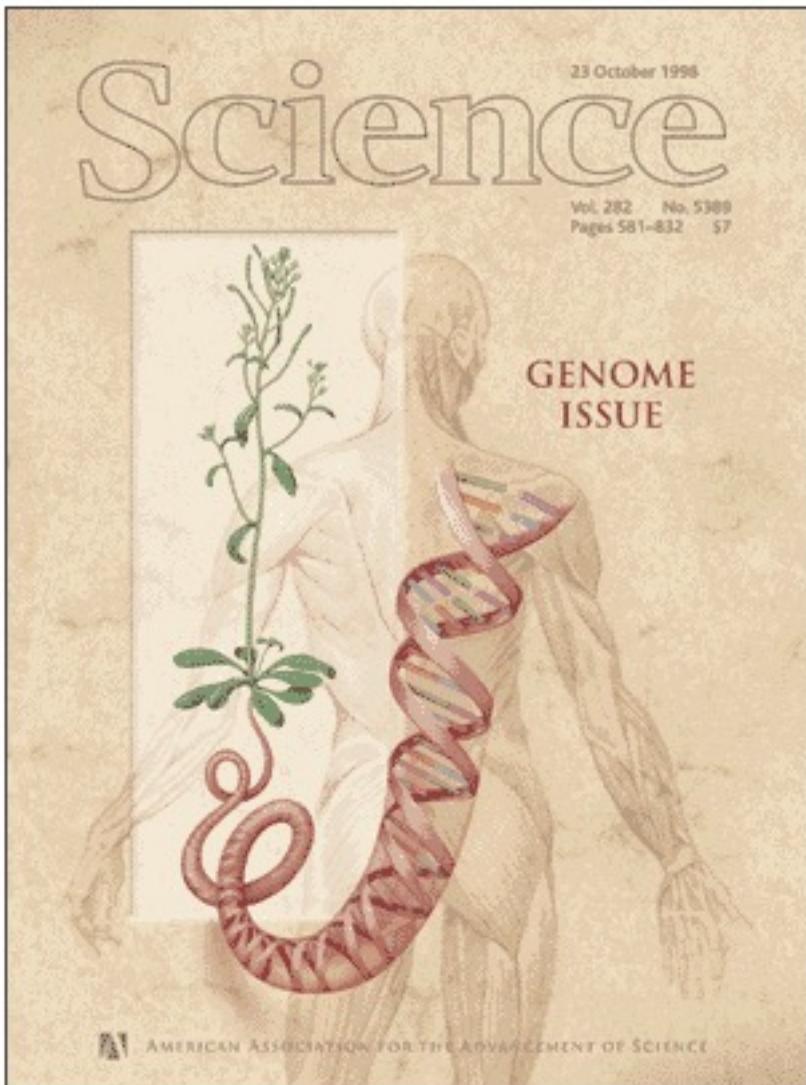
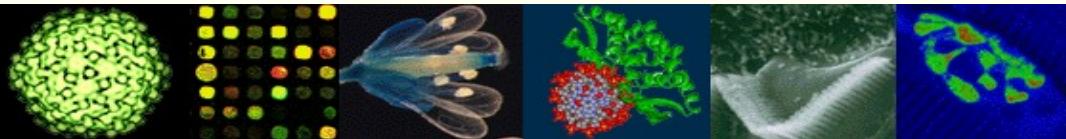


## Genomics (GENE3370)



# Evolution of Genomes I & II

**Prof Martha Ludwig**

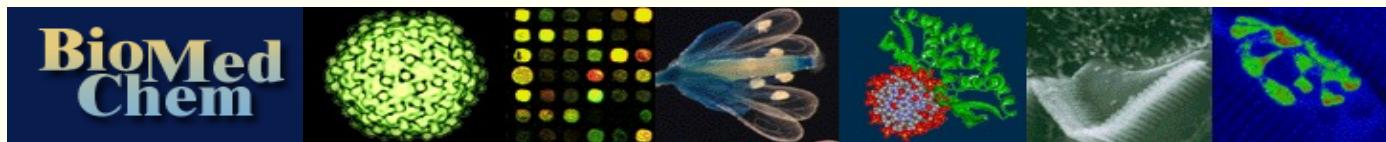


## Evolution of Genomes I & II – Learning Outcomes

At the end of these lectures, students should be able to describe and / or explain:

- know the major factors affecting nuclear genome evolution.
- know the types of gene duplication events that have occurred over evolutionary time.
- know the fates of duplicated genes.
- be able to describe how multigene families arise and know the difference between paralogues and orthologues.
- be able to explain the effects of domain shuffling on genome evolution.
- know the endosymbiotic theory of the origin of the nuclear, chloroplast and mitochondrial genomes, including the proposed ancestral genomes of each.
- be able to explain the product specificity corollary of the endosymbiotic theory, and examples supporting/refuting it.
- be able to explain the current hypotheses for why many endosymbiont genes were transferred to the host cell nucleus and why some have been retained in the organelles (including examples supporting and refuting each).
- be able to explain NUMTs and NUPTs and how they are believed to remodel nuclear genomes.
- be able to describe how DNA is thought to be transferred between intracellular compartments.
- be able to explain the mechanism that facilitates the integration of organelle DNA into the nuclear genome.
- be able to define horizontal gene transfer and give an example.

After these lectures students should also be able to apply and analyse concepts presented in the lectures.



## Evolution of Genomes I & II

### Suggested Reading

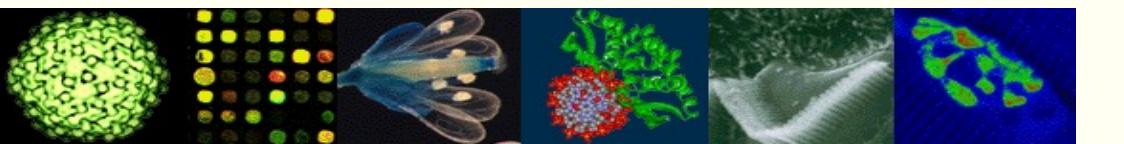
Brown TA (2018) Chapter 18 (relevant parts) *Genomes 4*, Garland Science

Shapiro JA (2014) Constraint and opportunity in genome innovation. *RNA Biology* 11: 186

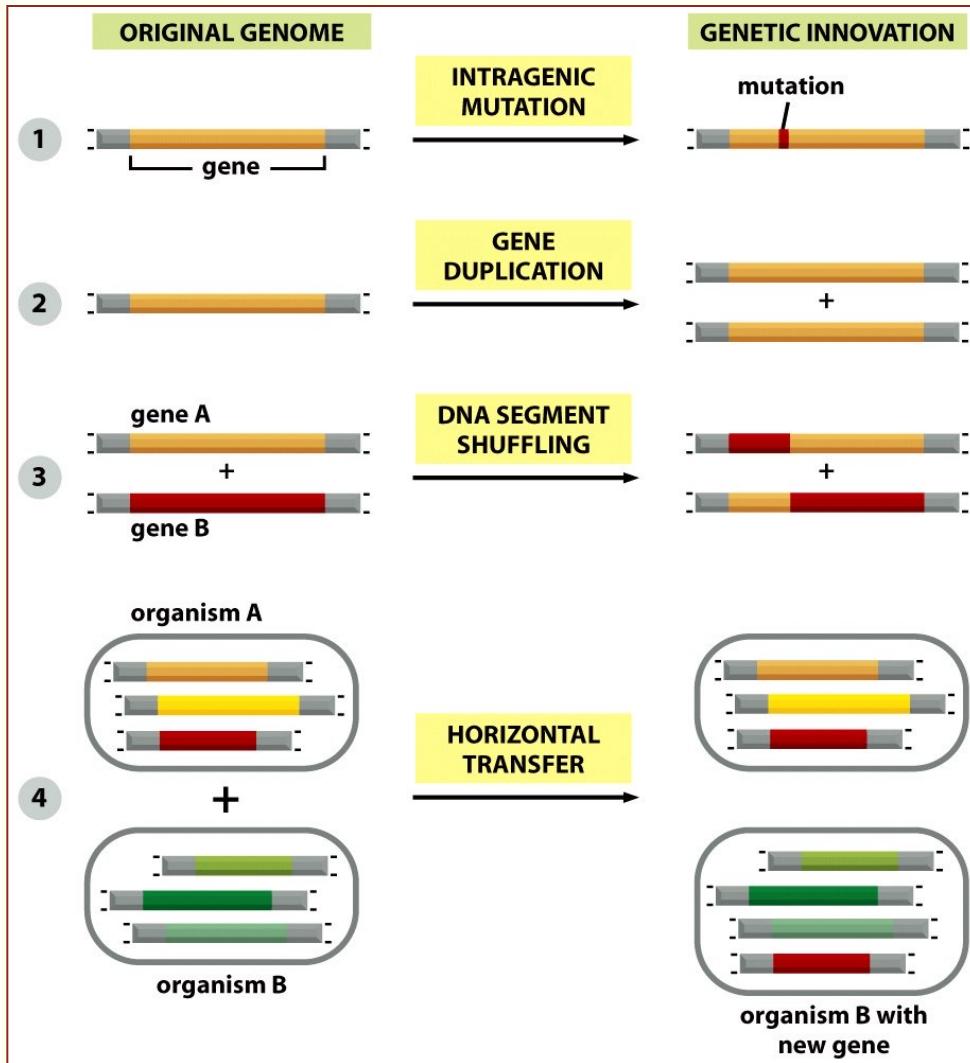
Imachi *et al.* (2020) Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* 577: 519

Tria *et al.* (2021) Gene Duplications Trace Mitochondria to the Onset of Eukaryote Complexity. *Genome Biol. Evol.* 13: 1

References in figure legends



# Factors Influencing Nuclear Genome Evolution



Mutation

Gene duplication and loss

Exon (domain) shuffling

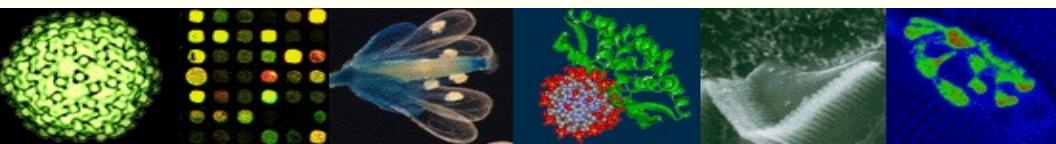
Repetitive DNA

Microsatellites

Transposable elements

Horizontal / Lateral gene transfer

Endosymbiosis



## Genome Evolution – Gene Duplication

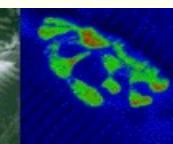
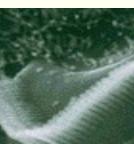
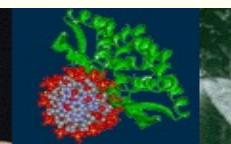
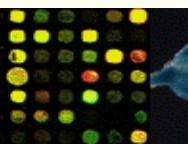
### Types of duplications

- entire genome duplications - polyploidy
- partial or entire chromosome duplications
- partial or entire gene duplications

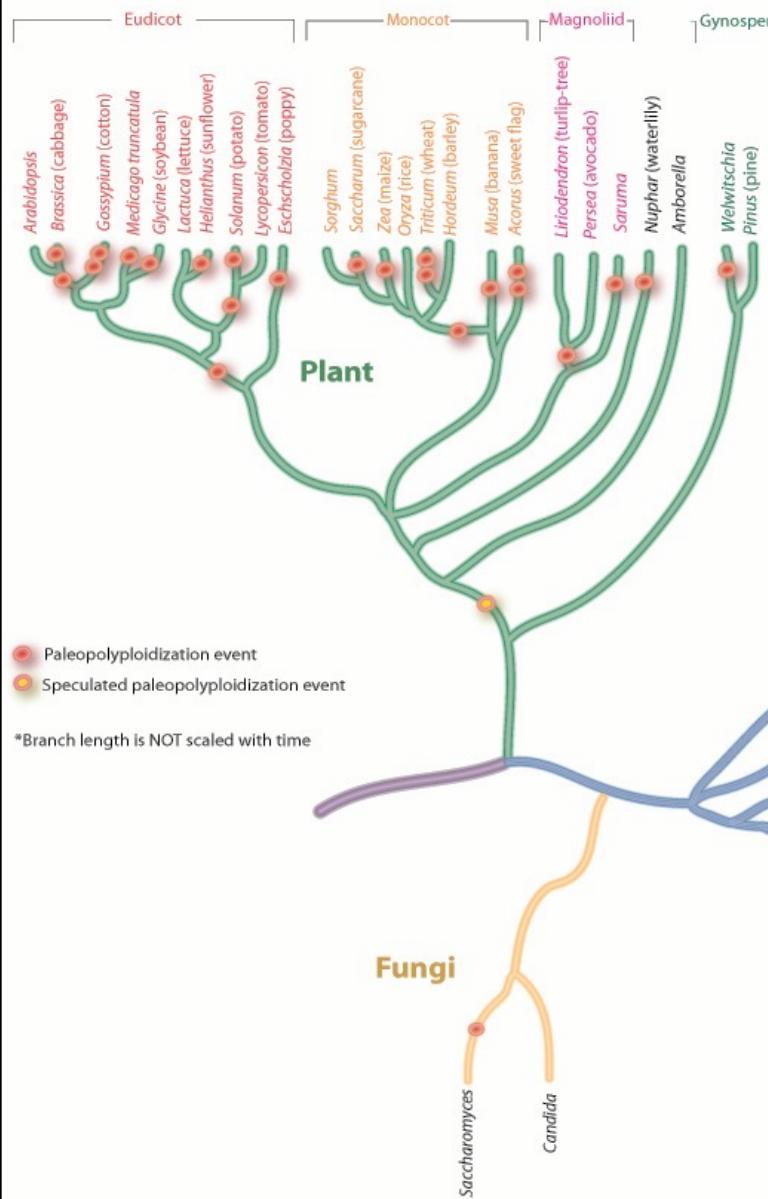
**Gene duplication allows novel gene functions and biochemical pathways to evolve, and the refinement of physiological processes**

**Table 1.** Comparative analysis of gene and repeat content of the human, fruitfly, worm and yeast genomes.

	Genome size (Mb) <sup>1</sup>	Total no. of genes/ % duplicated genes <sup>2</sup>	Fraction of proteome with human homologs	No. of domain types <sup>3</sup>	Distinct domain architectures (two or more domains) <sup>4</sup>
S. cerevisiae	12	6680/30%	46%	973	470
C. elegans	100	20 060/49%	43%	1183	1248
D. melanogaster	133	14 039/41%	61%	1218	1702
H. sapiens	3,093	23 224/38%	n/a	1865	3433



### Known Paleopolyploidy in Eukaryotes



## Most Eukaryotic Lineages Have Undergone Whole Genome Duplication (WGD) Events

### Plant Genomes

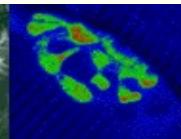
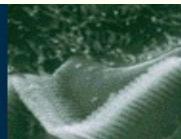
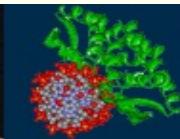
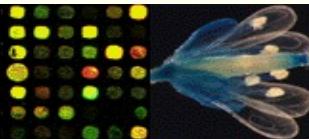
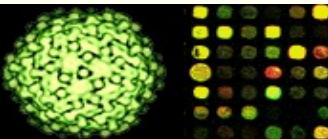
- appear to have gone through cycles of gene duplication events and gene loss
  - leads to increased fitness (survival) advantages that are lost over time and favour new WGD

