E-content

M.Sc. Zoology (Semester II)
CC8- Biochemistry

Unit: 3.5a

Origin of new genes

Dr Gajendra Kumar Azad
Assistant Professor
Post Graduate Department of Zoology
Patna University, Patna
Email: gajendraazad@outlook.com

Background

The formation of new genes is a primary driving force of evolution in all organisms. It is evident that approx. 30,000 different functional genes in mammalian genomes must have evolved from a much lower number in the earliest ancestors of living organisms.

Recent research has focused on identifying the mechanisms that generate new genes, and scientists have found that these mechanisms involve a variety of molecular events, all of which must occur in the germ line to be inherited by the next generation.

After the germ-line mutational event, the new gene (such as a new gene duplicate located on human chromosome 2) will be polymorphic in the population; in other words, not all second chromosomes in the population will carry the duplication.

Subsequently, the two most likely outcomes for the new gene are fixation (i.e., the new gene will reach a frequency of 100%) or extinction (i.e., the new gene will be lost).

Mechanisms of New Gene Generation

There are several mechanisms by which new genes are generated. These includes:

1. Gene duplication
2. Exon shuffling
3. Mobile elements or transposable elements
4. lateral gene transfer
5. gene fusion/gene fission
6. *de novo* origination.

1. Gene Duplication

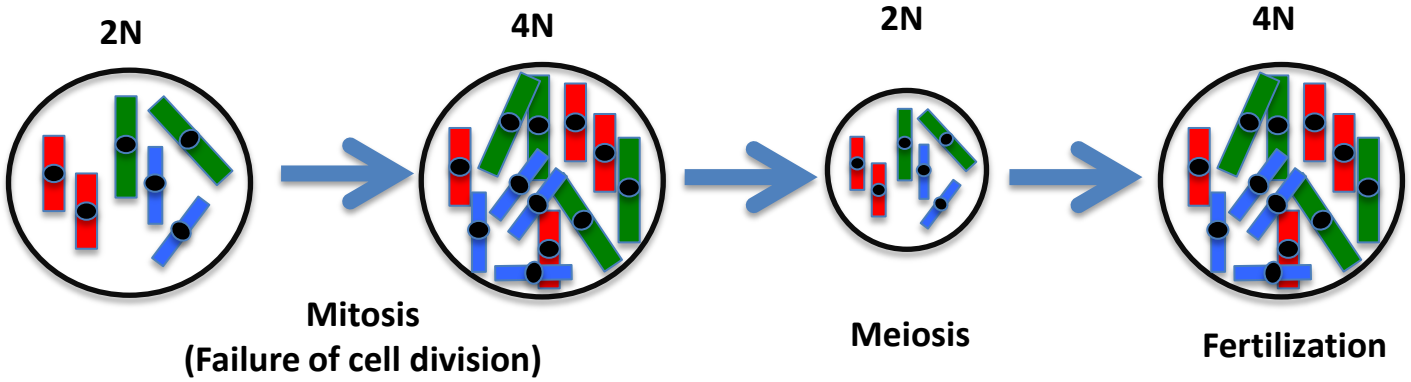
Gene duplication was the first mechanism of gene generation and this process is most common way of creating new genes.

Duplications are typically classified according to the size of the portion of the genome that is duplicated; thus, a duplication may be described as involving an entire genome, large segments of a genome, individual genes, individual exons, or even specific parts of exons.

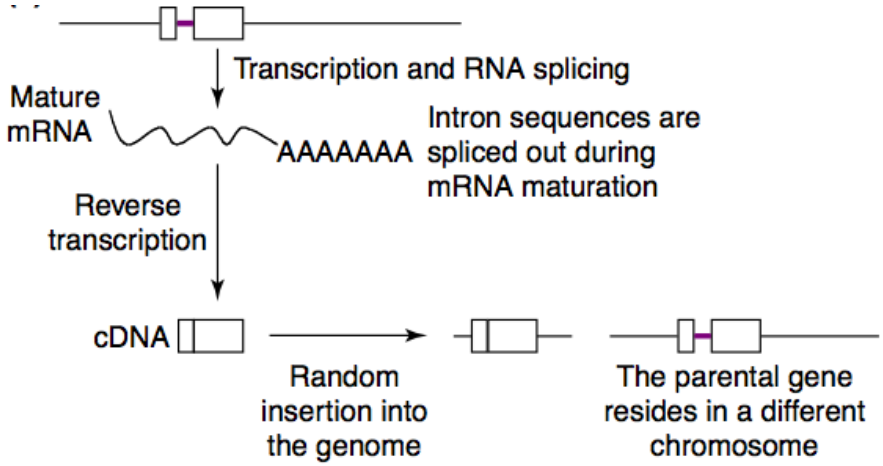
The mechanisms that generate duplicate genes are diverse. These mechanisms includes

