

PLNT2530  
2024

Unit 5

# Genomes: Organization and Comparisons

Unless otherwise cited or referenced, all content of this presentation is licensed under the Creative Commons License Attribution Share-Alike 2.5 Canada



# Genome

- the sum of all genes and intergenic DNA on all the chromosomes of a cell represents the cellular genome.
- Plants possess a **plastid**, a **mitochondrial**, and a **nuclear** genome while animals have only the latter two.
- Like other eukaryotes, plants have linear chromosomes, each containing hundreds or thousands of genes.
- Nuclear chromosomes consist primarily of **non-coding, repetitive DNA**. Genes coding for proteins make up only a small percentage of genomes of most higher plant species.

Organism	Chromosome number (n)	Genome size (bp)	Gene No.
<i>Chlamydomonas reinhardtii</i> (green alga)	17	$1.11 \times 10^8$	17,742
<i>Arabidopsis thaliana</i>	5	$1.19 \times 10^8$	27,549
<i>Fragaria vesca</i>	7	$2.80 \times 10^8$	25,050
<i>Brassica rapa</i>	10	$2.84 \times 10^8$	41,174
<i>Oryza sativa</i>	12	$4.66 \times 10^8$	58,000
<i>Glycine max</i>	20	$1.10 \times 10^9$	46,430
<i>Zea mays</i>	10	$2.80 \times 10^9$	63,000
<i>Triticum aestivum</i>	21	$1.70 \times 10^{10}$	41,910

# Genomes

## Dependence and independence of cellular genome components

- separate transcription and translation machinery for each organelle
- nuclear genome possess overall control

## Relative size (average)

	relative size*
plastid genome	1
mitochondrial genome	3
nuclear genome	30,000

*\*bp/haploid genome*

## C-value paradox

1C value refers to the haploid genome size

bacteria < fungi < plants/animals

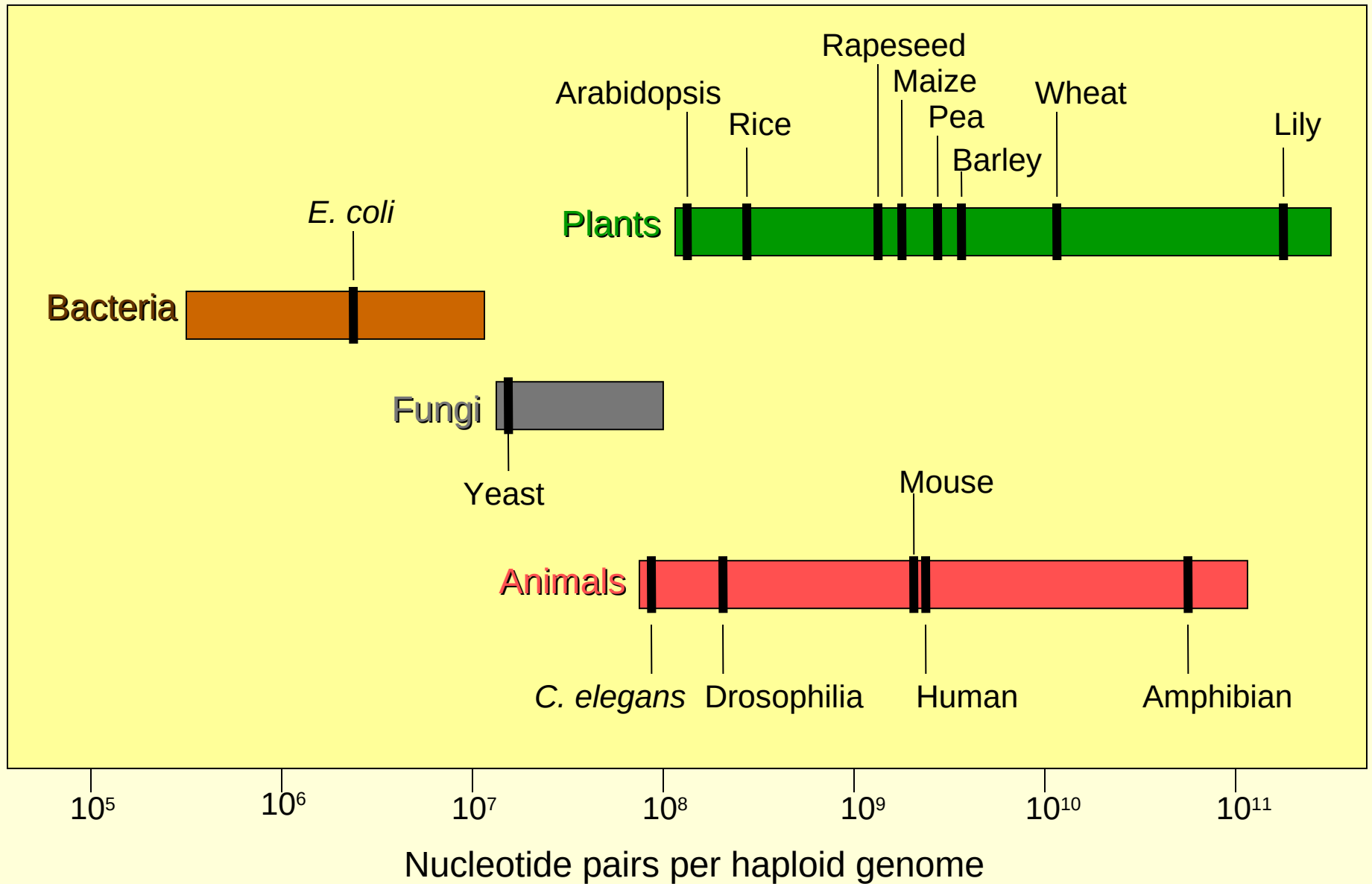
The paradox: Organism's complexity does not follow genome size completely.