### Vertebrate Genomes

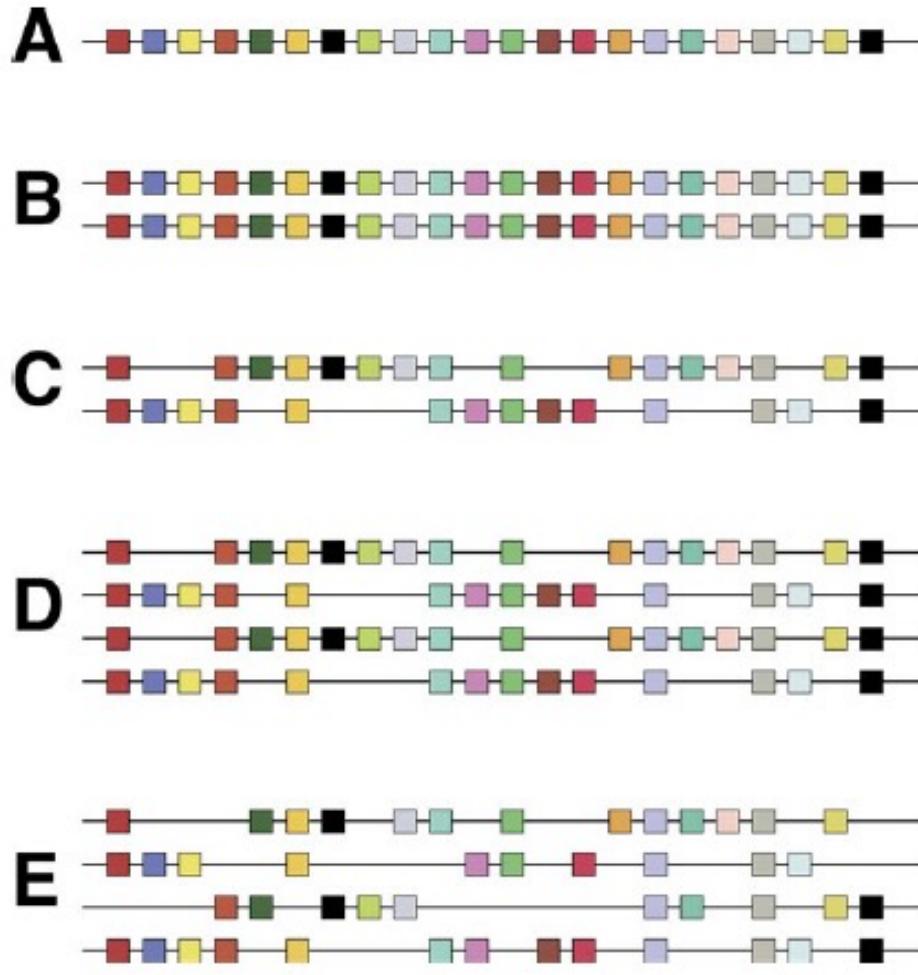
- ancestor went through 2 WGDs

*Saccharomyces cerevisiae* (baker's yeast)

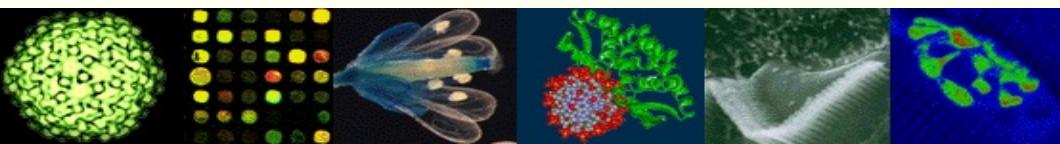
- ~10% of genes derived from WGD event, ~100 million years ago



## Gene Duplication – Fate of Duplicated Genes



- A. Hypothetical genome with 22 genes
- B. Duplication event – all genes duplicated; paralogues
- C. Many paralogues obtain disabling mutations, become pseudogenes and are lost
- D. Second duplication event – all paralogues duplicated. Multigene families – some with four copies, others with two
- E. Again, many paralogues obtain disabling mutations, become pseudogenes and are lost. (Transpositions and other duplication events, e.g. single gene duplications can occur)



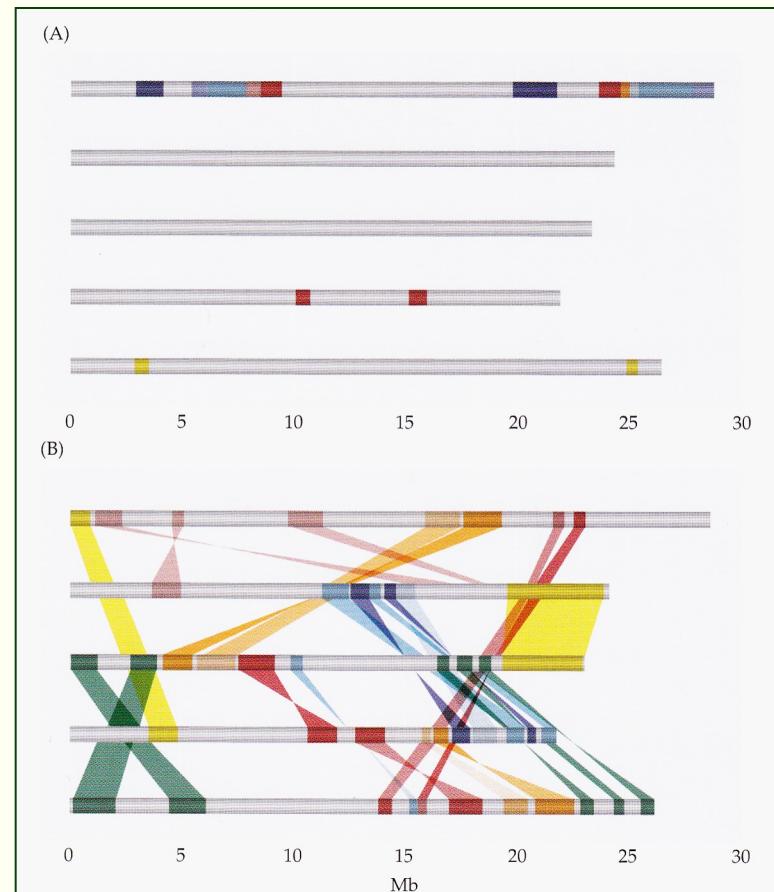
## Eukaryotic Genomes Contain Multiple Segmental Duplications

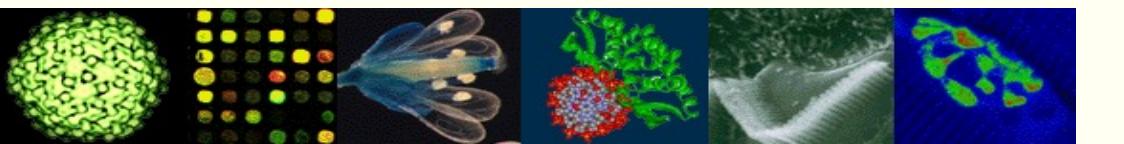
Genomes of multicellular eukaryotes have duplicated regions of more than 1000 bp

- intrachromosomal (most common) and interchromosomal duplications
  - result in multiple copies of genes
    - may become established as multigene families

Arabidopsis genome

- 30 segmental duplications
- intrachromosomal duplications (A) in 3 of the 5 chromosomes
- interchromosomal segmental duplications (B)
  - some inverted relative to one another (twisted bands)

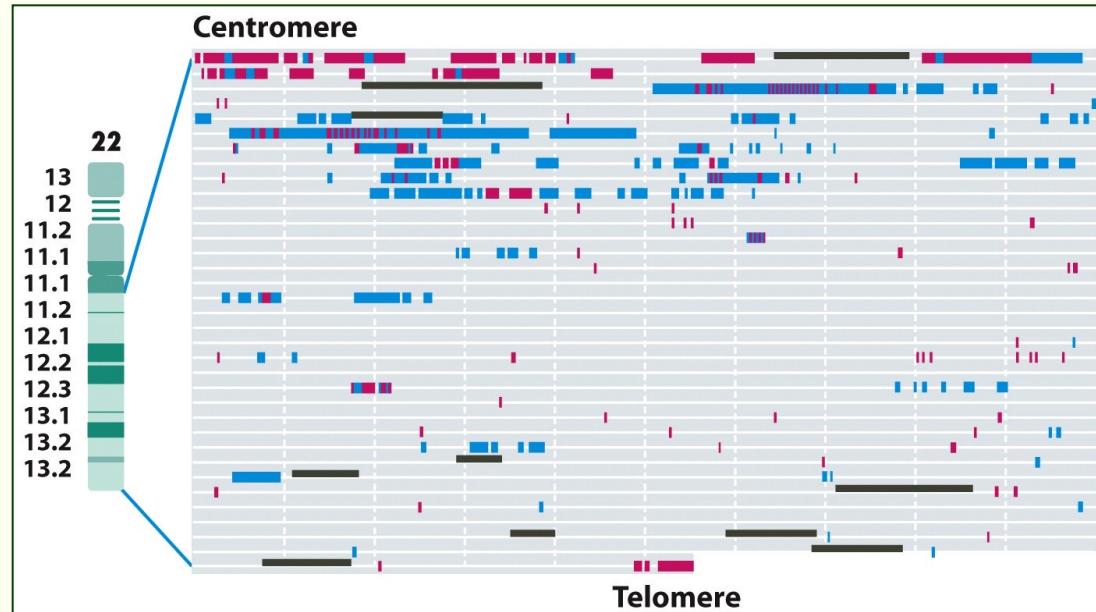




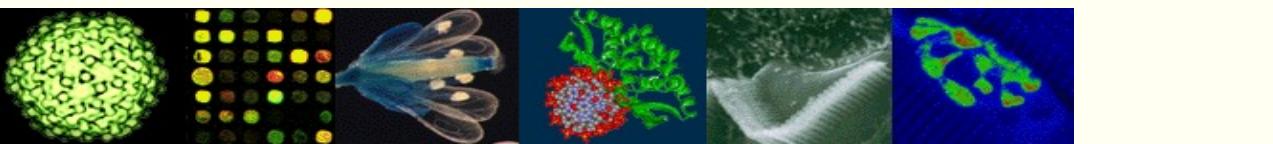
## Eukaryotic Genomes Contain Multiple Segmental Duplications cont' d

Approximately 4% of the human genome is composed of segmental duplications, averaging 15,000 bp in size

- duplication events on the long arm of chromosome 22
- nearly 200 segments (over 10%) of the chromosome arm have resulted from duplications
- duplication bias toward regions close to the centromere



pink bars = regions duplicated within this chromosome  
blue bars = regions duplicated from other chromosomes  
black bars = gaps in sequence when analysis was done

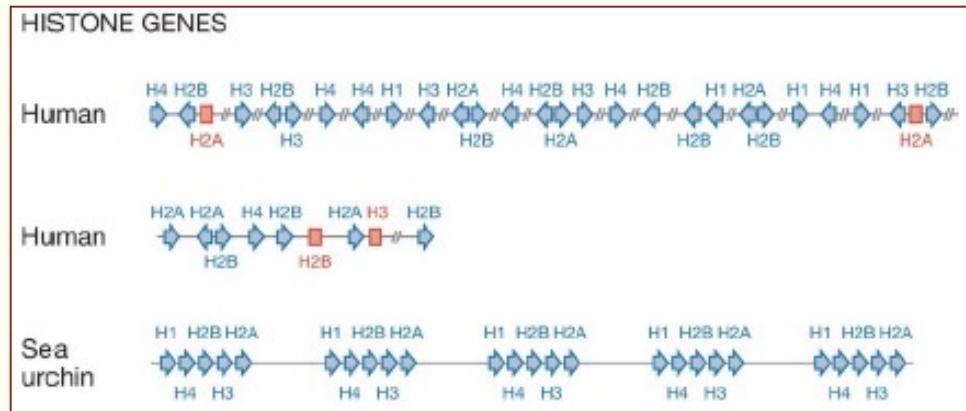


## Eukaryotic Genomes Contain Multigene Families

Single copy genes in multicellular eukaryotes make up 25-50% of protein-coding genes

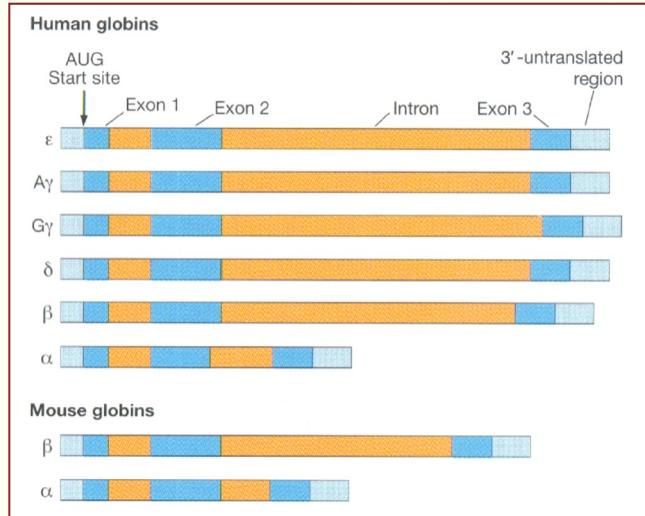
Simple (aka classical) gene families

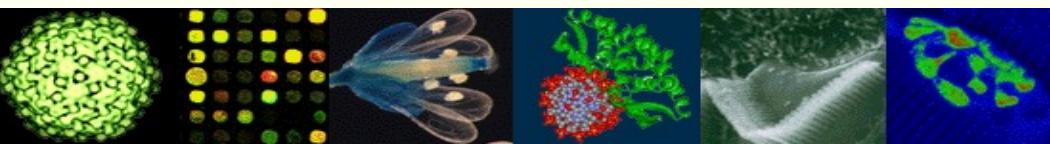
- all members have identical or nearly identical sequences



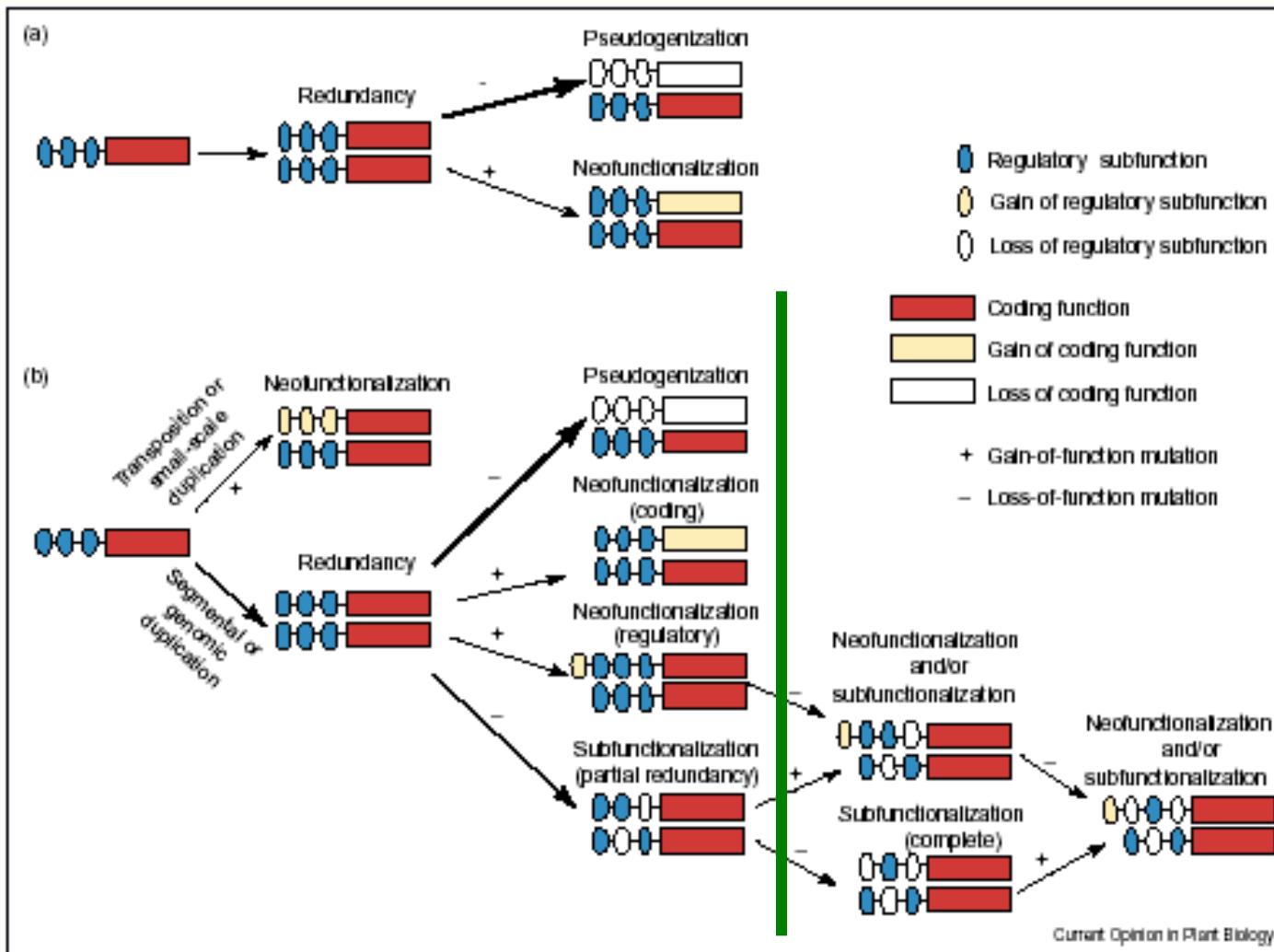
Complex gene families

- members have similar sequences
- different enough to code for gene products with different properties

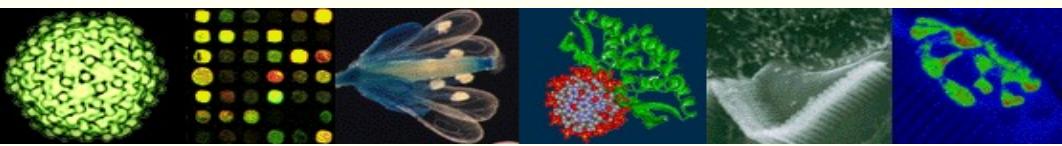




## Gene Duplication – Fate of Duplicated Genes



(a) Classic Ohno model of duplicate gene fates. Mechanisms of duplication and fates of genes are indicated. Thickness of arrows indicate relative frequency of possible fates. (b) Recent theoretical work supports a much more complex model for the fates of duplicate genes.