a) whole-genome duplications originating through nondisjunction
b) retrotranspositions originating through the retrotranscription of an RNA intermediate
c) transpositions involving transposable elements
d) tandem duplications originating through unequal crossover
e) duplications occurring after rearrangements and subsequent repair of staggered breaks.

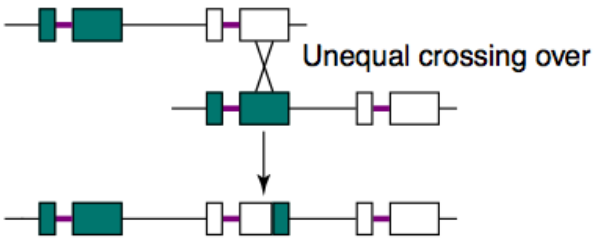
Mechanisms of Gene Duplication



whole-genome duplications



Retrotranspositions originating through the retrotranscription of an RNA intermediate

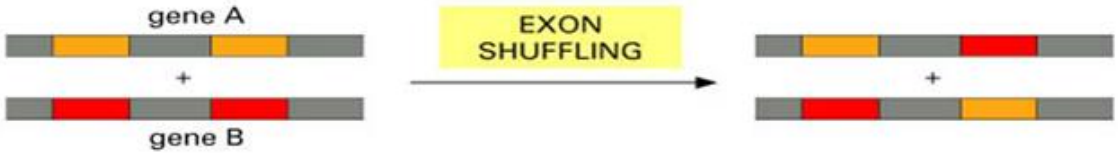


Tandem duplications originating through unequal crossover

Trends in Ecology & Evolution , 2003, 18: p. 292.

2. Exon shuffling

Two or more exons from different genes can be brought together ectopically, or the same exon can be duplicated, to create a new exon–intron structure.



Since exons are flanked by long introns then misalignment of introns can introduce exon dulpications. The exons of genes can generates an individual useful units that can be mixed and matched through exon shuffling to generate new, useful combinations. Duplication of exons leads to additional domains in the protein.

The 30,000 human genes are proposed to have arisen by duplication and shuffling of just a few thousand distinct exons. Domain complexity increases with organismal complexity