But Arabidopsis is not 100 times less complex than wheat and 1000 times less than lily

	<u>Genome size</u>	<u>Gene#</u>
E. coli	$\sim 4 \times 10^6$ bp	$\sim 4,000$ genes
Arabidopsis	$1.20 \times 10^8$ bp	$\sim 25,400$ genes

Gene # estimates for higher plants 24,000 - 60,000 genes

# Genome size comparison among organisms




# What portion of the genome is made up by genes?

Assume the average gene size = 3000 bp (promoter, coding region but not introns)

$$\frac{3000 \text{ bp} \times 25,400 \text{ genes}}{1.20 \times 10^8 \text{ bp}} \times 100 = 61\% \quad \text{Arabidopsis}$$

$$\frac{4000 \text{ bp} \times 35,000 \text{ genes}}{4.5 \times 10^9 \text{ bp}} \times 100 = 3.1\% \quad \text{Pea}$$

pea genome size 

# What accounts for the extra DNA among different species?

1. Repetitive sequences (coding and non-coding)

2. Polyploidy

-Autopolyploid (all chromosome sets are homologous)

Eg. Sugar beet autotriploid (AAA)

-Allopolyploid (subgenome sets of chromosomes arising from 2 or more species or genera)

Eg. canola *Brassica napus* (n=19) allotetraploid (AACCC)