## Orthologues and Paralogues

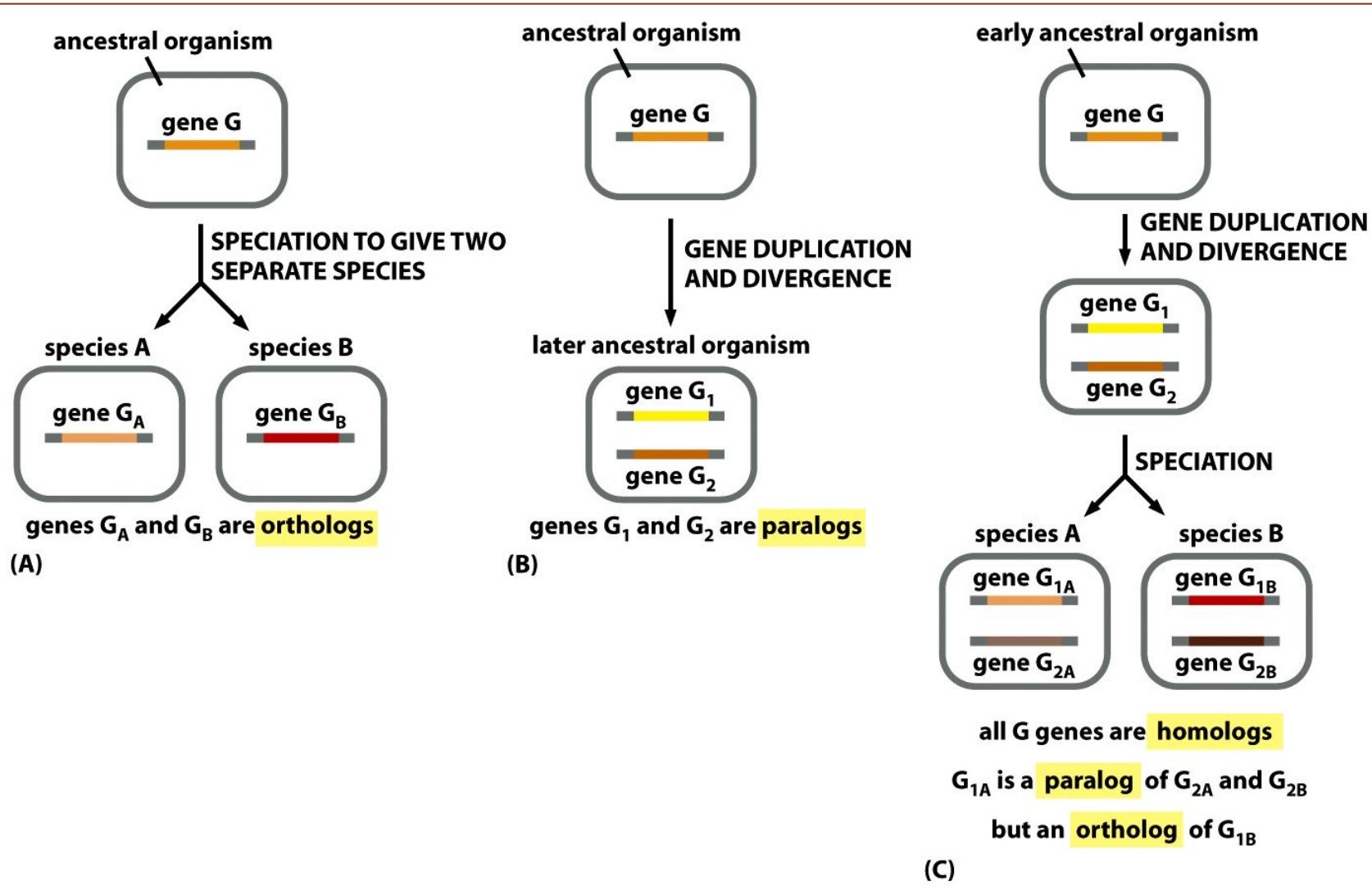
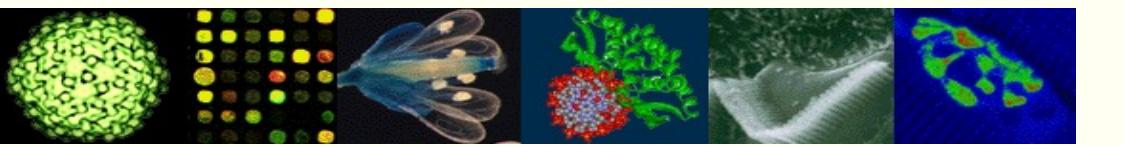


Fig 1-21 Alberts et al. (2015) Molecular Biology of the Cell

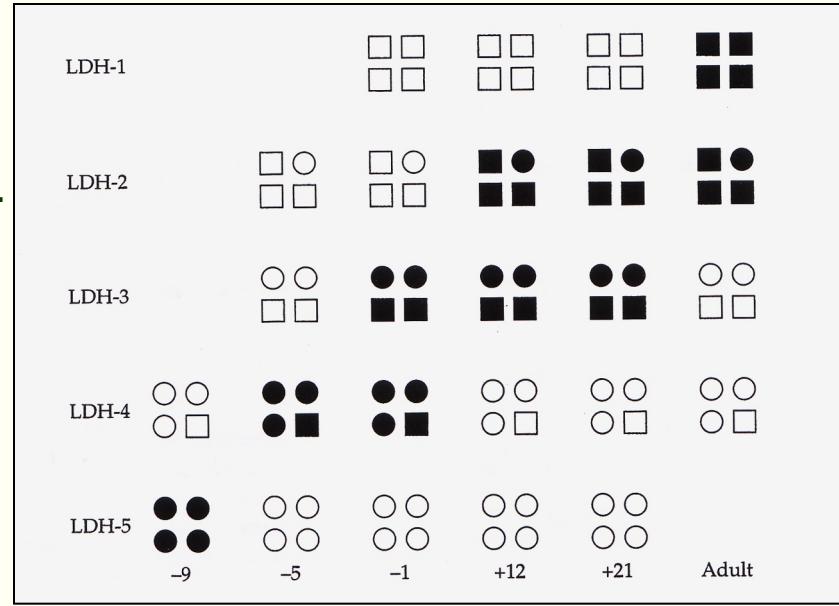


## Multigene Families – Fine-tuning & Expansion of Physiology

Isozymes = enzymes that catalyse the same biochemical reaction, but in different tissues, or at different times, or with different properties (e.g. kinetics)

e.g. lactate dehydrogenase (LDH) isozyme expression during rat heart development

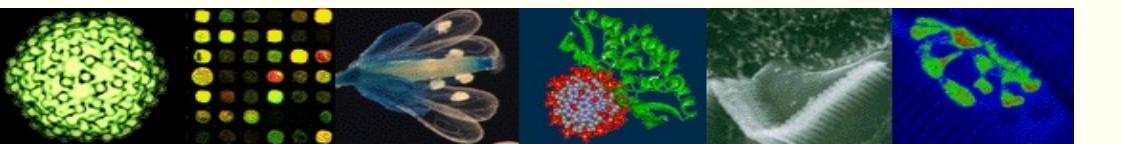
- as heart develops, isoenzymes with largely B subunits (associated with aerobically active tissues) are expressed



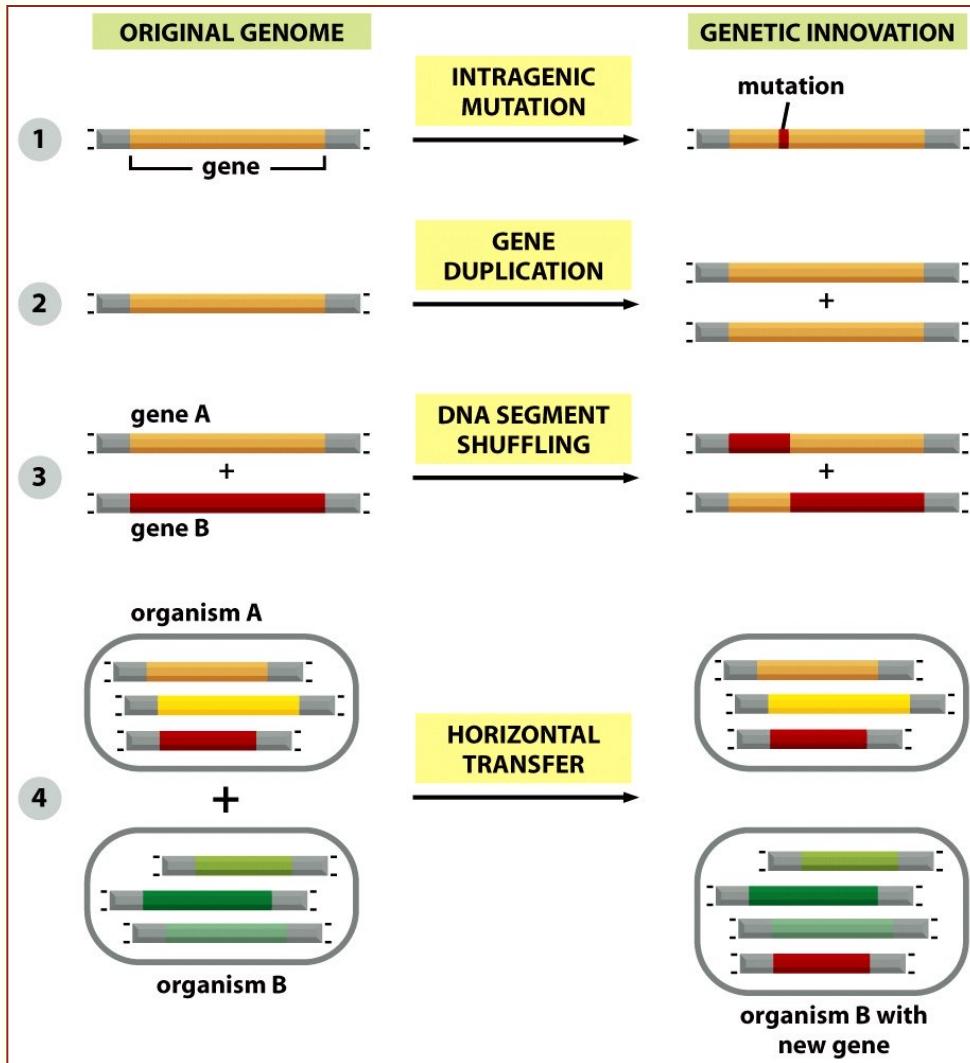
numbers indicate days before (-) & after (+) birth  
 circles = A subunits; squares = B subunits  
 solid symbols = predominant form of subunit

Gene losses and gene family expansions in different evolutionary lineages resulted in variability in the relative contributions of individual gene families to the total coding DNA seen across species

- e.g. human vs mouse genomes
  - roles in cellular ion and metabolite transport
  - second messenger signaling
  - synaptic transmission



# Factors Influencing Nuclear Genome Evolution



Mutation

Gene duplication and loss

Exon (domain) shuffling

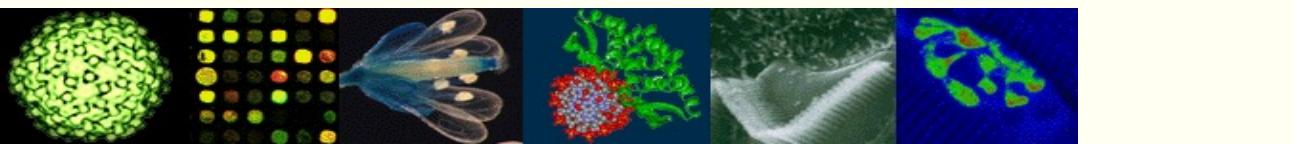
Repetitive DNA

Microsatellites

Transposable elements

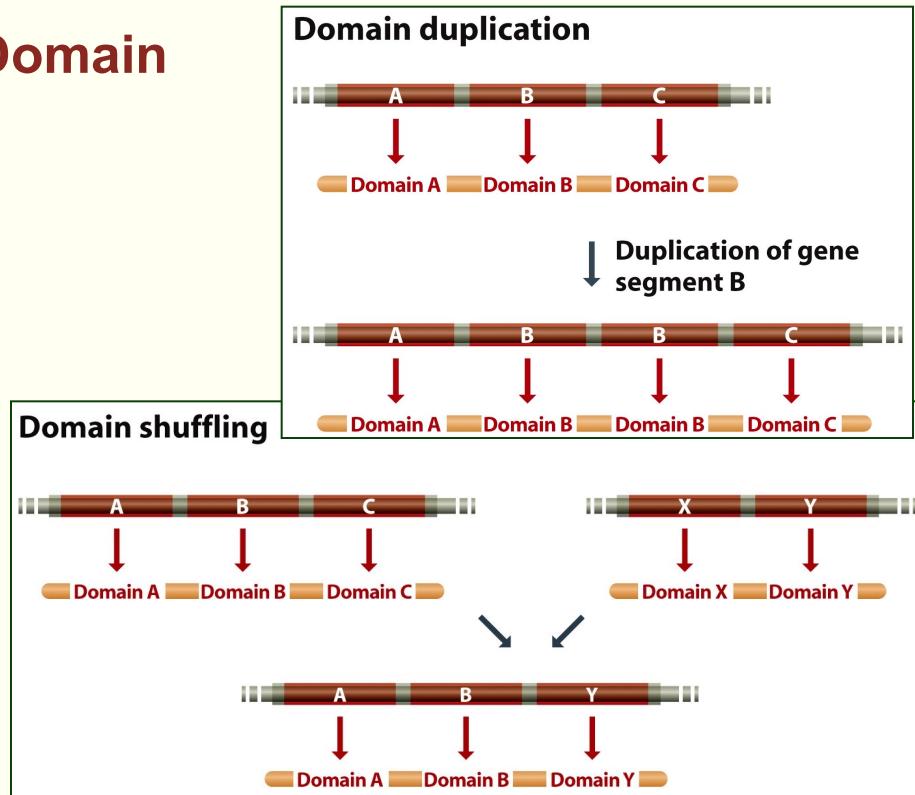
Horizontal gene transfer

Endosymbiosis



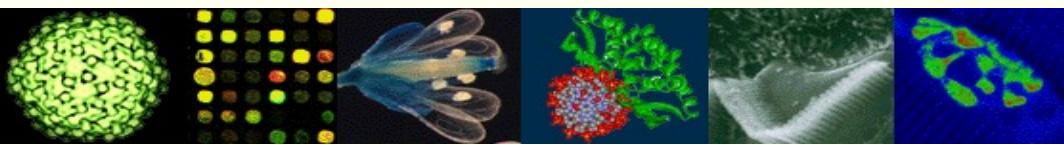
## Genome Evolution – Polypeptide Domain Shuffling