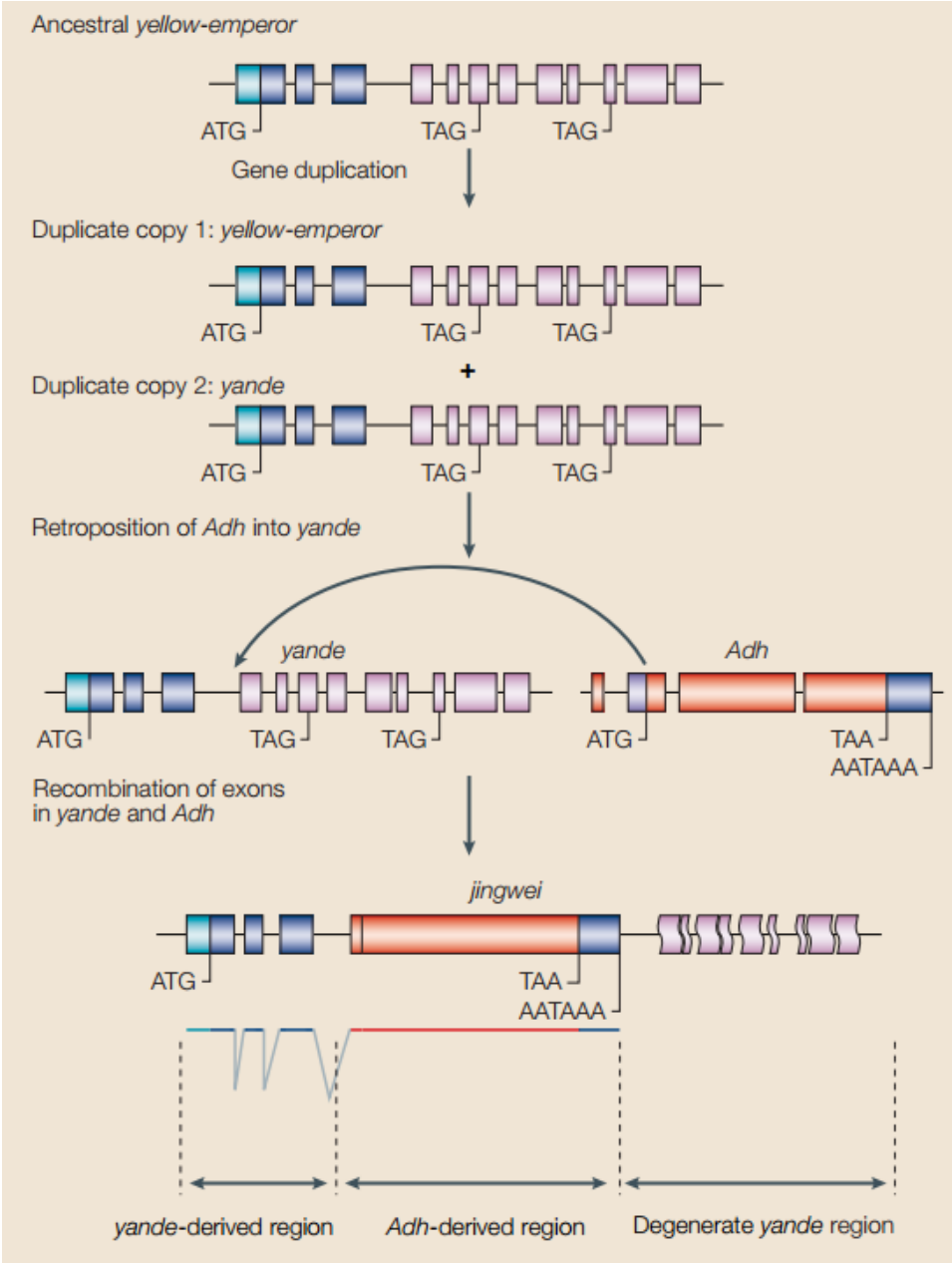
Rubin *et al.*, *Science, 2000*

Table 6. Proteins in *D. melanogaster*, *C. elegans*, and *S. cerevisiae* with multiple different InterPro domains. Individual InterPro domains are counted only once per protein, regardless of how many times they occur in that protein.

Unique InterPro domains per protein	*D. melanogaster* (number of proteins)	*C. elegans* (number of proteins)	*S. cerevisiae* (number of proteins)
2	1474	1248	402
3	413	335	95
4	156	114	23
5	52	38	4
6	8	9	1
7 or more	4	3	0

A. Example of formation of new gene by exon shuffling:
Origination of *jingwei* from the following two genes:

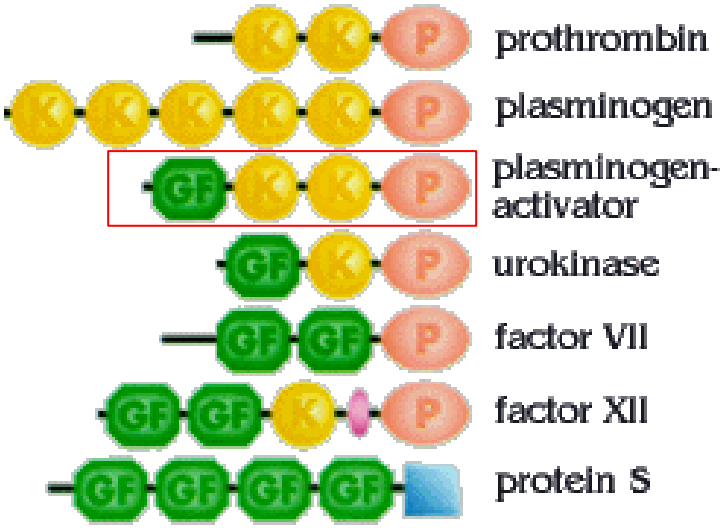
- *Yellow-emperor*
- Adh: a pseudogene



Long MY, *Science*. 1993
Long MY, *et al*. *Nat Rev Gent*. 2003

B. Formation of combinations of domains as a result of exon/domain shuffling.

Blood coagulation factors family members contain similar domains in various combinations and numbers.



