Journal of Pharmacognosy and Phytochemistry

Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
JPP 2019; 8(2): 2394-2405
Received: 21-01-2019
Accepted: 25-02-2019

Sowjanya BA
Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwad, Karnataka, India

Narayana BD
Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwad, Karnataka, India

Shreyas S
Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwad, Karnataka, India

Correspondence
Sowjanya BA
Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwad, Karnataka, India

Origin and evolution of new genes
Review paper

Sowjanya BA, Narayana BD and Shreyas S

Abstract
Evolution is the change in the inherited characteristics of biological populations over successive generations. Charles Darwin was the first to formulate a scientific argument for the theory of evolution by means of natural selection. Available molecular techniques and rapidly expanding genome data from many organisms revealed great variation in the number of gene across the organisms. Several mechanisms are known to be involved in the origin of new genes such as gene duplication- duplication of chromosomal segments containing whole genes or gene fragments. Retro position-new gene duplicates created in new genomic positions by reverse transcription. Exon shuffling-exons from different genes brought together ectopically. Lateral gene transfer-process by which an organism incorporates genetic material from another organism without being a direct descendant of that itself. Gene fusion-fusion of two previously separate source genes into a single transcription unit. De-novo originationof new genes from previously non-coding sequences.And combined mechanisms-new genes can be created by the mechanisms discussed above, either individually or in combination**.** Among which transposable elements play a key role in the origin and evolution of both protein –coding genes and non-coding RNA sequences.

Keywords: Evolution, new genes

Introduction
Genetic modifications of preexisting ancestral genes that can lead to differences in their (Protein or RNA) sequences or activities, new genes with novel functions have significantly contributed to the evolution of lineage or species specific phenotypic traits. It helps in underlying of adaptive evolutionary innovations. Consequently, the process of the ‘‘birth’’ and evolution of novel genes has attracted much attention from biologists in the past. Based on cytological observations of chromosomal duplications, Haldane (1933) and Muller (1935) already hypothesized in the 1930s that new gene functions may emerge from refashioned copies of old genes, high lighting for the first time the potential importance of gene duplication for the process of new gene origination. Previous efforts to study the origin of new genes have been sporadic and have focused on the evolution of duplicate and chimeric genes, even though these are often hundreds of millions of years old. Analyses of genes that have been identified by such an approach have provided some exciting insights. However, a more efficient approach is the direct observation of young genes when they are at an early stage in their evolution. The study of ancient genes has established the antiquity of some of the molecular mechanisms used to generate new genes. Studies from the genomics era will accelerated the discovery of fascinating novel mechanisms underlying the emergence of new genes. These include the origin of new protein coding and RNA genes ‘‘from scratch’’ (that is, from previously nonfunctional genomic sequences), various types of gene fusions, and the formation of new genes from RNA intermediates. All mechanisms have significantly contributed to functional genome evolution and phenotypic change, which further underscores the importance of novel genes for organismal evolution.
Several molecular mechanisms are known to be involved in the creation of new gene structures they are

1. Mutation
2. Mobile elements(Retro transposition)
3. Gene duplication
4. Exon shuffling
5. Horizontal gene transfer
6. Gene fusion
7. Gene origination from scratch

8. De novo origination
9. Combined mechanisms
10. Origin of RNA gene

Mutation

Mutation is a change in phenotype, which is sudden, heritable and is not produced due to segregation or recombination. Mutation is the ultimate source of all the genetic variation existing in any organism.

In the history of evolutionary biology, Hugo de Vries is known as a proponent of the mutation theory of evolution, in which new species are believed to arise by single mutational events. This theory is based on the breeding experiment he conducted for 13 years with the evening primrose *Oenothera lamarckiana* and its mutant descendants.

Small-scale mutations affecting one or a few nucleotides include:

1. Point Mutations - Substitution of one nucleotide for another
2. Insertions -Addition of one or more nucleotides
3. Deletions- Removal of one or more nucleotides

Point mutations occurring within a protein-coding region of the genes may be classified into three kinds, depending upon what the altered codon codes for.