- domains may be either structural or functional
- involves combining existing domains into novel gene architectures
- may involve domain duplication – leads to elongation of gene; evolution of novel domain through mutation
- may involve domain insertion from a different gene – leads to mosaic genes
- ~20% of eukaryotic exons involved



**Table 1.** Comparative analysis of gene and repeat content of the human, fruitfly, worm and yeast genomes.

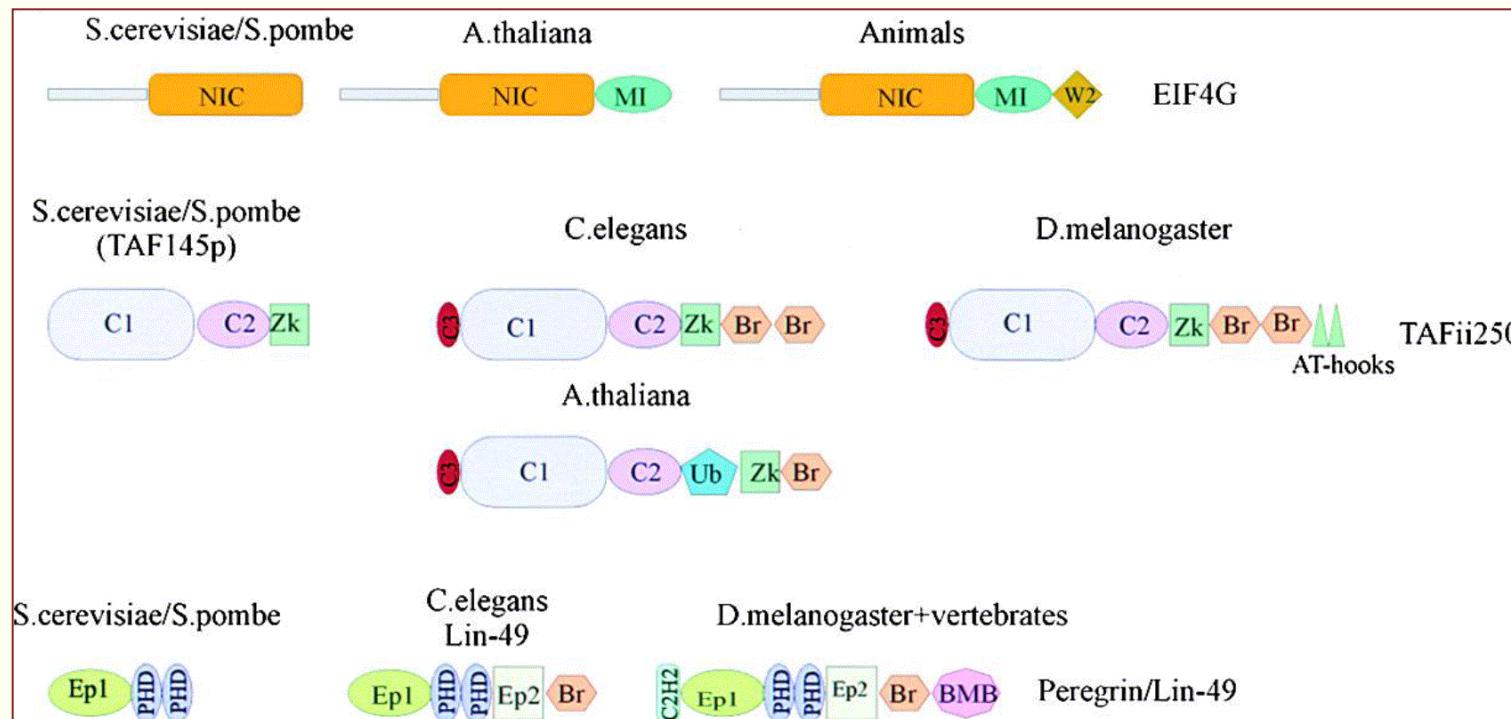
Genome size (Mb) <sup>1</sup>	Total no. of genes/ % duplicated genes <sup>2</sup>	Fraction of proteome with human homologs	No. of domain types <sup>3</sup>	Distinct domain architectures (two or more domains) <sup>4</sup>
S. cerevisiae	12	6680/30%	46%	973
C. elegans	100	20 060/49%	43%	1183
D. melanogaster	133	14 039/41%	61%	1218
H. sapiens	3,093	23 224/38%	n/a	1865
				3433

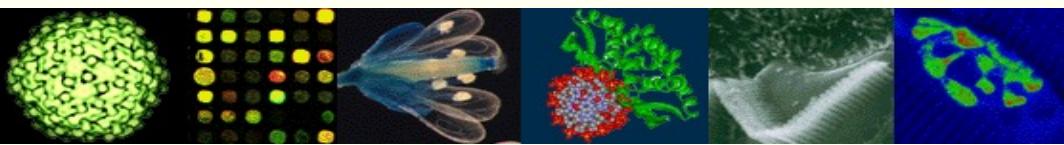


## Domain Shuffling – Increasing Gene Architecture Complexity

Gene architecture complexity increases with organism complexity

- Drosophila, vertebrate, and flowering plant genes have accumulated more domains and evolved more complex architectures than genes of worms or fungi
- Think about genome evolution in terms of evolution of domains making up genes

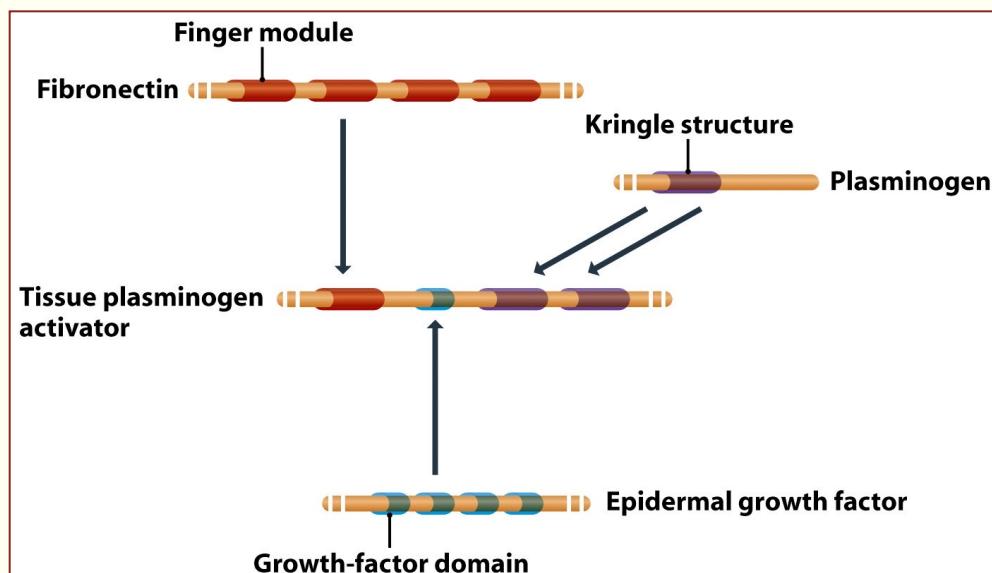




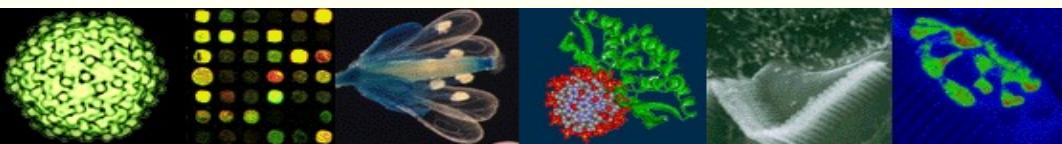
## Domain Shuffling – Increasing Gene Architecture Complexity cont'd

### Tissue plasminogen activator (TPA)

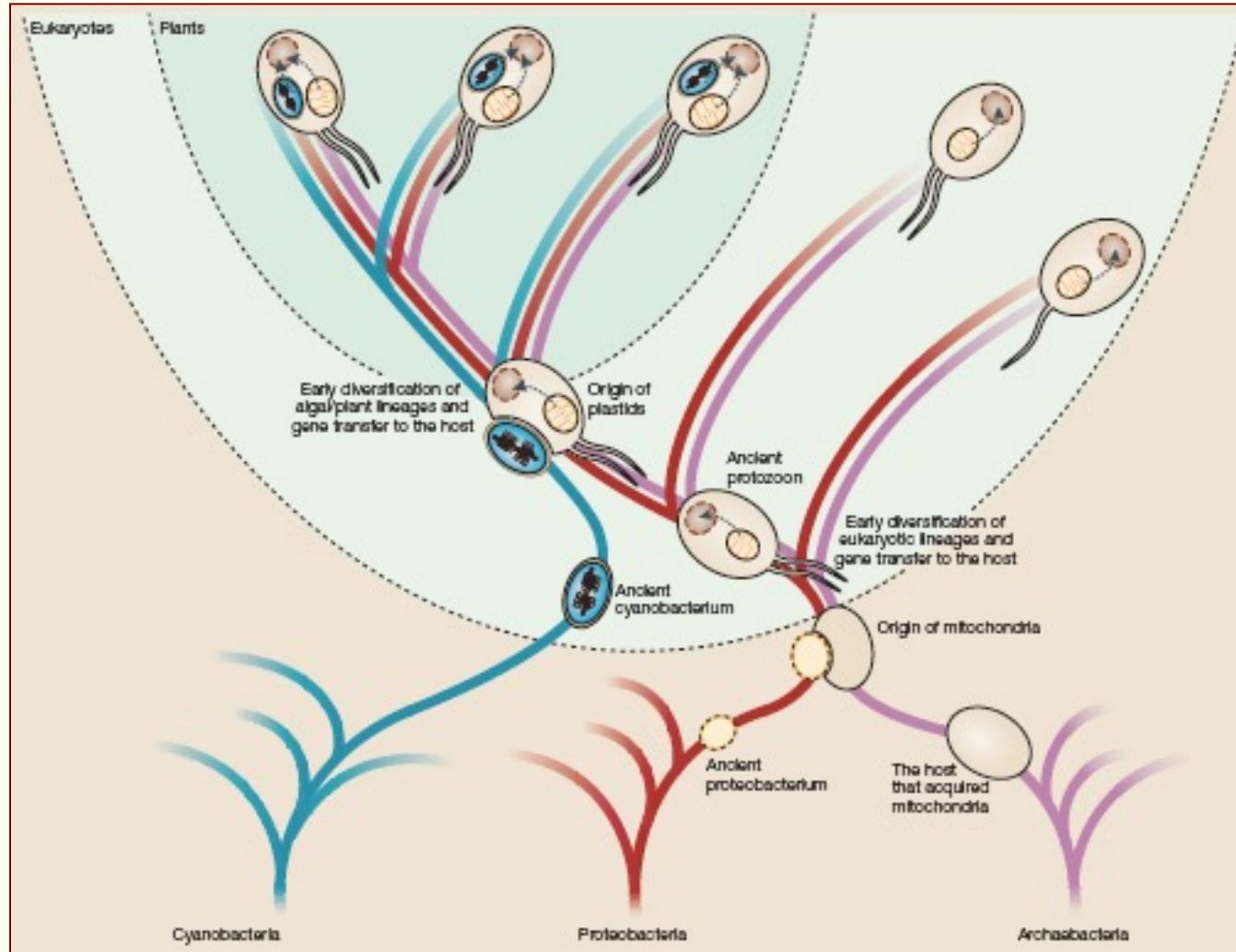
- involved in blood clotting
- one of the first discovered mosaic proteins
- TPA gene has four exons
  - 1<sup>st</sup> derived from gene encoding fibronectin
  - 2<sup>nd</sup> derived from epidermal growth factor gene
  - 3<sup>rd</sup> and 4<sup>th</sup> derived from gene encoding plasminogen



- borders of the protein domains match perfectly with exon-intron boundaries – supporting idea that during evolution exons may be transferred from one gene to another



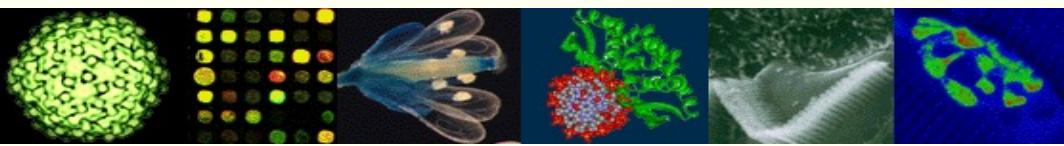
## Origin of the Eukaryotic Genome – Endosymbiosis



Nucleus

Mitochondrion

Chloroplast

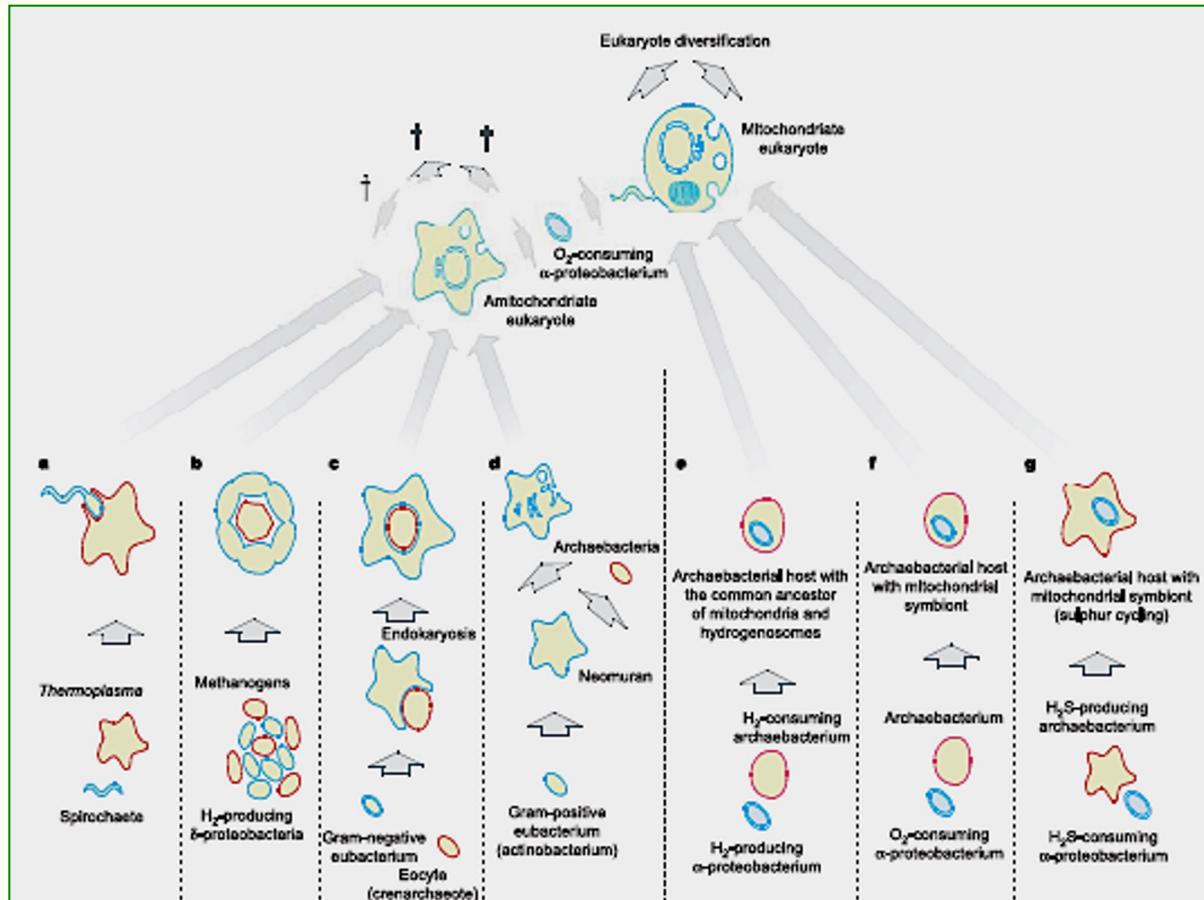


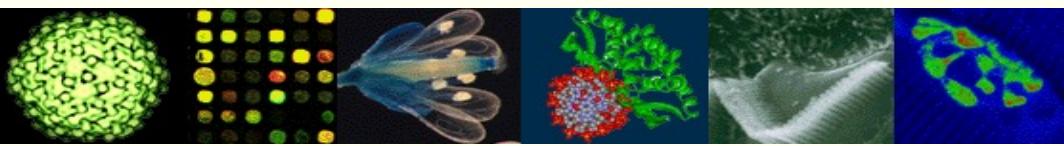
## Origin of the Nucleus

Highly debated

Prior to or concurrent with the acquisition of mitochondrion?