P=protease domain,
GF=growth factor domain,
K="kringle"-domain

3. Transposable Elements

Transposable elements (TEs) are so-called "selfish" segments of DNA that can move themselves within a genome.

There are two types of TEs: DNA transposons and retrotransposons.

DNA transposons are able to excise themselves out of the genome and be inserted somewhere else, whereas retrotransposons copy themselves through an RNA intermediate.

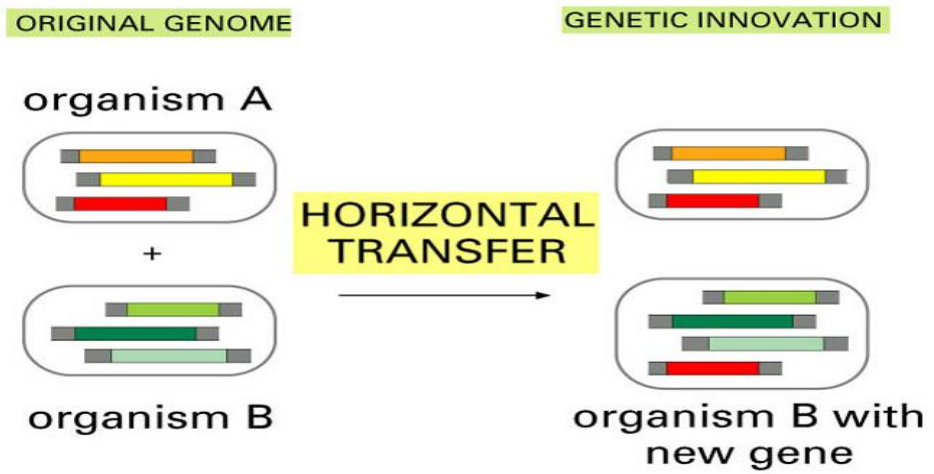
Transposons are mobile DNA elements akin to plasmids in bacteria. They are present in large numbers (500,000 Alu-like transposons in human genome) They are constantly moving around the genome.

TE insertions cause mutations and contribute to increased genome size, but they usually do not encode cellular proteins.

4. Lateral gene transfer (LGT)

Lateral gene transfer is the gene transferred between different organisms. Often in Prokaryotes, rare in eukaryotes.

The term "lateral gene transfer" to refer to the case in which a gene does not have a vertical origin (i.e., direct inheritance from parent to offspring) but instead comes from an unrelated genome. The Gene transfer across species is termed as LGT. It is also referred as Horizontal gene transfer.



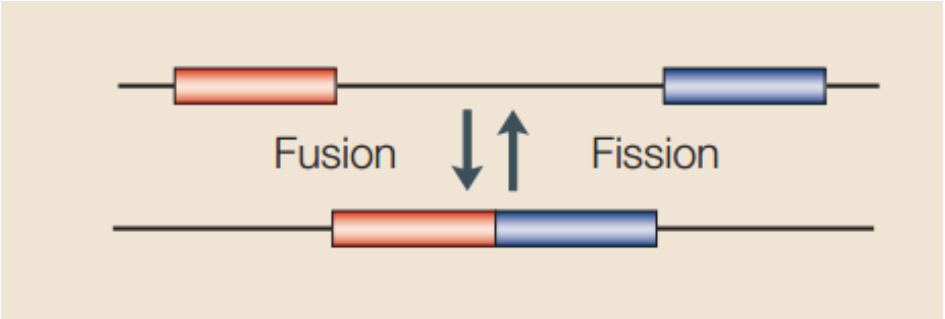
It is well known that this sort of transfer occurs between bacteria, and that it also has taken place between the genomes of the cellular organelles (mitochondria and chloroplasts) and the nuclear genomes.

Examples:

acytylneuraminate lysase, Escherichia coli, mutU and mutS

5. Gene Fusion and Fission

In this mechanism, the existing gene can fuse (i.e., two or more genes can become part of the same transcript) or may undergo fission (i.e., a single transcript can break into two or more separate transcripts), which results in the formation of a new gene (Long et al., 2003). The gene fusion or fission occurs due to the mutations on stop codon or initial codon of the gene.