*B. oleracea* (n=9) CC

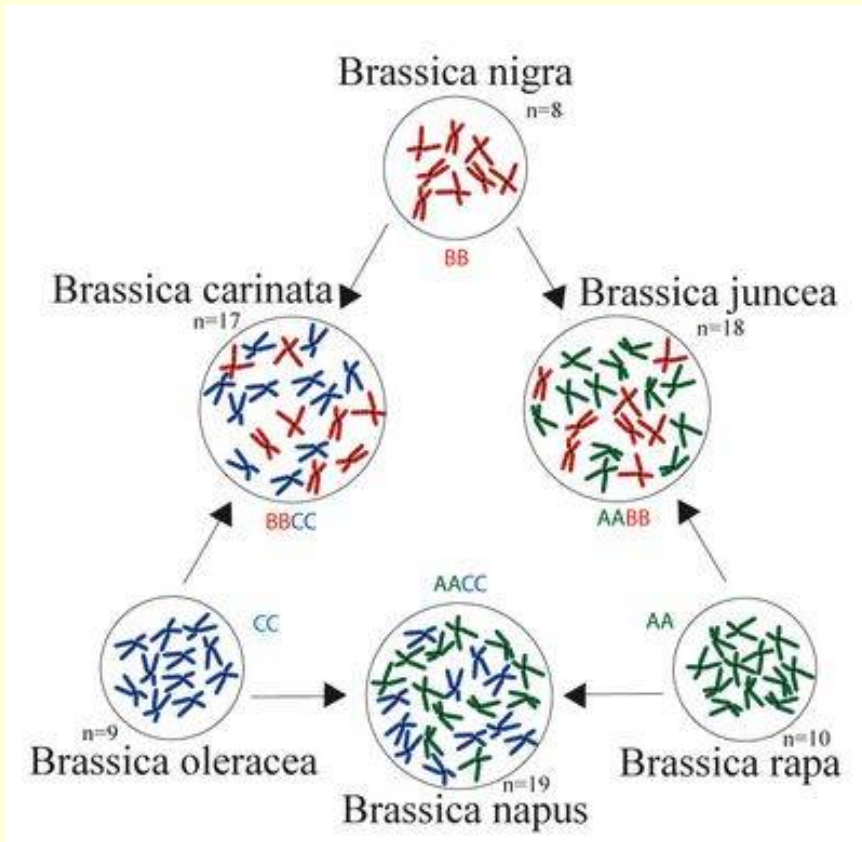
*B. rapa* (n=10) AA

(wheat, oat, tobacco)

3. Different size and frequency of introns in genes

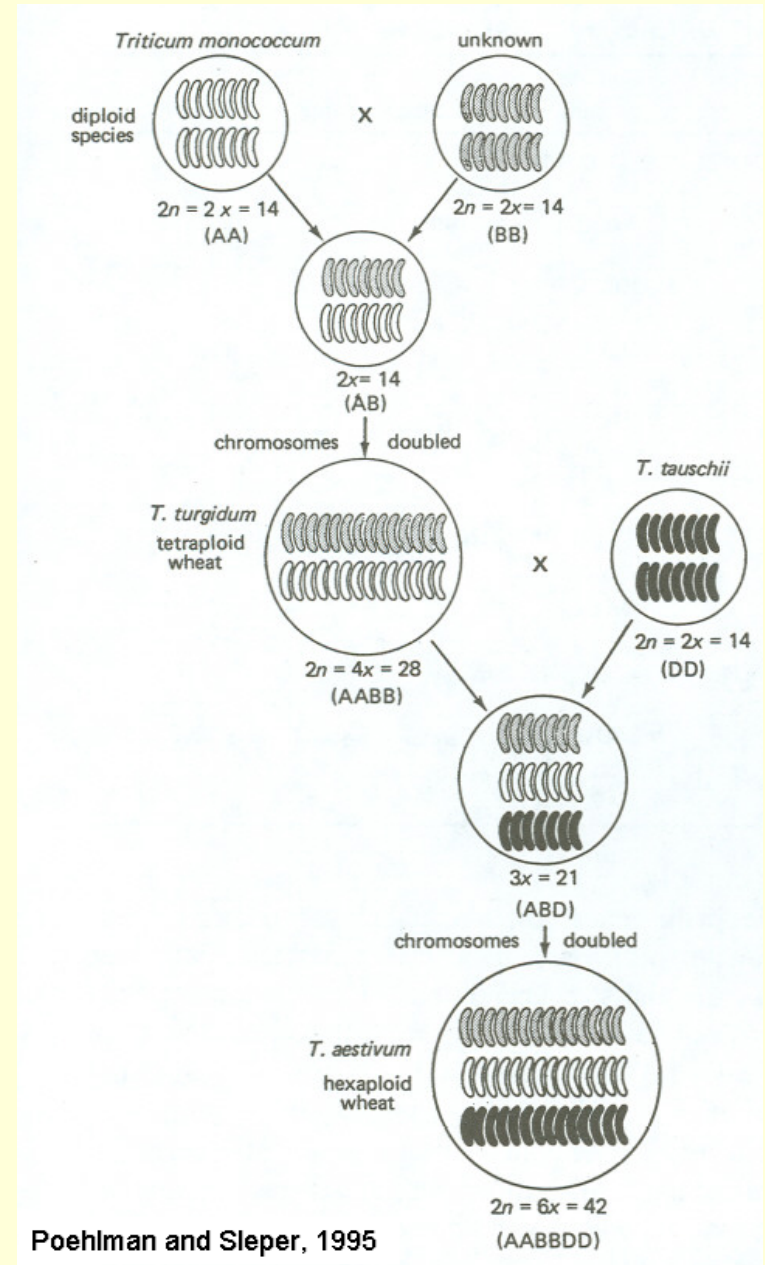


# Canola/rapeseed



Overview of species relationships in the genus 'Brassica'. Original work by [Mike Jones](#), for Wikipedia.