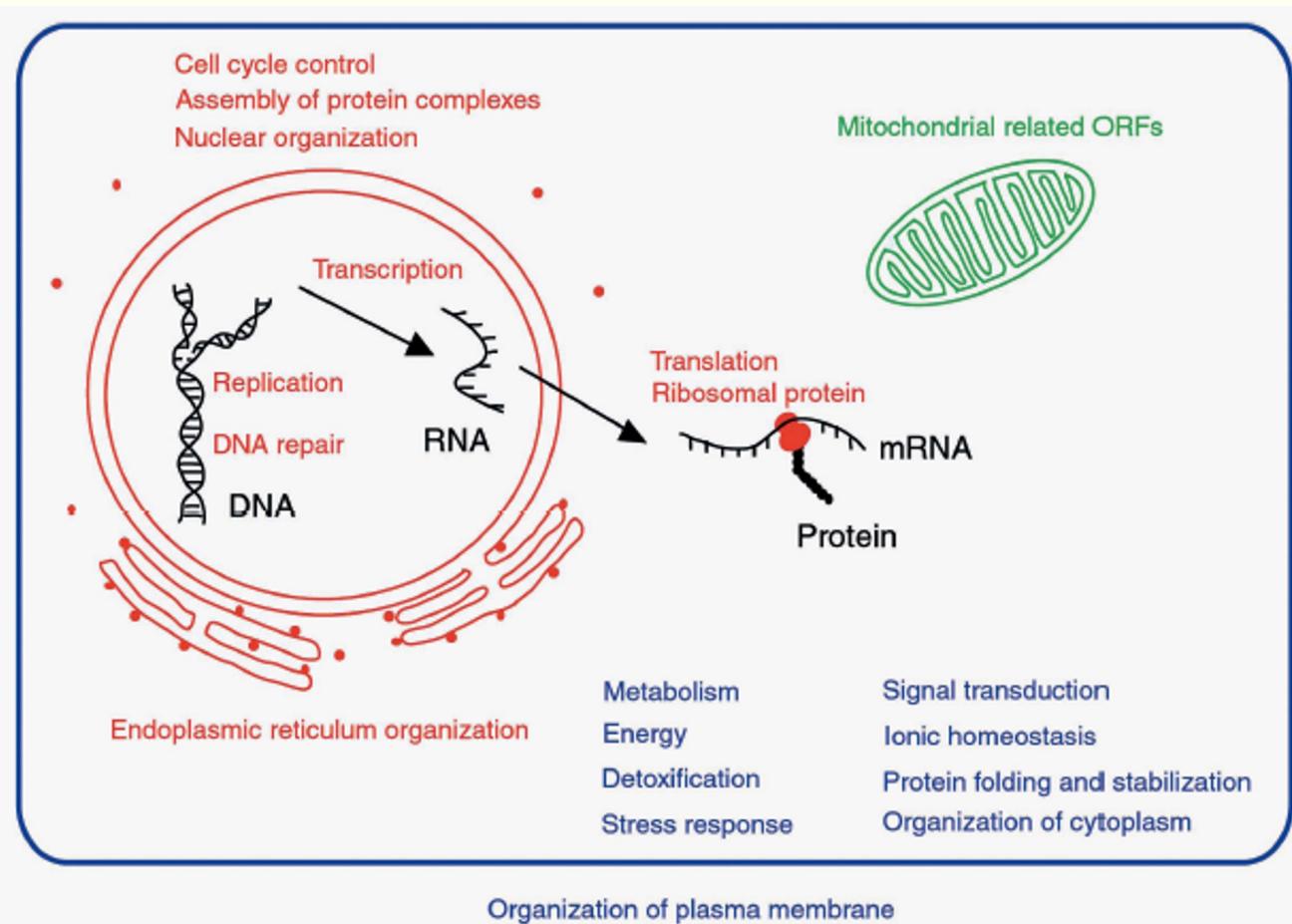
- for latest, see Tria *et al.* (2021) *Genome Biol. Evol.* 13: 1





## Origin of the Nucleus cont'd

Comparison of yeast nuclear open reading frames (excluding those encoding mitochondrial proteins) with those of eight archaea and 12 eubacteria

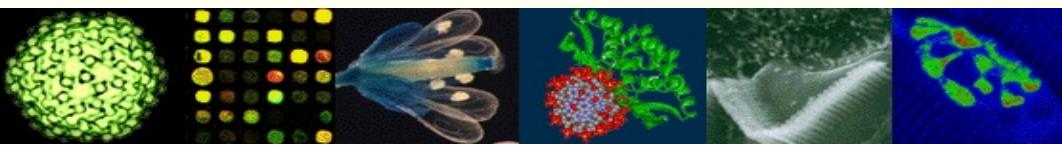


red – ORF groups originating from archaea

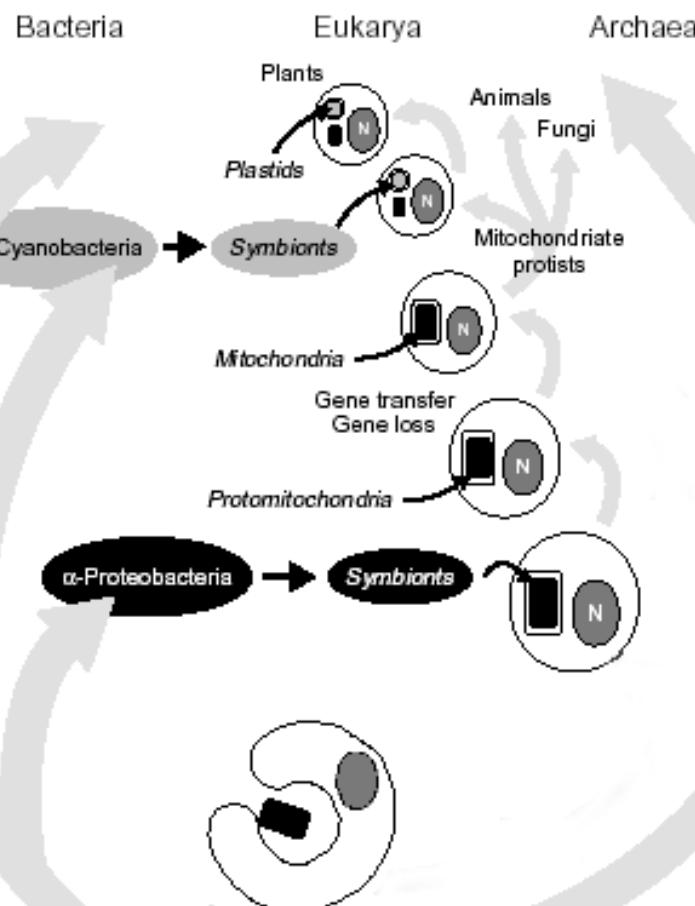
blue – ORF groups originating from eubacteria

Concluded: an archeal cell entered a eubacterial cell and evolved into a nucleus

Now identity of archaeon being debated, e.g., Imachi et al. (2020) Nature 577: 519



## Origin of Mitochondria



Free-living  $\alpha$ -proteobacterium

- typical genome size - 1 to 9 Mbp

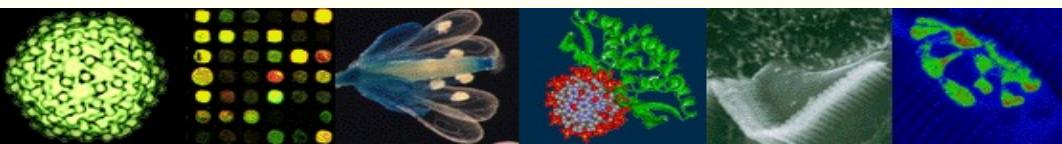
Mitochondrial genome size:

typically 15-60 kbp  
extremes

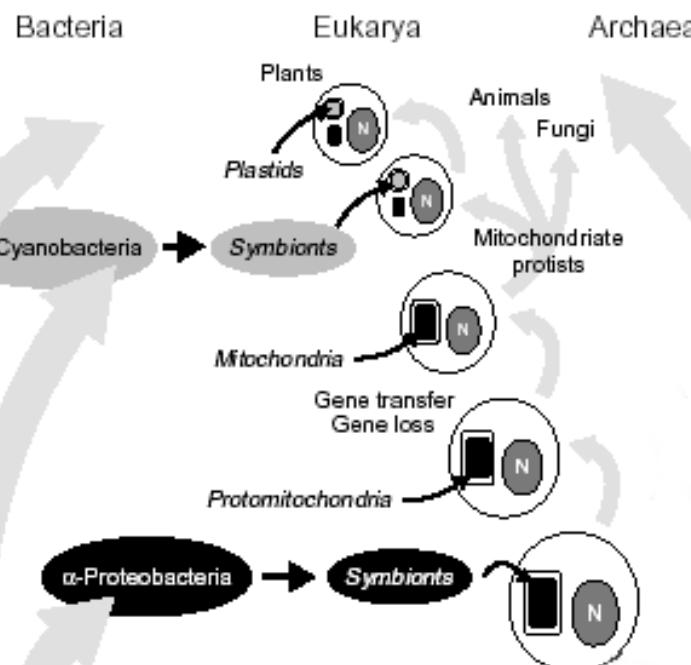
- *Plasmodium* sp. - 6 kbp
- curcubit plants - 2000 kbp

Mitochondrial genome coding capacity:  
on average - 50-60 genes

Reduction in genome content due to  
gene transfer and gene loss



## Origin of Chloroplasts



Free-living cyanobacterium

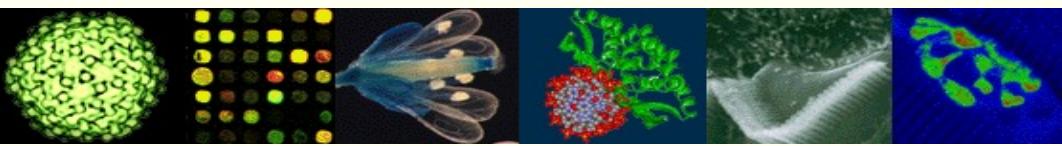
- typical genome size - 1.5 to 8 Mbp

Chloroplast genome size:

typically 120-160 kbp  
extremes

Chloroplast genome coding capacity:  
on average – 50-200 genes

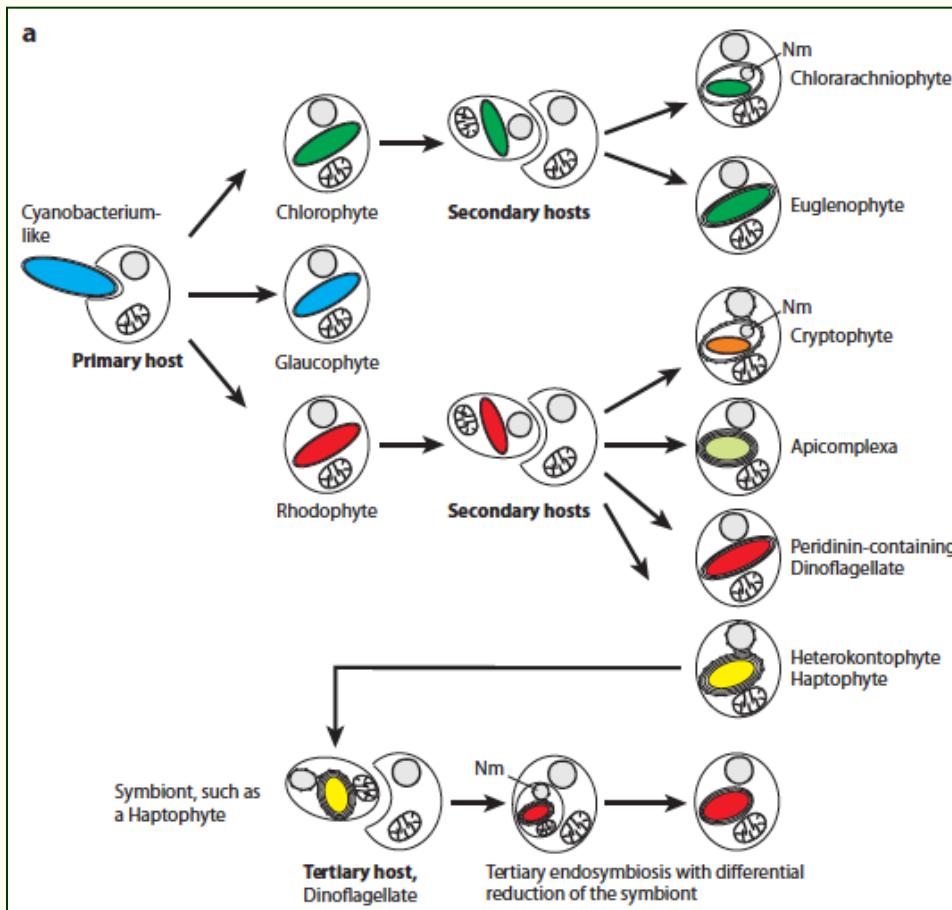
Reduction in genome content due to  
gene transfer and gene loss

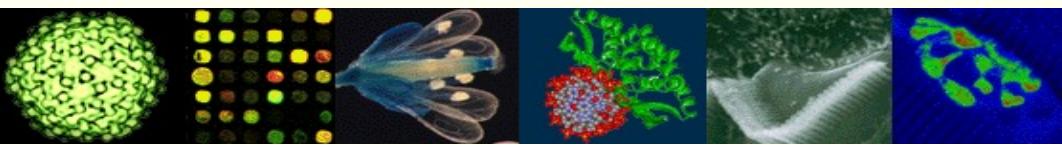


## Secondary and Tertiary Origins of Chloroplasts!

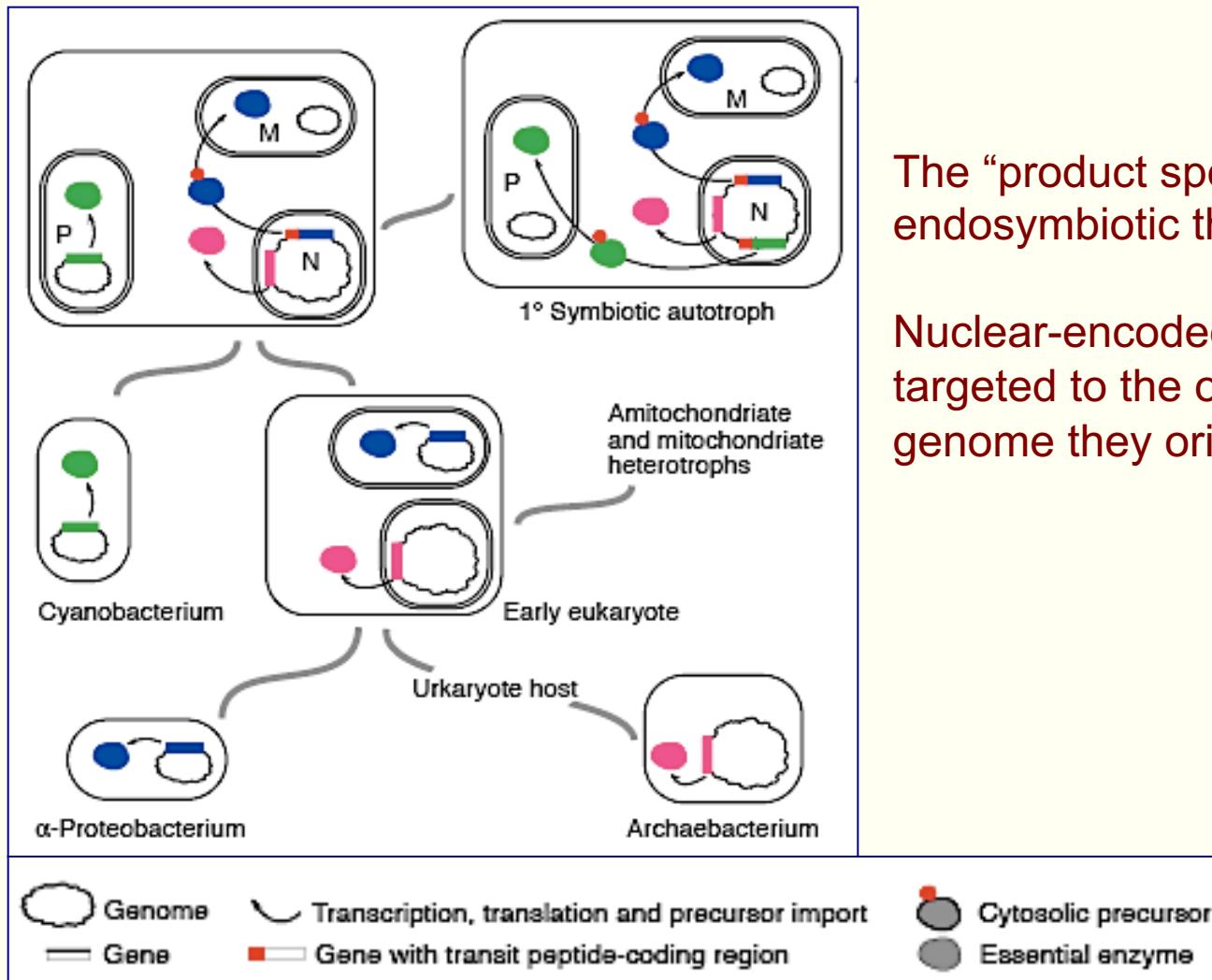
Eukaryote-eukaryote symbioses

Common in the photosynthetic protists (algae)