(i) Synonymous (or silent) mutations: code for the same amino acid

(ii) Nonsynonymous (or Missense) mutations: code for a different amino acid

(iii) Nonsense mutations: code for a stop codon, can truncate the protein

Synonymous Substitutions (K_S) as a Proxy for Evolutionary Time

Synonymous (silent) substitutions are thought to be largely neutral or invisible to natural selection because they do not change the amino acid sequence. Fraction of synonymous substitutions (K_S) for a pair of sequences approximates the time since divergence.

	Met	Phe	Arg	Ser	Pro	Thr
Duplicate Copy 1	AUG	UUU	CG**A**	U**C**C	CC**G**	ACC
Duplicate Copy 2	AUG	UUU	CG**U**	U**G**C	CC**C**	ACC
	Met	Phe	Arg	Cys	Pro	Thr

K_S = Fraction of synonymous substitutions per synonymous site = 2/5 = 0.40 or 40%.

K_A = Fraction of nonsynonymous substitutions per nonsynonymous site = 1/13 = 0.077 or 7.7%.

This helps in understanding of role of synonymous mutation in evolutionary time, It explain the neutral theory of evolution, which says that only small amount of mutations (Atnon synonymous) have role in the evolution. Those mutations which are fixable and altered form in the population they can able to formation of new genes in future. But mutation at synonymous site doesn't have any significant role in the evolution and origin of new genes, they remain as neutral in the evolution. In the above example fraction of synonymous substitution is maximum compared to the non-synonymous substitution which clearly gives an idea about very small amount of mutation have role in the evolution.

Mobile elements (Retro transposition)

Mobile elements found in most eukaryotic genomes in humans – Alu (SINE), (LINEs), which contribute to genome evolution in several ways. They are,

- Exon shuffling
- Insertion mutagenesis
- Homologous and non-homologous recombination

Makalowski *et al*. (1994) [6] were the first to describe the integration of an Alu element into the coding portion of the human decay-accelerating factor (DA) gene. They found that mobile element derived diversity was not limited to the human genome or to the Alu family15 (TABLE 1). Further analyses of human genome sequences16 and vertebrate genes17 have shown that the integration of mobile elements into nuclear genes to generate new functions is a general phenomenon.

3.1 Retro transposition

This mechanism creates duplicate genes in new genomic positions through the reverse transcription of expressed parental genes (TABLE 1). As a retroposed gene copy does not usually retropose a promoter copy from its parental gene, it has to recruit a new regulatory sequence to be functional or it will die out as a processed pseudogene. So, a functional retroposed gene has a chimeric structure either a retroposed coding region with a new regulatory sequence or a retroposed coding region with a new protein fragment that is recruited from the targeted site that leads to it having a different function to its parental gene. In mammals, the L1 retro-element is responsible for retroposing nuclear genes.

The observation of numerous functional retrogenes in various genomes immediately raises the question of how retrocopies can obtain regulatory sequences that allow them to become transcribed a precondition for gene functionality. Studies that sought to address this question uncovered various sources of retrogene promoters and regulators and therefore also provided general insights into how new genes can acquire promoters and evolve new expression patterns (Kaessmann *et al*. 2009) [4].

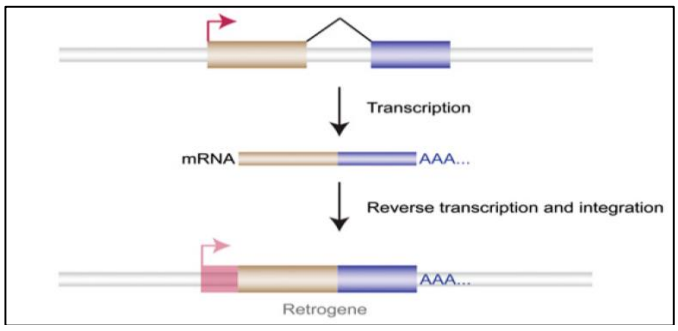

Fig 1: RNA based duplication (termed retroposition or retroduplication). New retroposed gene copies may arise through the reverse transcription of messenger RNAs (mRNAs) from parental source genes. Functional retrogenes with new functional properties may evolve from these copies after acquisition or evolution of promoters in their 5' flanking regions that may drive their transcription. (Pink rightangled arrow) transcription start site, (transparent pink box) additionally transcribed flanking sequence at the insertion site.