Many cases of gene fusion and fission have been identified in prokaryotic genomes as well as in higher eukaryotes. An example of fusion gene in human is, KUA-UEV, in which the ubiquitin E2 variant domain of tumour susceptibility gene (UEV1) and a newly identified gene known as KUA were fused (Thomson et al., 2000). Other example is *Fatty-acid synthesis enzymes.*

Interestingly, it has been observed that chimeric fusion genes sometimes involve two copies of the same gene (e.g., the alcohol dehydrogenase gene in *Drosophila*), and when that happens, the resulting genes undergo parallel evolution, in which they shift away from the functions of their parental genes.

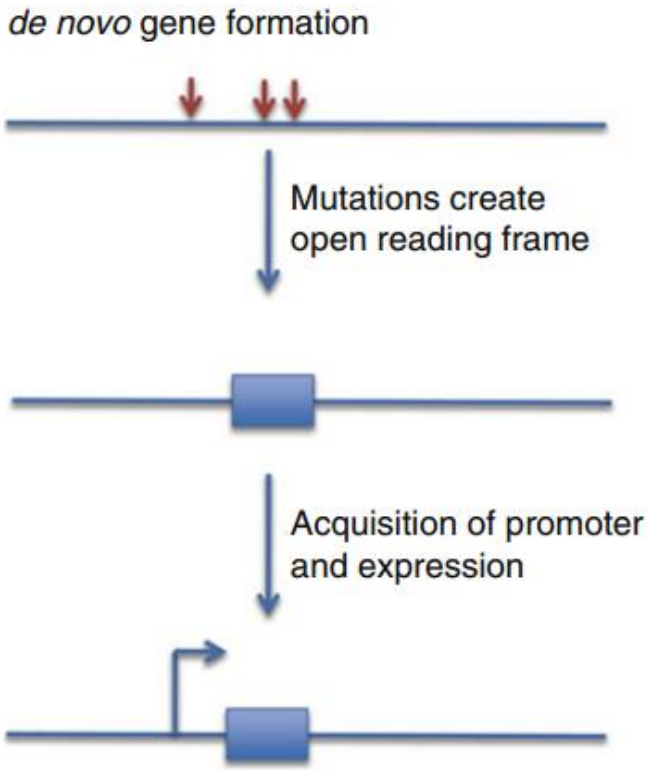
6. de novo origination

A new gene can form by the de novo origination process from the noncoding region of DNA molecule, which is neither transcribed, nor translated. In other words, *de novo* genes refer to events, where a coding region originates from a previously non-coding region.

Several novel genes derived from noncoding DNA have recently been described in *Drosophila*. For these recently originated *Drosophila* genes with likely protein-coding abilities, there are no homologues in any other species.

These new genes sometimes originate in the X chromosome, and they often have male germ-line functions.

Examples:
AFGPs, BC1RNA, BC200RNA



Long MY, *et al*. *Nat Rev Gent.* 2003
Cardoso-Moreira M and Long MY. 2004

Fate of newly formed Genes

All these new sequences add to the complexity and diversity of genomes. As with any mutation, when new genes become fixed in a genome, they add to the differences between species and serve as the raw material for evolution.

This is easy to see in the case of gene duplication. Gene duplication results in two or more copies of a gene: one that can maintain its original function in the organism, and other(s) that can be "played with" to take on new functions.

As a consequence, new duplicates are a main source of genome innovation and often evolve under positive selection, in which rapid changes in the protein encoded by the new gene occur to gain a new function. This process is referred to as neofunctionalization of the new gene.