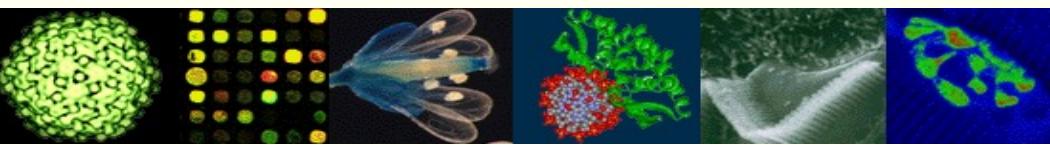


## Evolution of Nuclear and Organelle Genomes



The “product specificity corollary” of the endosymbiotic theory:

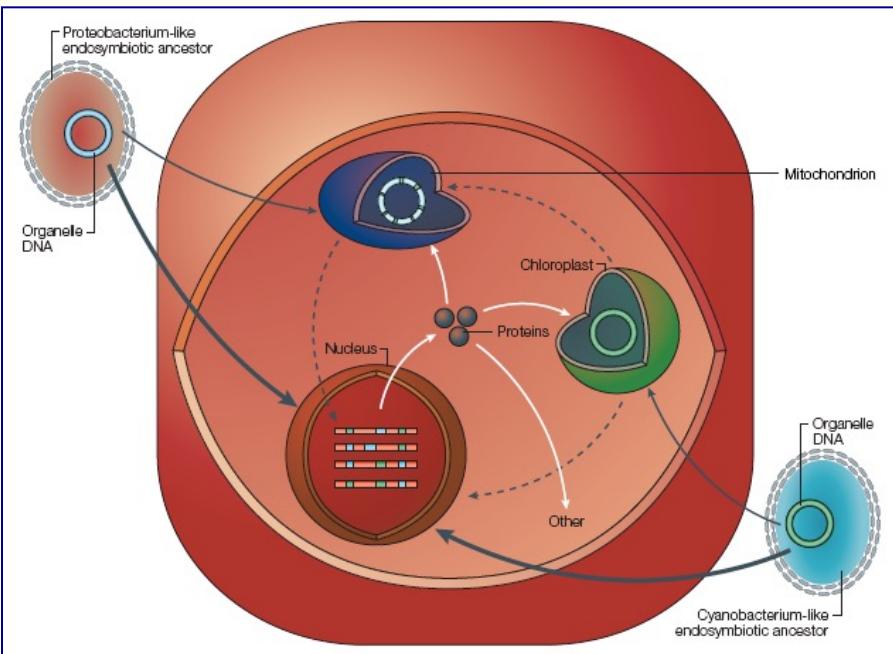
Nuclear-encoded gene products are re-targeted to the organelle from whose genome they originated



## Evolution of Nuclear and Organelle Genomes cont'd

Chloroplasts and mitochondria import ~90% of their proteins from the cytosol

HOWEVER: many mitochondrial proteins are not of  $\alpha$ -proteobacterial origin, and many chloroplast proteins are not of cyanobacterial origin!

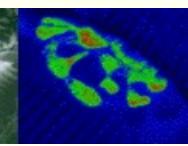
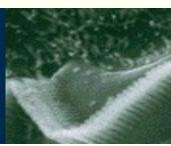
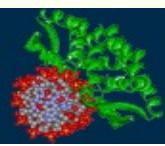
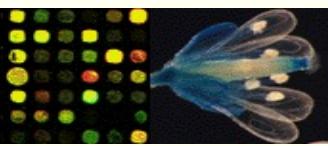
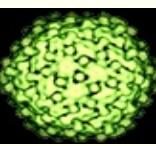


Timmis et al. (2004) Nature Reviews Genetics 5: 123

The product specificity corollary of the endosymbiotic theory is not supported

<30% of the proteins encoded by genes originating from  $\alpha$ -proteobacterium are predicted to be targeted to yeast and human mitochondria

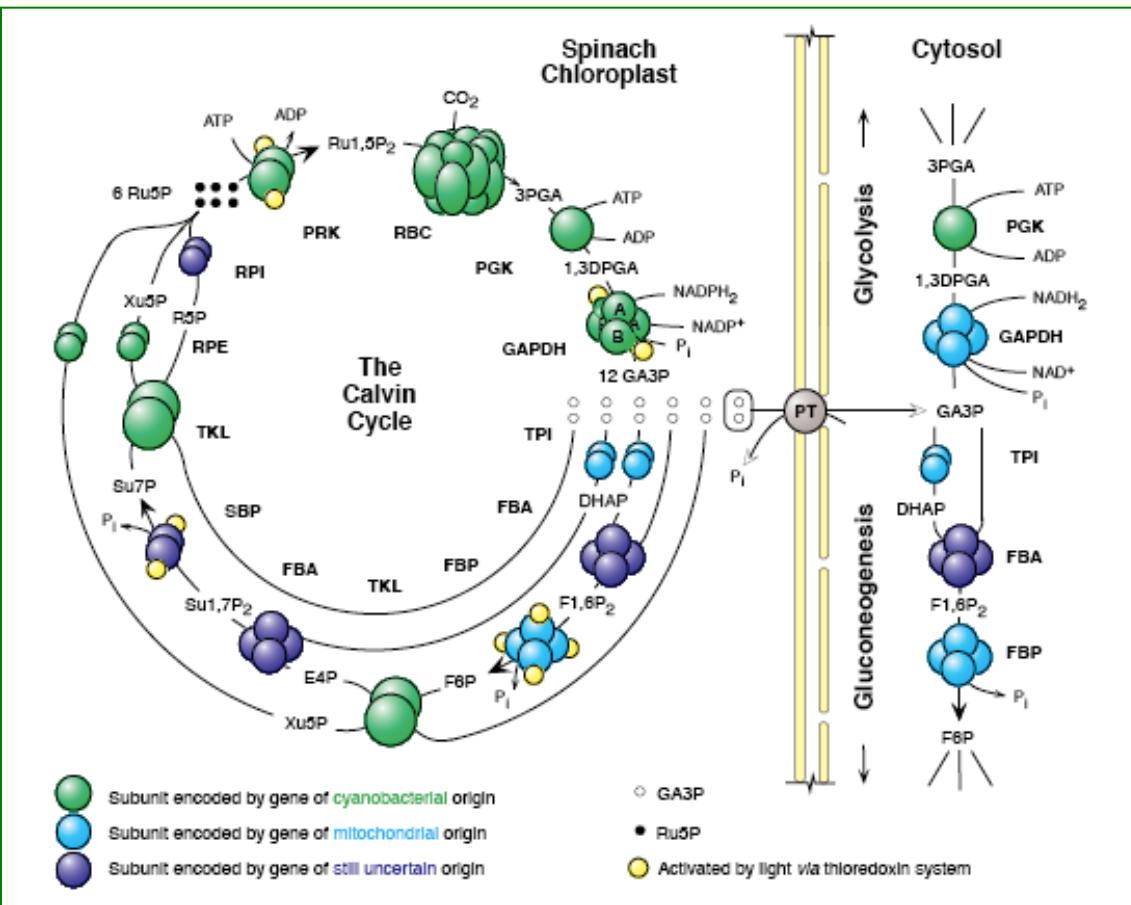
<50% of the proteins encoded by genes originating from cyanobacterium are predicted to be targeted to *Arabidopsis* chloroplasts



## Evolution of Nuclear and Organelle Genomes cont' d

### Calvin cycle

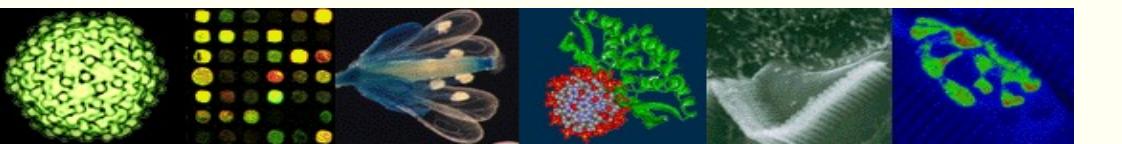
- CO<sub>2</sub> fixation pathway in plant chloroplasts
- enzymes of pathway are encoded by nuclear genes of cyanobacterial and α-proteobacterial origin



Cytosolic isoforms of these enzymes are typically encoded by genes of endosymbiont origin!

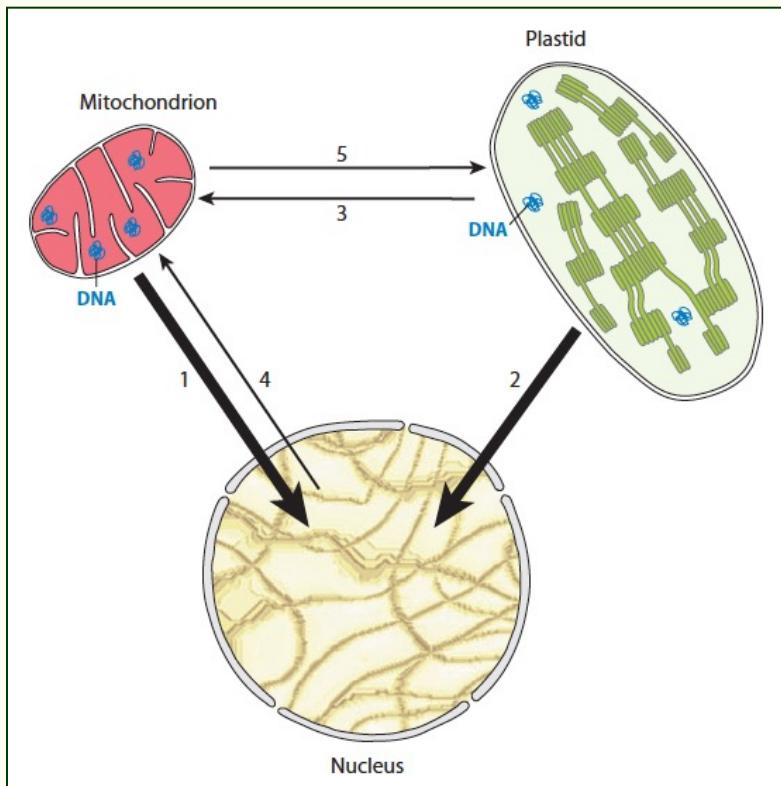
### Outcomes

- opportunity for evolution of novel function
- co-evolution of nuclear, mitochondrial and chloroplast genomes in eukaryotic cells



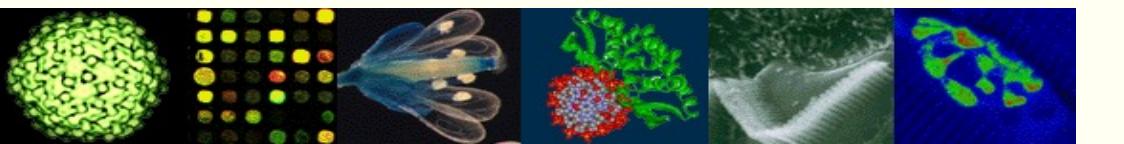
## Evolution of Nuclear and Organelle Genomes cont' d

Why were endosymbiont genes transferred to the host cell nucleus?



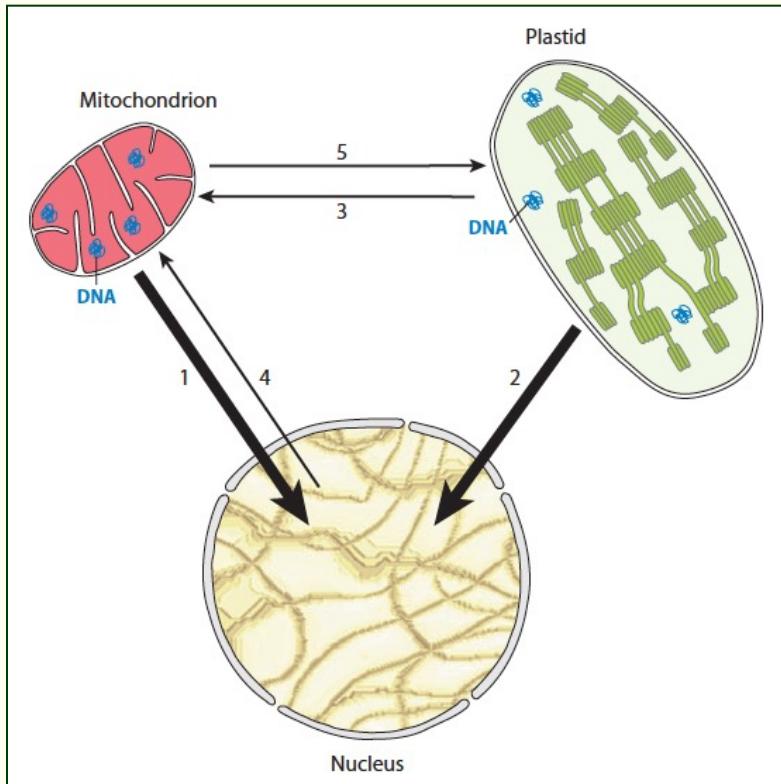
Evolutionarily advantageous to:

- isolate genes from sites where mutagenic radicals are formed
- have genes in a sexual population instead of asexual population (avoidance of Muller's Ratchet)



## Evolution of Nuclear and Organelle Genomes cont' d

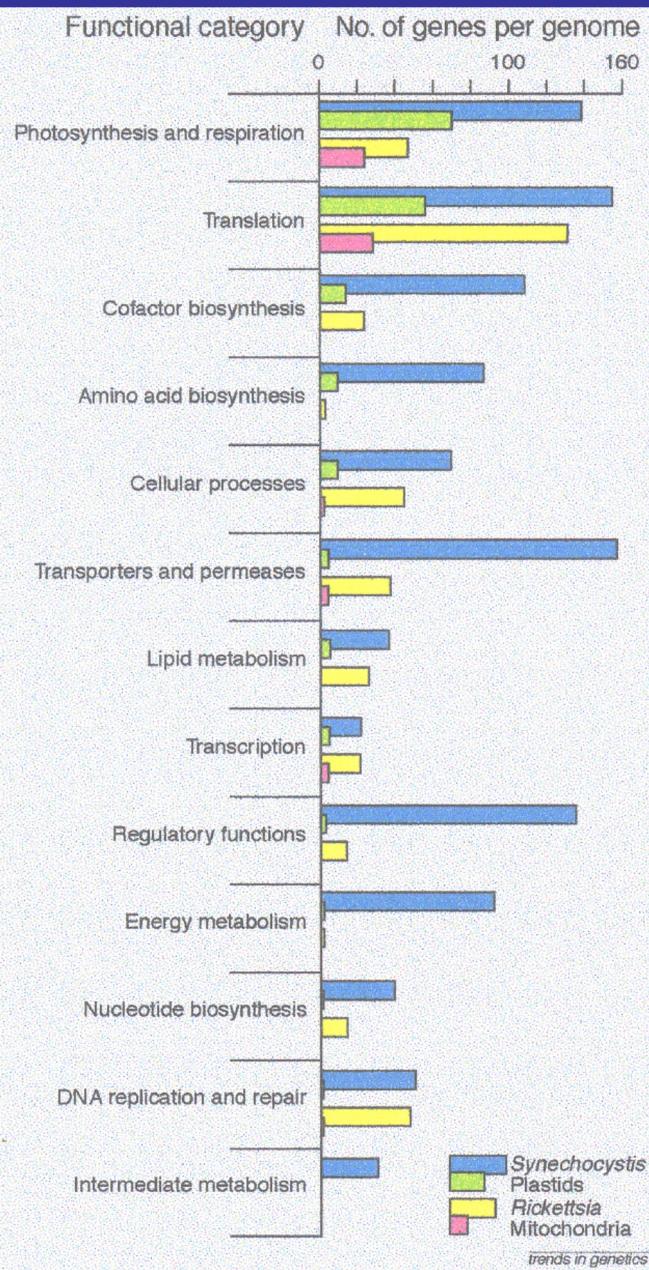
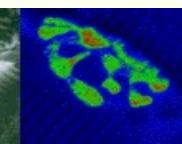
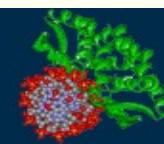
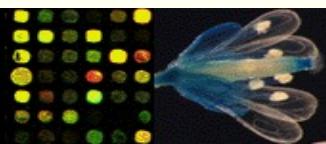
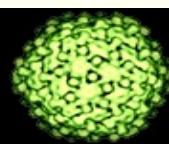
Why have not all endosymbiont genes been transferred to the host cell nucleus?