4. Gene duplication

This classical model creates a duplicate gene that can evolve new functions, whereas the ancestral copy maintains its original functions (TABLE 1). Many new gene functions have evolved through gene duplication and it has contributed

tremendously to the evolution of developmental programmes in various organisms. Also, duplications at the segmental chromosomal and genome levels, which are abundant in plants, have been shown to contribute to the evolution of new functions in humans. DNA mediated duplication mechanisms include small-scale events, such as the duplication of chromosomal segments containing whole genes or gene fragments (termed segmental duplication), which are essentially outcomes of misguided recombination processes during meiosis. However, they also include duplication of whole genomes through various polyploidization mechanisms.

4.1 Mechanisms of duplication

DNA-based duplication

1. Ectopic recombination,
2. Aneuploidy
3. Polyploidy (whole genome duplication)
4. Replication slippage

   RNA-based duplication
5. Retrotransposition events

4.1.1 Ectopic recombination

Ectopic recombination is an aberrant form of recombination in which crossing over occurs at non-homologous, rather than along homologous, loci. Such recombination often results in dramatic chromosomal rearrangement, which is generally deleterious. Ectopic recombination does not require loci involved to be close to one another; it can occur between loci that are widely separated on a single chromosome, and has even been known to occur across chromosomes. Main factor affecting the frequency of this cross over event is the zygosity of the alleles in the loci of interest. If the alleles are heterozygous, then ectopic recombination is relatively likely to occur, whereas if the alleles are homozygous, they will almost certainly undergo allelic recombination. The role of transposable elements in ectopic recombination is an area of active inquiry. Transposable elements repetitious sequences of DNA that can insert themselves into any part of the genome can encourage ectopic recombination at repeated homologous sequences of nucleotides.

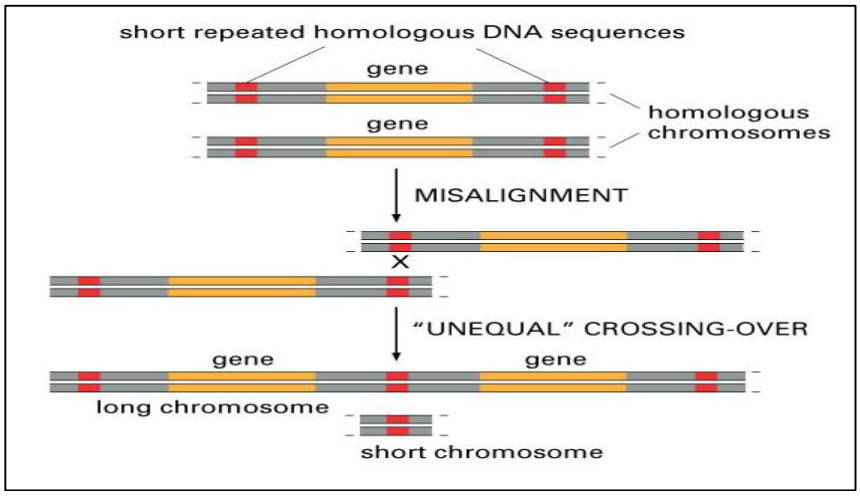

Fig 4: Mechanism of ectopic recombination.

4.1.2 Aneuploidy

Aneuploidy is a condition in which the chromosome number is not an exact multiple of the number characteristic of a particular species. Aneuploidy originates during cell division when the chromosomes do not separate properly between the two cells. This generally happens when cytokinesis begins while karyokinesis is still under way. *Non-disjunction* usually occurs as the result of a weakened mitotic checkpoint, as these checkpoints tend to arrest or delay cell division until all components of the cell are ready to enter the next phase. If a checkpoint is weakened, the cell may fail to 'notice' that a chromosome pair is not lined up on the mitotic plate, for example. In such a case, most chromosomes would separate normally (with one chromatid ending up in each cell), while others could fail to separate at all. This would generate a daughter cell lacking a copy and a daughter cell with an extra copy. Aneuploidy alters gene dosage which is detrimental to the organism so that it is unlikely to spread through populations. Plants are able to tolerate aneuploidy better than animals.