# Hexaploid wheat



Poehlman and Sleper, 1995

# Major Classes of Genomic DNA

Genic and non-genic DNA

## A. Genic DNA (<5% of total genome in large genomes)

### 1. “Single” copy genes

- between 1 - 10 copies per haploid genome
- represents the majority of genes

### 2. Duplicated and divergent genes

- functional multi-gene families
- may occur at a single or several loci on different chromosomes
- in Arabidopsis (fully sequenced) [996 gene families ] and [8,331 genes]

Specific examples: Calcium dependent protein kinase (1 family - 34 genes)

Cytochrome P450 (69 families with 256 genes)

wheat seed storage proteins (gliadens) (40-60)

# Major Classes of Genomic DNA

## A. Genic DNA (cont'd)

### 3. Tandemly repeated genes (non-divergent)

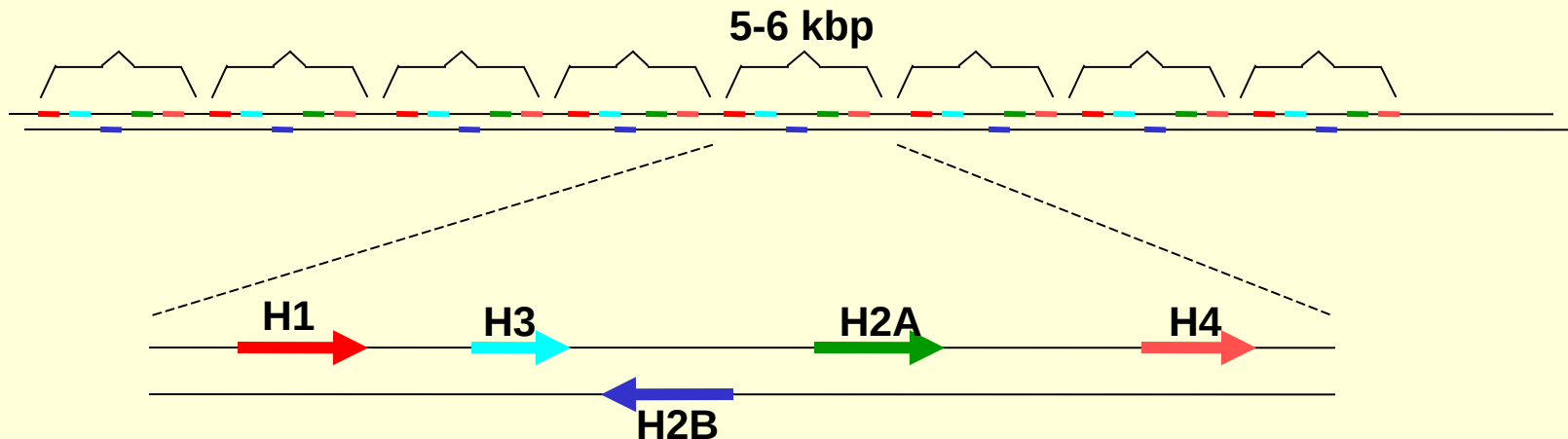
#### a) Histones

5 histone genes are clustered in a 5-6 kb unit

Unit is tandemly repeated



Most multicellular organisms (50-500 tandem repeats/copies)

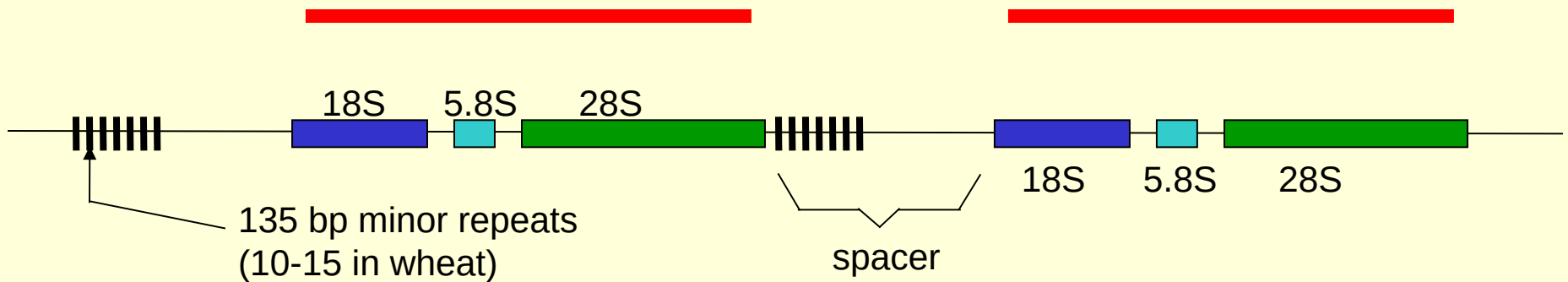
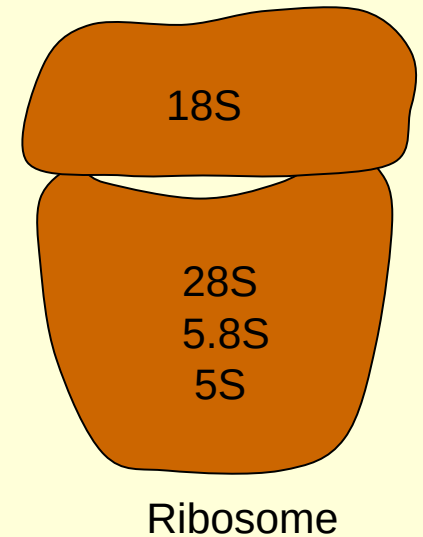