**Hydrophobicity Hypothesis (HH):**  
Some gene products are too hydrophobic to be imported into organelles.

**Code Disparity Hypothesis (CDH):**  
Genetic code used in organelles and nucleus differ, which would lead to incorrectly translated proteins.

**Co-location for Redox Regulation (CORR):**  
Proximity of genes to their products' sites of activity allows a rapid response in gene expression to changes in metabolism.



## What is Left in Organelle Genomes?

Two main functional categories of genes retained:

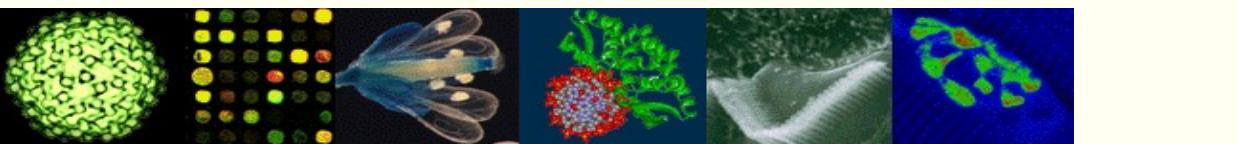
### 1) Proteins of the bioenergetic membranes

- photosynthetic
- respiratory

### 2) Components of gene expression

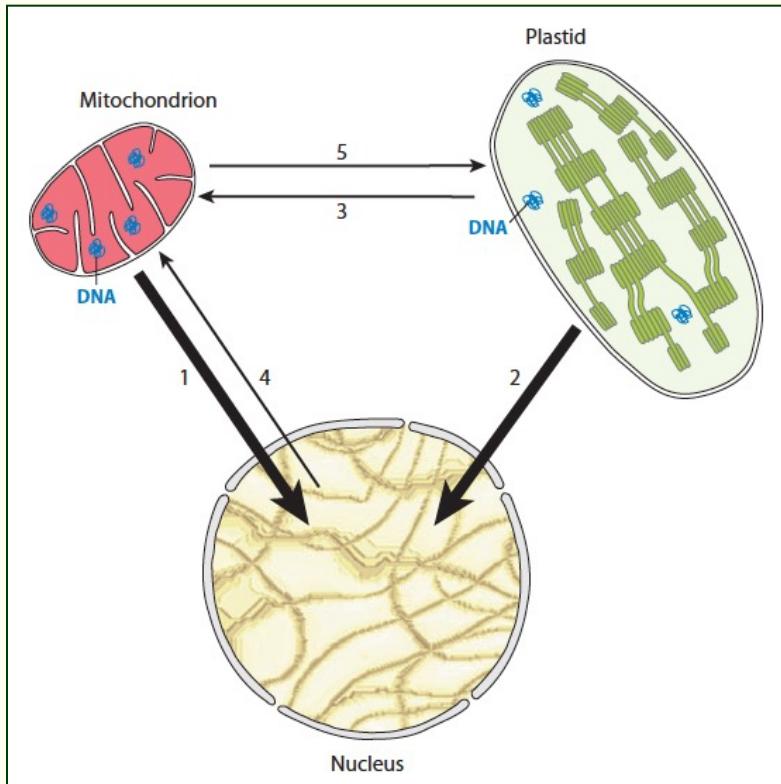
- rRNAs
- tRNAs

So do these data support any one hypothesis over the others?



## Evolution of Nuclear and Organelle Genomes

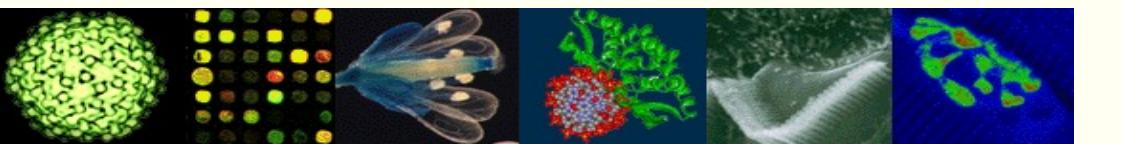
Why have not all endosymbiont genes been transferred to the host cell nucleus?



**Hydrophobicity Hypothesis (HH):**  
Some gene products are too hydrophobic to be imported into organelles.

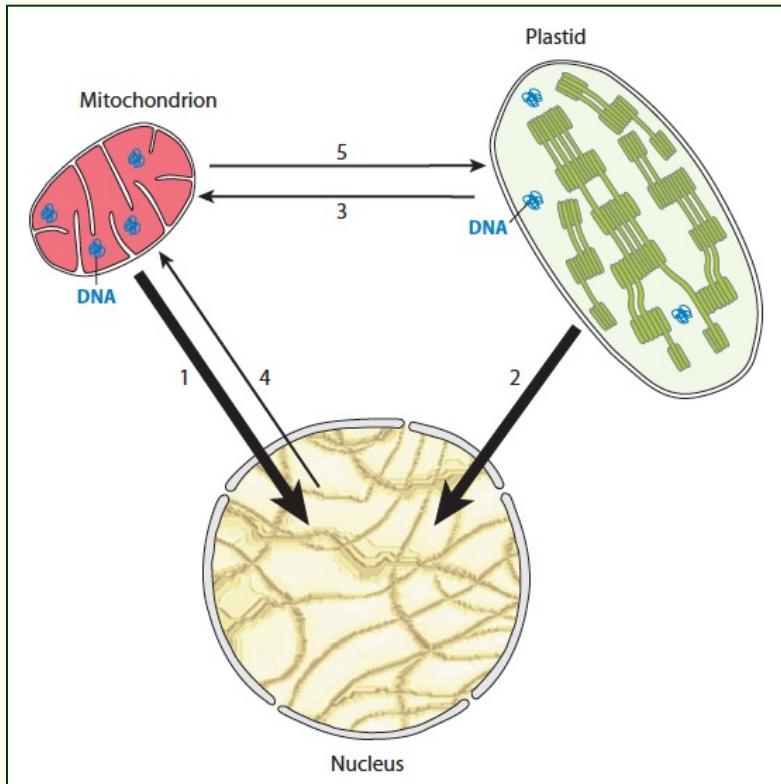
Animal and fungal proteins encoded in the mitochondrion support this hypothesis.

Nuclear-encoded proteins of the mitochondrial respiratory chain and chloroplast light harvesting proteins argue against this hypothesis.



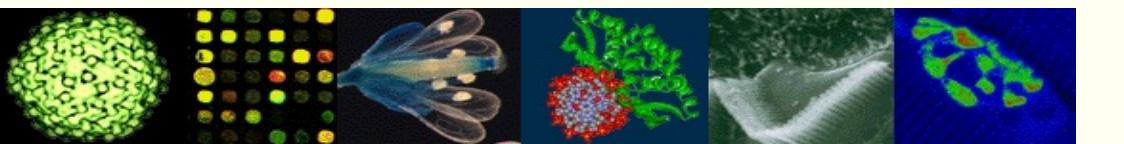
## Evolution of Nuclear and Organelle Genomes cont' d

Why have not all endosymbiont genes been transferred to the host cell nucleus?



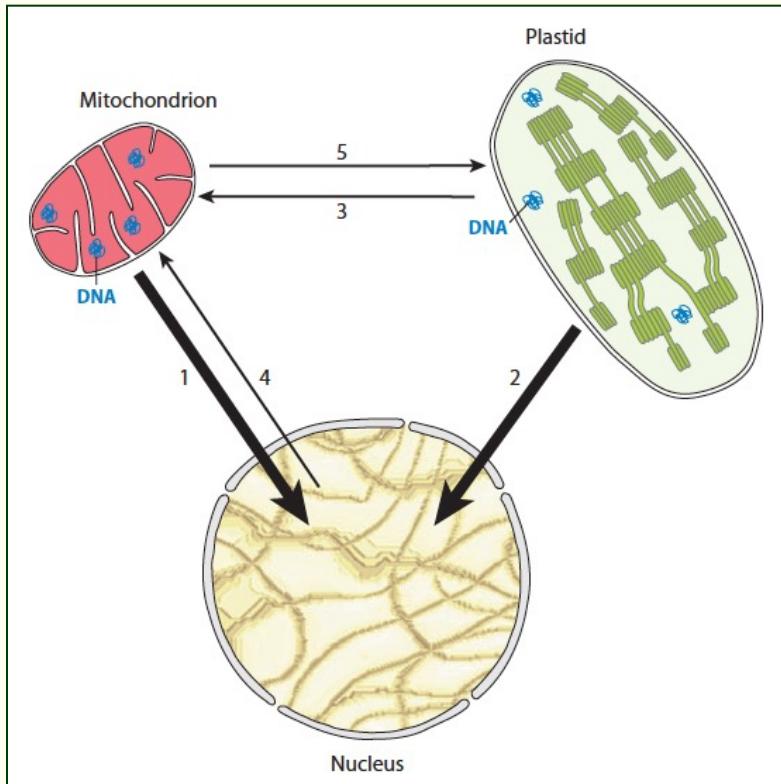
**Code Disparity Hypothesis (CDH):**  
Genetic code used in organelles and nucleus differ, which would lead to incorrectly translated proteins.

Most organelle genomes use the standard code; however, disparity exists between nuclear and mitochondrial codes in animals and protists.



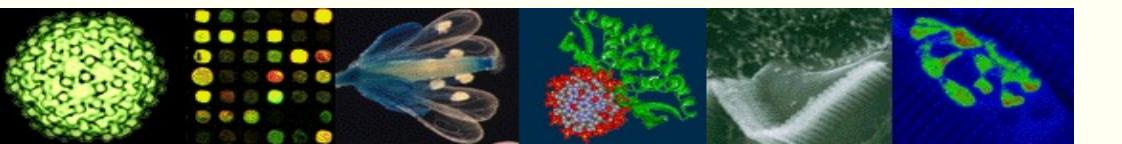
## Evolution of Nuclear and Organelle Genomes cont' d

Why have not all endosymbiont genes been transferred to the host cell nucleus?



**Co-location for Redox Regulation (CORR):**  
Proximity of genes to their products' sites of activity allows a rapid response in gene expression to changes in metabolism.

Evidence supporting this comes from studying the effects on gene transcription when the redox poise of the chloroplast changes.



## Transfer of Organelle DNA to the Nucleus Continues: Promiscuous DNA!

Comparative genomics has identified:

Copies of genes still in organelle genomes are also in other cellular compartments

- NUMTs – nuclear mitochondrial DNAs
- NUPTs – nuclear plastid DNAs

Range of sequence similarity indicates recurrent transfers – fragments with most similar sequences are most recent transfers

### EXAMPLES

#### yeast nuclear genome

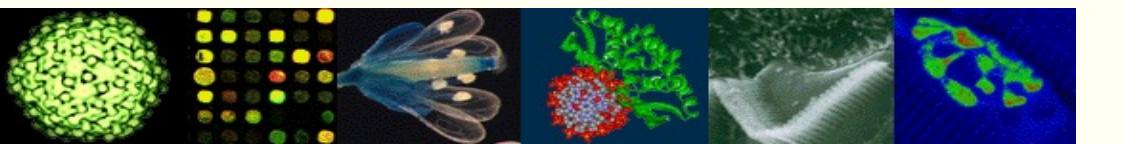
- fragments with 80-100% similarity to mitochondrial genes
- size range: 22 to 230 bp
- integrated at 34 sites

#### rice nuclear genome

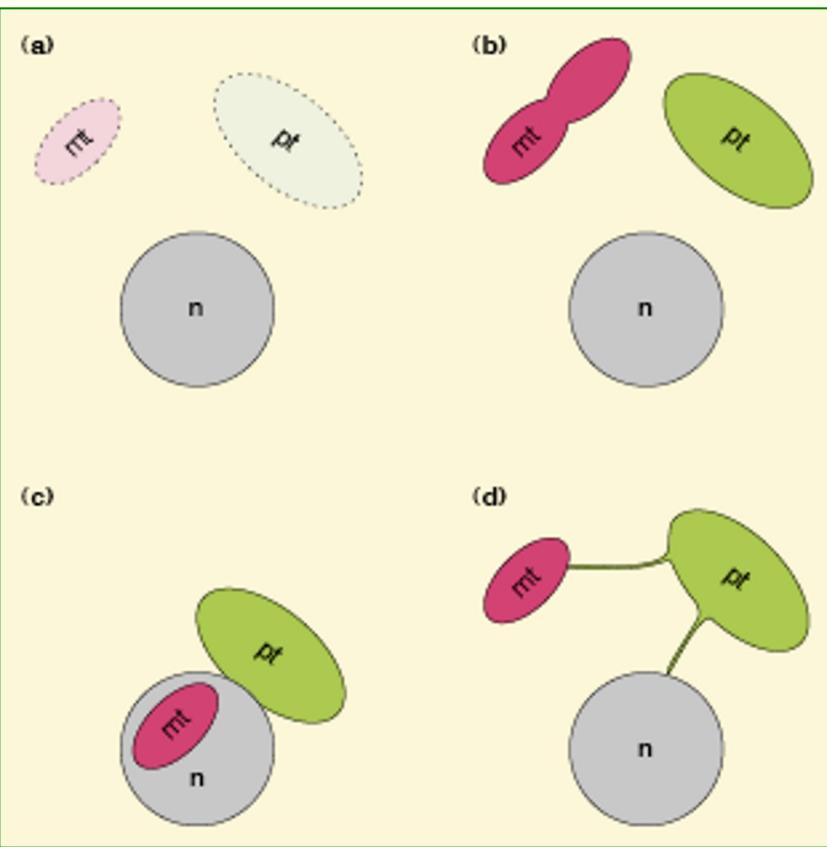
- chromosome 10 contains 33 kb insert of chloroplast DNA and a 131 kb insertion that represents nearly the entire chloroplast genome

#### human nuclear genome

- 59 fragments of mitochondrial DNA >2 kb in length totaling 280 kb
  - one 14.6 kb fragment encoding nearly the entire human mitochondrial genome

