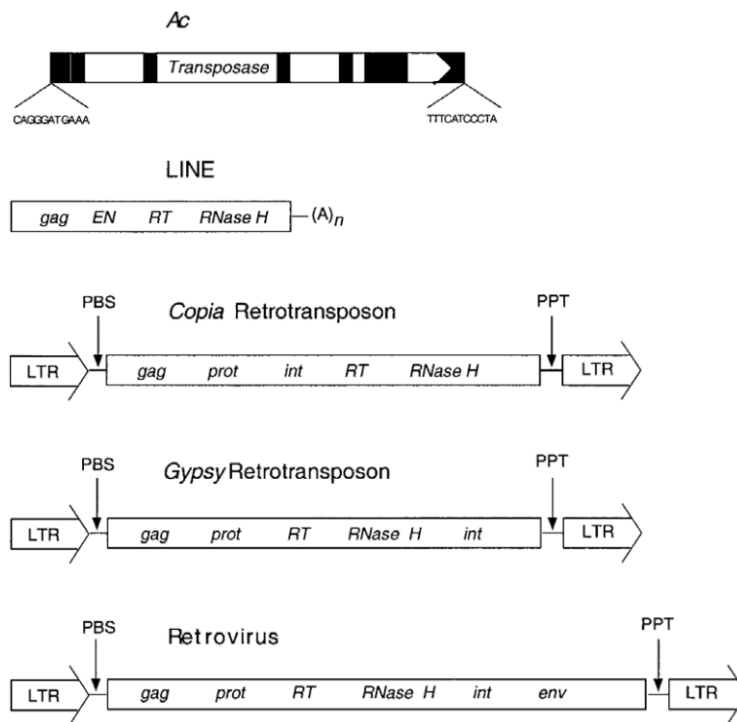
- Gene order
 - gag/ RT/RNase H/int
- Location
 - Pericentromeric and heterochromatic regions

Retroviruses

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

Structures of retroelement related sequences

from: 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
 - Transposon and amplification
 - Element replicated
 - Pol makes DNA copy
 - New DNA copy is integrated
 - Copy number
 - 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

Intronless mRNA pseudogenes

- Pol II generates mRNA that is spliced
- Trans-acting Pol and Int functions required
 - Create DNA copy (Pol)
 - Integrate the DNA copy (Int)

Localization of TEs

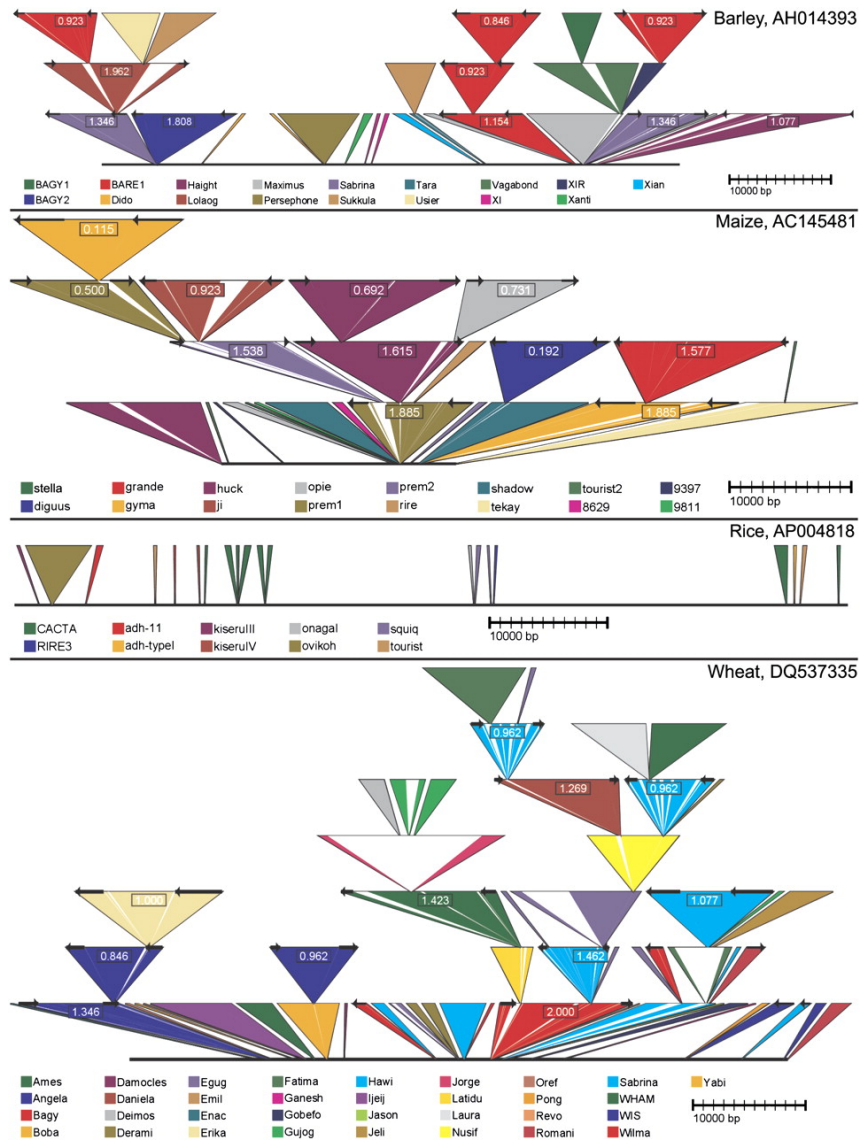
- **DNA elements**
 - Genetically active genomic regions
 - MITES
 - 5' and 3' regions of genes
 - Near matrix attachment regions (MARs)
 - MARs
 - Insulate genes from neighboring DNA elements
 - Prevent spread of unmethylated region into genetically active region
 - Prevent genetic interactions with 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

- Elements integrate into other elements



Kronmiller and Wise Plant Physiol. 2008;146:45-59

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	45.42

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%

How Transposable Elements Can Modify Genes

Notes from

Vicient and Casacuberta (2017) Impact of transposable elements on polyploid genomes. Annals of Botany 120:195.

Effects of Transposition

- 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 genes
 - *Athila* TE of Arabidopsis
 - Induces siRNAs (small interfering RNA)
 - siRNA regulates an RNA-binding protein

TEs and Plant Genome Structure

Location

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

Why this location distribution?

- Selection of 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
 - 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