Example of Neofunctionalisation:
-*GLUD2* in primates has a new role in neurotransmitter flux.
-Thrombin (cleaves fibrinogen during clotting) and trypsin (digestive enzyme) are derived from a complete gene duplication.
-Lactate dehydrogenase can be converted into malate dehydrogenase with a single amino acid replacement (out of total protein length of 317 amino acids)

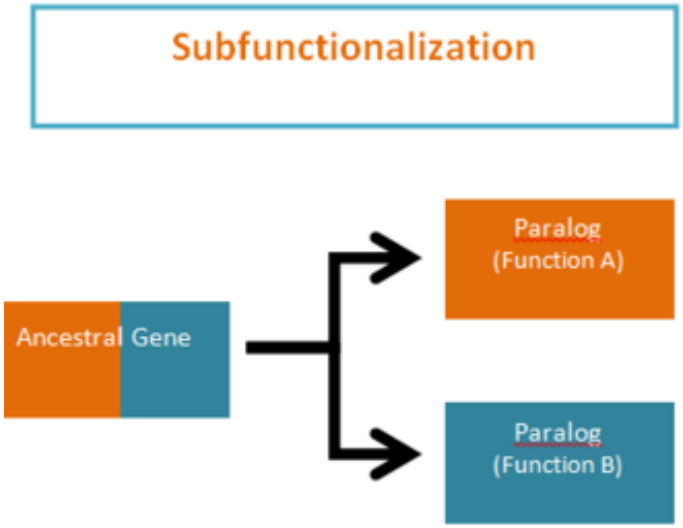
Fate of newly formed Genes

Other possible fate of newly formed genes are:
1. Functional compensation: Many duplicated genes shelter the organism from deleterious mutations in the other copy (shown in yeast and worm)

2. gene loss or pseudogenization:

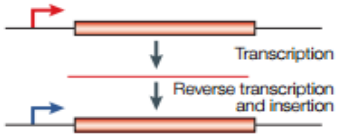
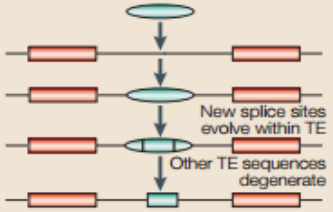
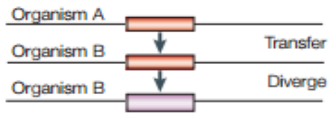
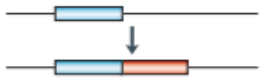
3. maintenance of both genes as a way to increase expression (Dosage increase) or to maintain multiple variants within individuals. Example: Esterase B in mosquito-increased gene dosage confers greater pesticide resistance.

4. the occurrence of subfunctionalization: Subfunctionalization is a neutral mutation process in which each paralog retains a subset of its original ancestral function. The figure illustrates that the ancestral gene (orange & blue) is capable of both functions before gene duplication. After gene duplication the functional capabilities are divided amongst the gene copies. After this divergence each paralog is capable of independently performing a distinct ancestral function.



Summary

Table 1 | **Molecular mechanisms for creating new gene structures**

Mechanism	Process	Examples	Comments
Exon shuffling: ectopic recombination of exons and domains from distinct genes		*fucosyltransferase*, *jingwei*, *Tre2*	~19% of exons in eukaryotic genes have been formed by exon shuffling
Gene duplication: classic model of duplication with divergence		*CGβ*, *Cid* , *RNASE1B*	Many duplicates have probably evolved new functions
Retroposition: new gene duplicates are created in new genomic positions by reverse transcription or other processes		*PGAM3*, *Pgk2*, *PMCHL1*, *PMCHL2*, *Sphinx*	1% of human DNA is retroposed to new genomic locations
Mobile element: a mobile element, also known as a transposable element (TE), sequence is directly recruited by host genes		*HLA-DR-1*, human *DAF*, *lungerkine* mRNA, *mNSC1* mRNA	Generates 4% of new exons in human protein-coding genes
Lateral gene transfer: a gene is laterally (horizontally) transmitted among organisms		*acytylneuraminate lysase*, *Escherichia coli* *mutU* and *mutS*	Most often reported in prokaryotes and recently reported in plants
Gene fusion/fission: two adjacent genes fuse into a single gene, or a single gene splits into two genes		Fatty-acid synthesis enzymes, *Kua-UEV*, *Sdic*	Involved in the formation of ~0.5% of prokaryotic genes
De novo origination: a coding region originates from a previously non-coding genomic region		*AFGPs*, *BC1RNA*, *BC200RNA*	Rare for whole gene origination; might not be rare for partial gene origination

Long MY, *et al*. *Nat Rev Gent.* 2003

References

Chandrasekaran C, *et al*. *Nature. 2008*

Long MY, *et al*. *Nat Rev Gent.* 2003

Cardoso-Moreira M and Long MY. 2004

Evolution by Douglas J. Futuyma

---------- End ----------