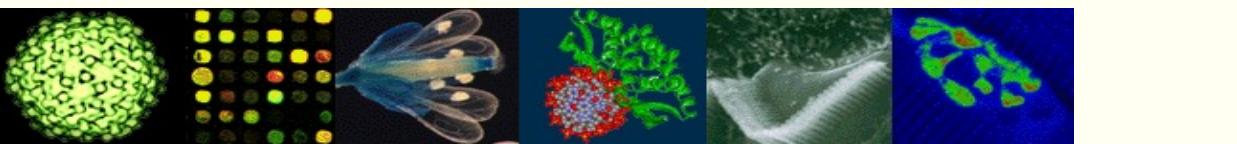


## How is DNA Transferred Between Intracellular Compartments?

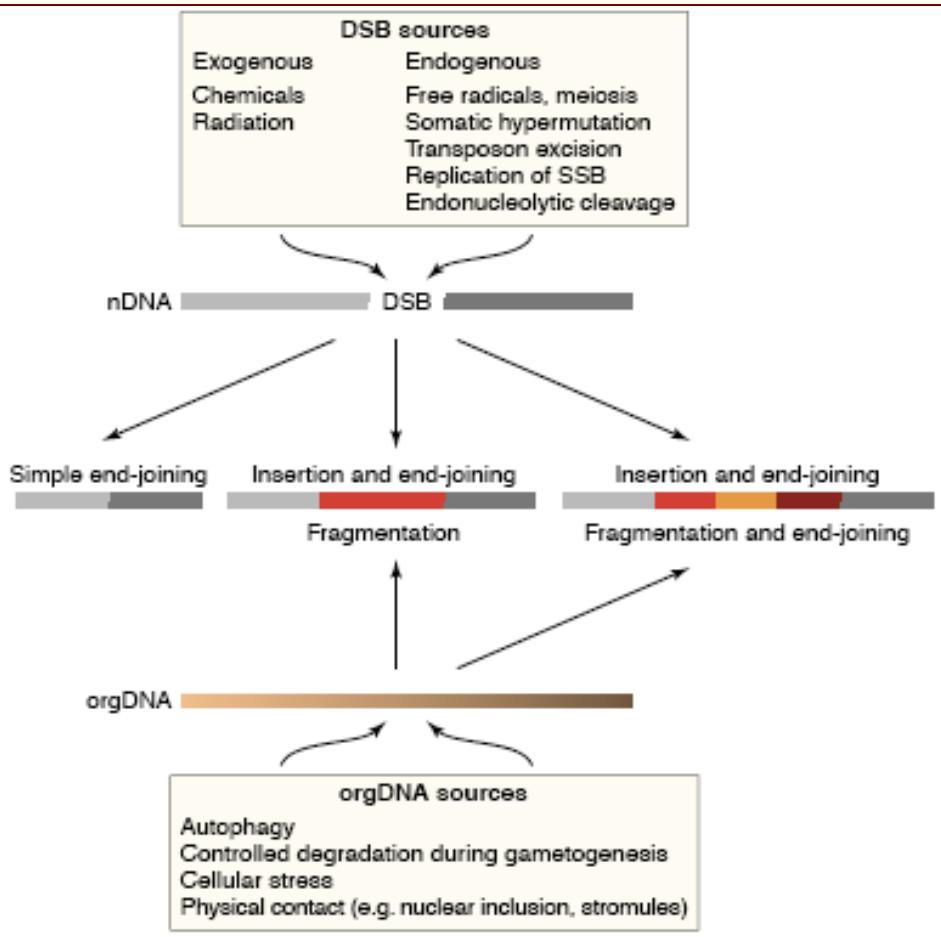


- a) Lysis of organelles and release of DNA
- b) Fusion of organelles (lateral gene transfer)
- c) Physical contact between organelles, or inclusion of mitochondria in nuclei
- d) Plastids (e.g. chloroplasts) form connections, called stromules, with other organelles

mt, mitochondrion; n, nucleus; pt, plastid



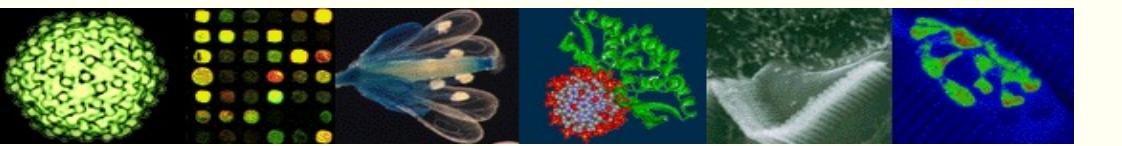
## How is Organelle DNA Integrated into the Nuclear Genome?



Nonhomologous end-joining (NHEJ): double-stranded-break (DSB) repair mechanism that requires no homology between ends (illegitimate repair)

Two general types of insertions identified:

- i) relatively long fragments of organelle DNA
- ii) mosaics of organelle DNA fragments



## Effects of Organelle DNA on Nuclear Integrity

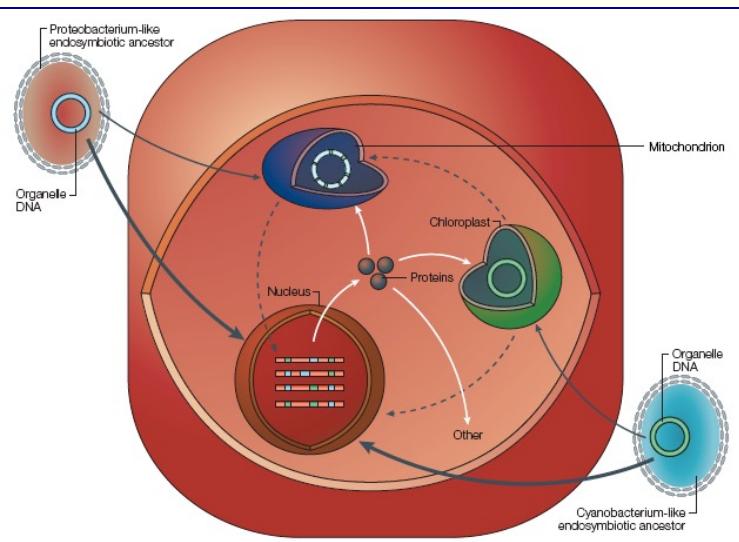
Most present-day transfers generate non-coding sequences (pseudogenes).

Some recent NUMTs show preferential insertion into genes – changes in exon-intron patterns.

- NUMTs associated with human disease, e.g. 251 bp insertion in coagulation factor VII gene – novel splice site, defective protein, bleeding disorder

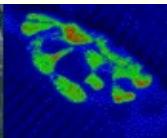
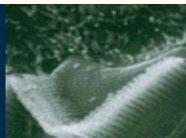
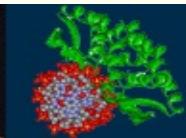
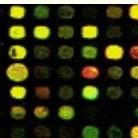
Insertions into non-protein-coding regions may affect gene regulation.

One study found ~40 nuclear genes across 4 species (yeast, rice, Arabidopsis and human) that have been remodeled by organelle DNA.



Three outcomes of remodeling observed:

- i) partial conservation of the original organelle ORF
- ii) replacement of the original ORF by a novel ORF
- iii) high rate of nonsynonymous substitutions – sequence diversification



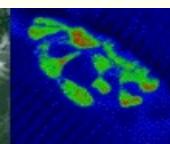
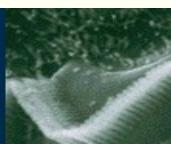
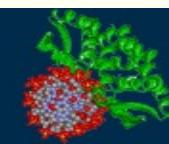
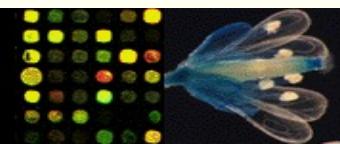
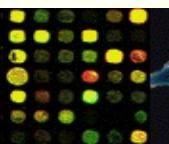
## Genome Evolution – Horizontal (Lateral) Gene Transfer

“The transmission of genetic material between the genomes of two individuals (that may belong to different species) by nonvertical inheritance” Feschotte & Pritham (07) Annu Rev Genet 41:331

### Examples

- endosymbiont genes transferred to host nucleus
- retroviruses in vertebrate genomes acquired through infection of the germ line
- prokaryote to prokaryote transmission of antibiotic resistance and virulence genes, and the genes for nitrogen fixation and photosynthesis
- acquisition of transposons

Few clear examples of genes transferred between unrelated multicellular eukaryotes



## Horizontal Gene Transfer – Photosynthesising Sea Slugs(!)

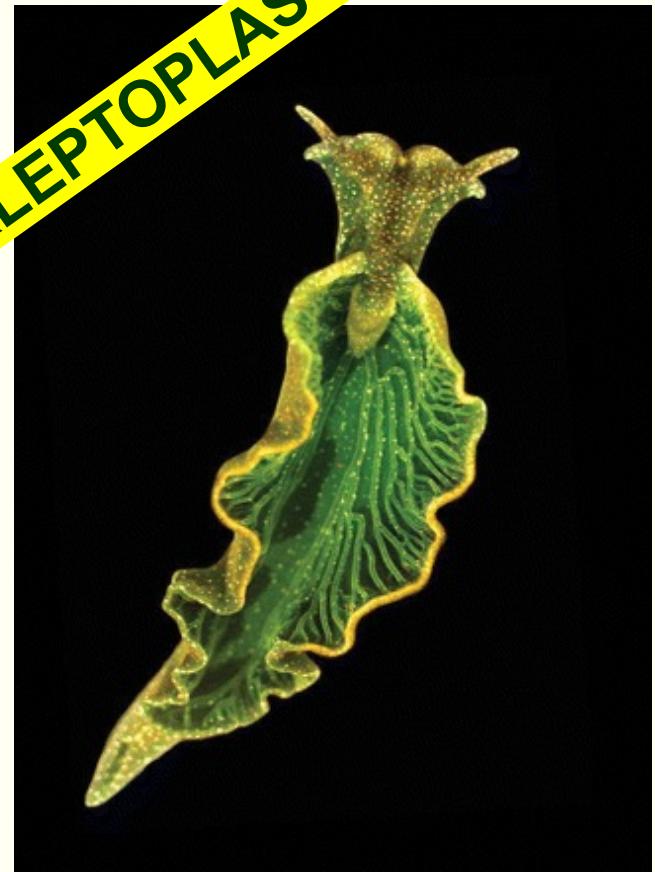
Sea slugs (*Elysia chlorotica*) acquire chloroplasts through feeding on an alga (*Vaucheria litorea*)

- the chloroplasts
  - inside cells lining the slug's gut
  - continue to photosynthesise, even without nuclei that normally provide >90% of their proteins
  - provide carbohydrates to slugs for  $\geq 10$  months
- numerous algal chloroplast and nuclear genes are expressed
  - transcriptome data

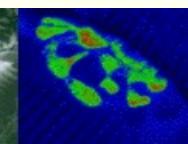
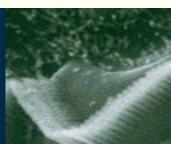
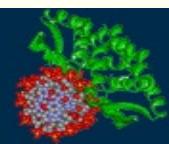
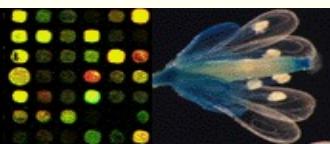
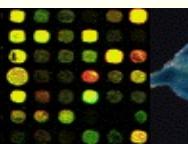
BUT sequencing of sea slug egg cell genome indicates no algal sequences in the germ line

A current hypothesis – algal genes exist as extrachromosomal DNA in feeding adult sea slugs

'KLEPTOPLASTY'



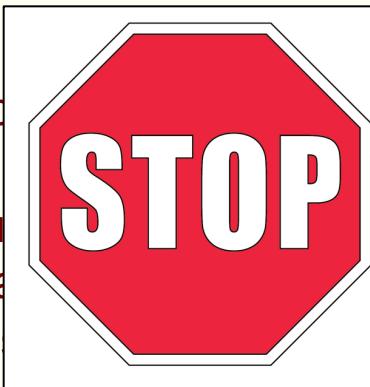
Rumpho et al., (2008) Proc Natl Acad Sci 105: 17867



## Horizontal Gene Transfer – Photosynthesising Sea Slugs(!)

Sea slugs (*Elysia chlorotica*) acquire chloroplasts through feeding on an alga (*Vaucheria litorea*)

- the chloroplasts
  - inside cells lining the slug's gut
  - continue to photosynthesise, even without nuclei that normally provide >90% of their proteins
  - provide carbohydrates to slugs for  $\geq 10$  months
- numerous algal chloroplast and nuclear genes are expressed
  - transcriptome data



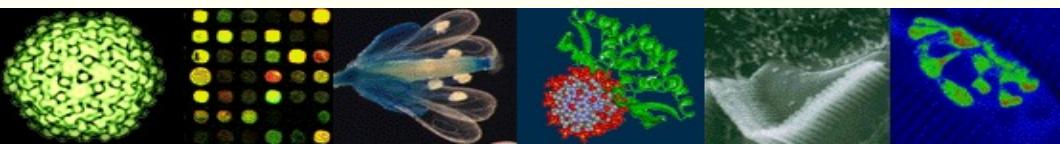
BUT  
indicates

A current  
extraordinary  
slug

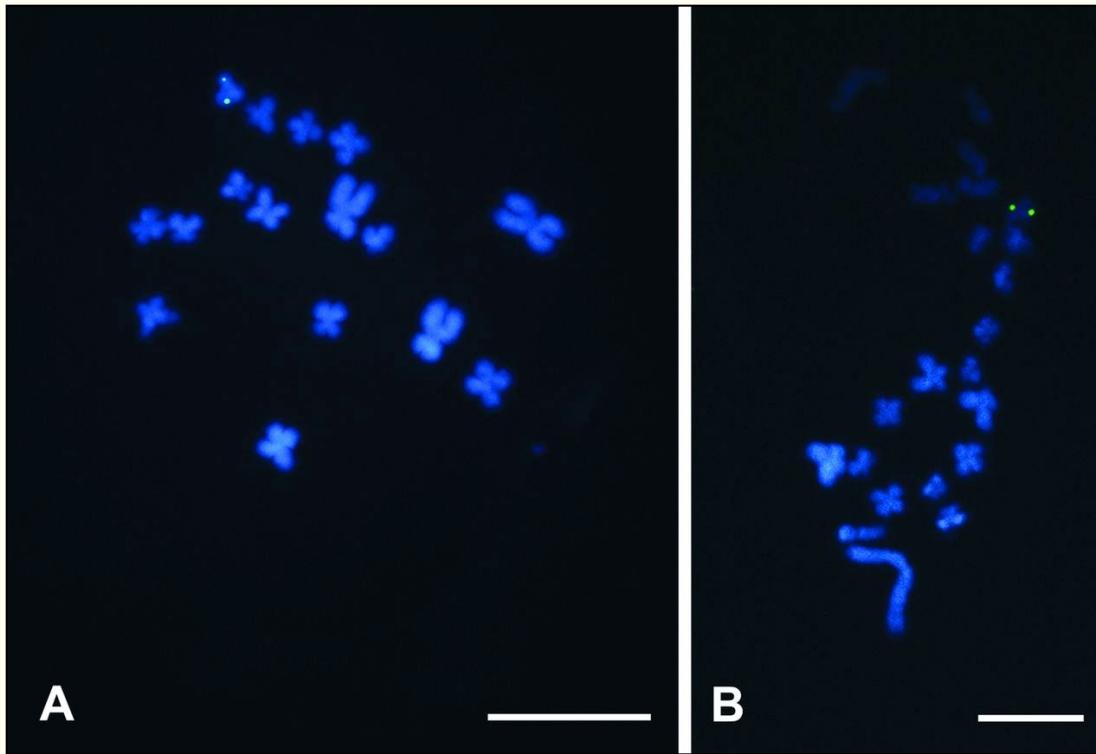
That was 2013. In 2014...



Jumpho et al., (2008) Proc Natl Acad Sci USA 105: 17867



## Horizontal Gene Transfer – Photosynthesising Sea Slugs(!) cont'd



Schwartz et al. (2014) Biol Bull 227: 300

*in situ* hybridsation of metaphase chromosome spreads from unhatched sea slug larvae (prior to feeding)

- DIG-labelled probe for *Vaucheria litorea prk* gene
  - *prk* = phosphoribulokinase – an enzyme in the CO<sub>2</sub> fixation pathway