# Tandemly Repeated DNA (cont'd)

## b) Pre-rRNA genes

Repeat unit includes the 18S, 5.8S and 28S RNA coding regions plus a non-coding spacer region

Repeat unit 8-11 kb depending on species



## b) Pre-rRNA genes (cont'd)

- Copies / haploid genome

- Arabidopsis                    570 copies
- Most animals                100-200 copies
  
- Hexaploid wheat            1A 1000 copies  
    1B 2600  
    6B 6250

## c) 5S-rRNA genes

100-2000 copies/haploid genome

2 size variants in wheat 410 bp and 500 bp

in flax    340    &    360 bp

# Impact of repetitive genic elements

**'Single' copy genes** 30,000 genes x 3000 bp (avg)/gene =  $9 \times 10^7$

## Repetitive genes

Histones 500 copies x 6000 bp/repeat =  $3 \times 10^6$

Wheat pre rRNA 9850 copies x 9000 bp/repeat =  $8.8 \times 10^7$

5S RNA 1000 copies x 500 bp/repeat =  $0.5 \times 10^6$

# Major Classes of Genomic DNA

## A. Non-genic sequence (>95% of genome)

### 1. Tandem repeats

simple sequence repeats ("satellite" sequences)

3-10 bp repeats repeated  $10^4 - 10^6$  times  $[\text{TTTAGGG}]_n$

(Arabidopsis) GAA (wheat) common at telomeres

centromere repeats are often multiple elements of short repeats

### 2. Dispersed repeats

Many families of repeats spread throughout the chromosomes

eg 150-300+ bases repeated  $10^3 - 10^4$  times throughout genome

eg Alu element (humans) (~300b)

Afa family of elements (wheat, barley, oat + related species

(340 b)

# Major Classes of Genomic DNA

Major source of dispersed repeats are transposable elements:

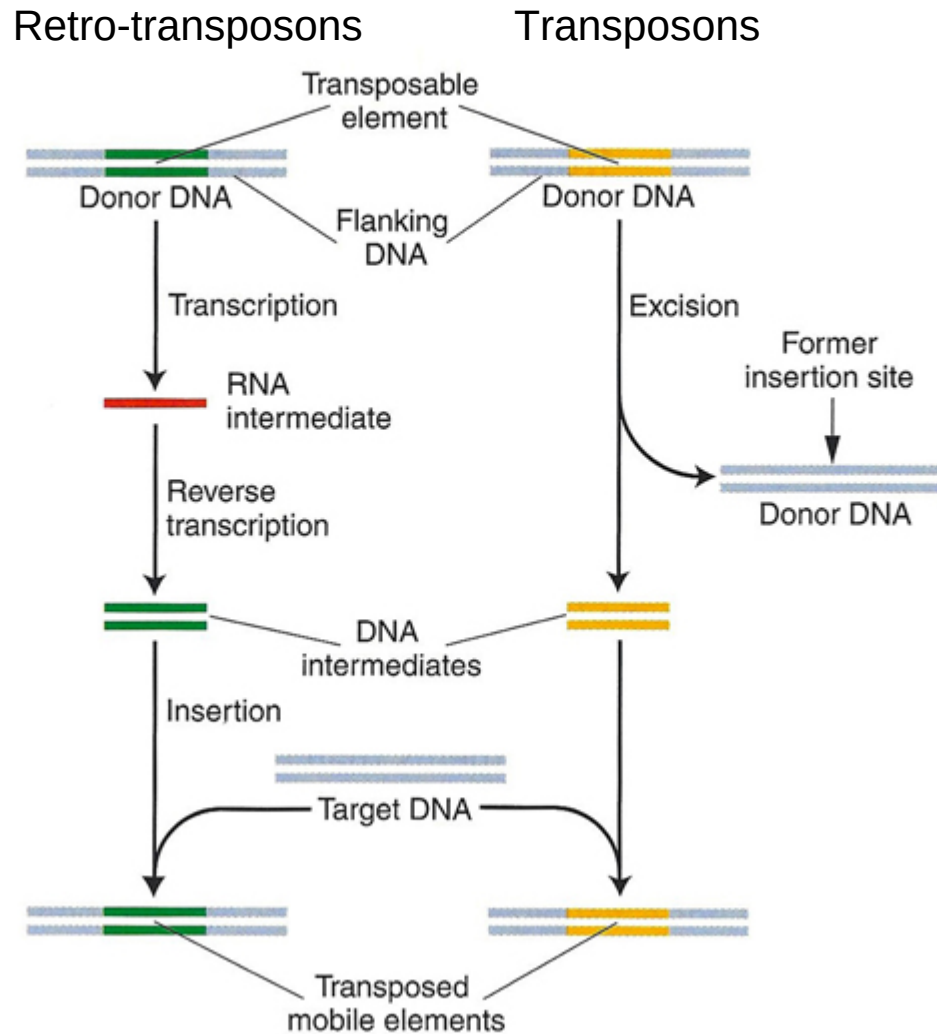
Transposable elements - sequences of DNA that can move around to different positions within the genome of a single cell (mobile genetic elements) (- Nobel Prize 1983 Barbara McClintock 1940s)

2 major classes

- **transposons** - replicate and move as DNA elements
- **retrotransposons** - (originally viral sequences) that replicate via an RNA intermediate method and are copied back into DNA by a reverse transcriptase



# Two classes of mobile elements



# Retrotransposons

- Particularly abundant in plants
- Maize 49-78% of genome composed of retrotransposons
- Wheat (90% repetitive DNA, 68% transposable elements)

Common elements (many are also found in plants, fungi and protists)

- LTR retrotransposons (with long terminal repeats 100bp - 5 kbp)
  - Ty1-copia retrotransposons
  - Ty3-gypsy retrotransposons
- Non-LTR retrotransposons (up to 250,000 copies in plant species)
  - LINEs long interspersed nuclear elements
  - SINEs short interspersed nuclear elements
- Human 42% of genome made up of retrotransposons
  - while DNA transposons account for about 2-3%