4.1.3 Polyploidy (whole genome duplication)

The chromosome number is an exact multiple of the basic or genome number of the species, these variation is called polyploidy. In Autopolyploid Genome duplication in one species, hybridization and duplication of the genomes of two different species (different species) is allopolyploids. It is a product of non-disjunction during meiosis which results in additional copies of the entire genome. Polyploidy is also a source of speciation because polyploidy species to interbreed with non-polyploidy organisms. Allopolyploids occurs widely in various genera of plants, it is estimated that about the one-third of angiosperms are polyploidy and vast majority of them are allopolyploids. Allopolyploids have been more successful as crop species than autopolyploid, and many of our present-day crops are allopolyploids.

4.1.4 Replication Slippage

Replication slippage, otherwise known as slipped-strand mispairing, is a form of mutation which leads to either a trinucleotide or dinucleotide expansion or contraction during DNA replication. A slippage event normally occurs when a sequence of repetitive nucleotides (tandem repeats) are found at the site of replication. Tandem repeats are unstable regions of the genome where frequent insertions and deletions of nucleotides can take place, resulting in genome rearrangements. DNA polymerase, the main enzyme to catalyze the polymerization of free deoxy ribonucleotides into a newly forming DNA strand, plays a significant role in the occurrence of this mutation. However, when DNA polymerase encounters a direct repeat, it can undergo a replication slippage.

Slippage occurs through five main stages:

1. In the first step, DNA polymerase encounters the direct repeat during the replication process.
2. The polymerase complex suspends replication and is temporarily released from the template strand.
3. The newly synthesized strand then detaches from the template strand and pairs with another direct repeat upstream.

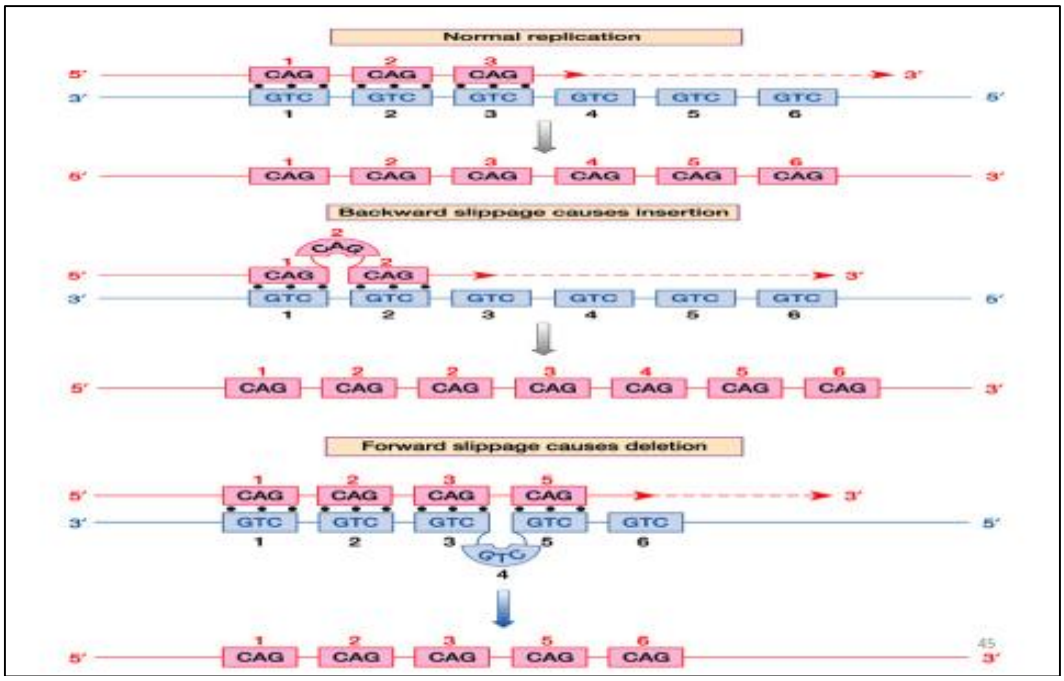

Fig 5: Mechanism of replication slippage.