# Maize genome 2 x 10<sup>9</sup> bp

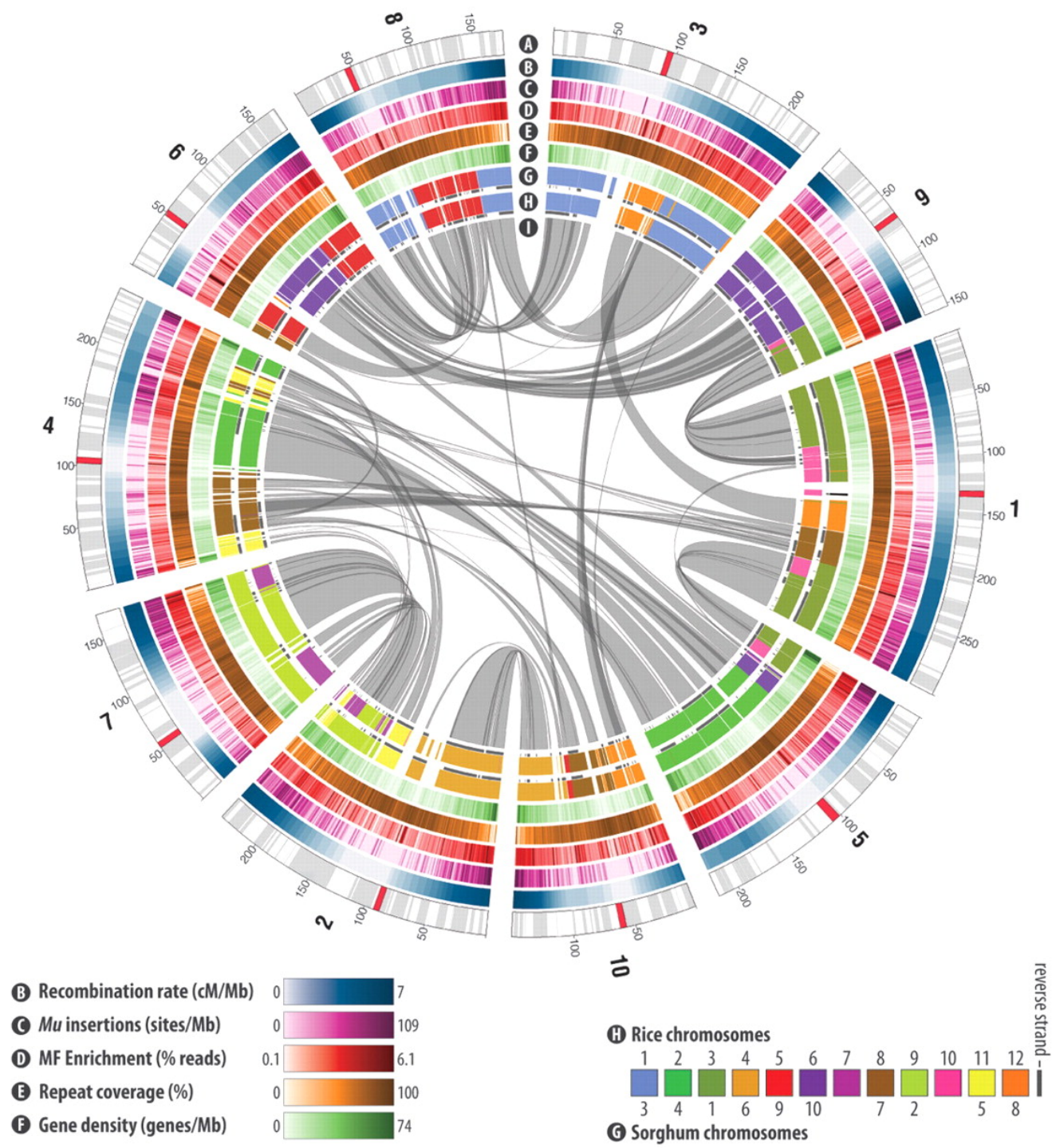
from Schnable PS et al. 2009

Science 326:1112-1115.

Main take home  
lesson from C,E  
and F:

Genes are  
islands in a sea  
of repetitive  
sequences.

E- repeats from previously  
characterized repeats,  
found by RepeatMasker  
or other programs.



# Non-genic DNA is not necessarily non-functional

## Functions of some repeat regions:

Scaffolding (Matrix) attachment regions ( $\sim 10^4$  in large genome)

Spindle fiber attachment sites (centromere)

Protection of chromosome termini (telomeres)

Origins of replication (in yeast  $\sim 250$  bp long)

Estimate of the # of origins in the human genome:

$\sim 50,000$  (Genome Res. 2013. 23: 1-11)

# Plastid Genome

All plant cell contain plastids of some type - same genome  
(chloroplasts, amyloplast, chromoplast, elaioplast) all from etioplast

Replicate by division. When a cell divides plastid copies are shared

In angiosperms (all common crop plants) during fertilization both plastid and mitochondrial genomes come from the ovary never from pollen.

Basis of maternal inheritance of certain traits.

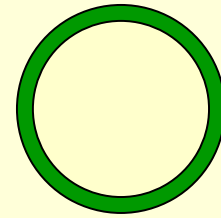
Significant consideration in some genetic engineering as it is a way of preventing transgene gene transfer by pollen flow.

# Plastid Genome

Similar in most plant species

120-270 kb per plastid genome

Circular dsDNA



Encodes 120-140 genes -mostly involved in protein synthesis or photosynthesis

Majority of proteins in the chloroplast are nuclear-encoded

Typical leaf 10-20 chloroplasts

5-10 copies of the plastid genome/chloroplast

Thus 50-200 copies of the genome/leaf cell

—

# Plastid Genome

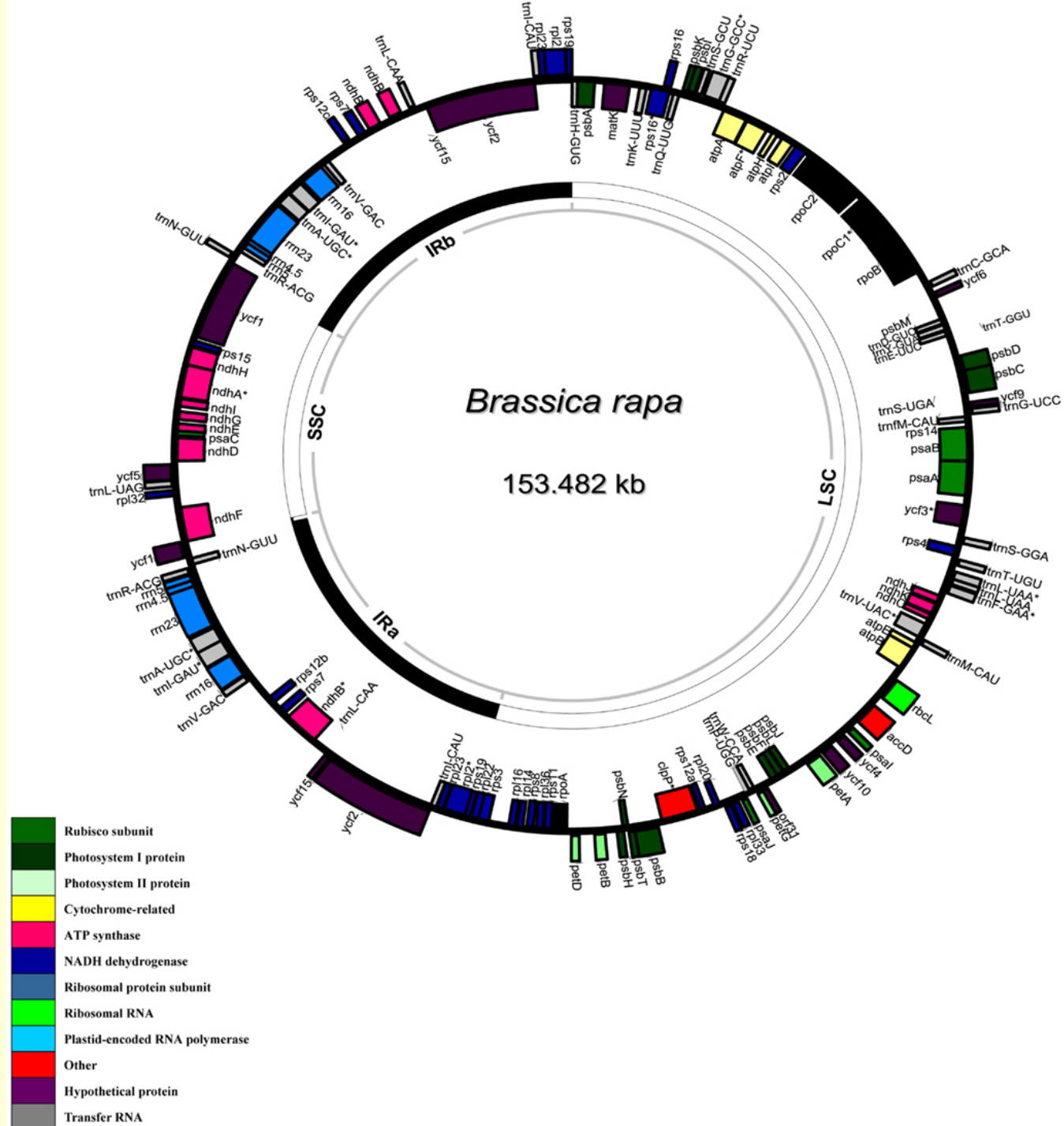
Genes transcribed on the forward strand are on the outer circle. Those transcribed on the reverse strand are on the inner circle.

Regions found in plastid genomes of higher plants:

LSC - large single-copy region

SSC - small single-copy region

IRa, IRb - inverted repeats; duplicate copies of genes are seen in inverted orientation



# Mitochondrial genome

Plant mitochondrial genome range 200-2500 kbp

Fewer genes than the chloroplast genome <100

Many mitochondria/per cell

Individual mitochondria divide and multiple organelles can re-fuse

Multiple copies of the genome/per mitochondrion

Genome exists as a series of subgenomic\* circles of dsDNA molecules which can interchange parts of the circles by recombination (unlike mammalian and fungal mitochondria)

The majority of proteins in the mitochondria are nuclear-encoded

subgenomic\* - portions of the full mitochondrial genome sometimes exist as separate circular molecules within the mitochondrion.



# Mitochondrial genome

circle 1: Large repeats containing blocks of genes with related functions, R1 - R4.

circles 2 and 3: Transcribed regions on forward and reverse strands.

circle 4: GC content