4. DNA polymerase reassembles its position on the template strand and resumes normal replication, but during the course of reassembling, the polymerase complex backtracks and repeats the insertion of deoxy ribonucleotides that were previously added. This results in some repeats found in the template strand being replicated twice into the daughter strand. This expands the replication region with newly inserted nucleotides. The template and the daughter strand can no longer pair correctly.
5. Nucleotide excision repair proteins are mobilized to this area where one likely outcome is the expansion of nucleotides in the template strand while the other is the absence of nucleotides. Although trinucleotide contraction is possible, trinucleotide expansion occurs more frequently.

4.2 Evolutionary Fates of Duplicate Genes

4.2.1 The Haldane Model

Gene duplication creates redundancy, which in turn enables functional diversification and adaptation. The classical view of the fate of gene duplications dates back to the work of J. B. S. Haldane and R. A. Fisher. They believed that, in the presence of recurrent mutation, one member of a duplicate pair eventually becomes nonfunctional; that is, most duplicates should eventually die out as pseudogenes. And other gene undergo neofunctionalization by acquiring new function.

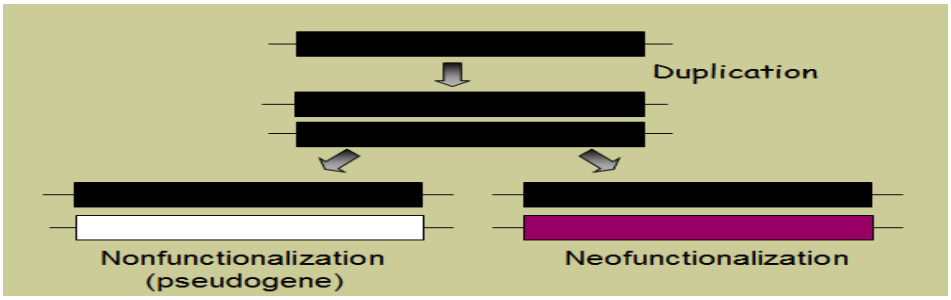

Fig 6: Evolutionary Fates of Duplicate Genes Haldane Model.

4.2.2 DDC (duplication-degeneration-complementation) model

In this model new duplicate gene in addition to pseudoginization and neo-functionalization, gene undergo sub-functionalization. The process of partitioning the ancestral functions of a locus among its duplicates. For example, if a single-copy gene that is normally expressed in two tissues subsequently duplicates, and each duplicate is then expressed in a different tissue, sub-functionaliztion has occurred.

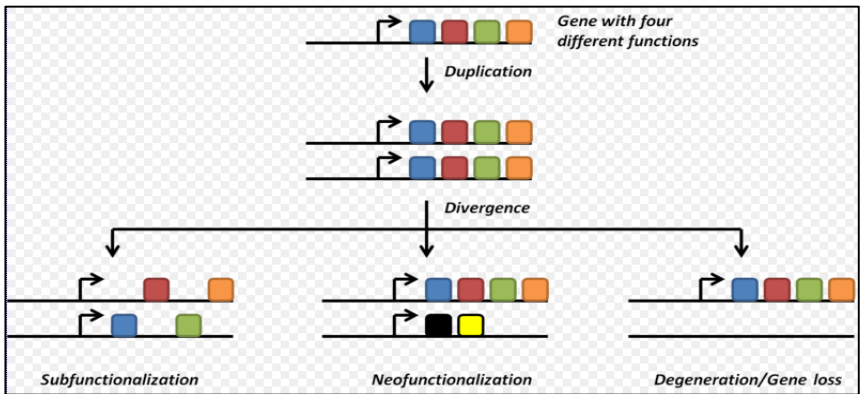

Fig 7: Evolutionary Fates of Duplicate Genes DDC (duplication-degeneration- complementation) model.

5. Exon shuffling

Exon shuffling is a molecular mechanism for the formation of new genes, first introduced Walter Gilbert (1978). It is a process through which 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.

5.1 Mechanisms of Exon shuffling

1. Crossover during sexual recombination of parental genomes
2. Illegitimate recombination
3. Long interspersed element (LINE)-1mediated exon shuffling
4. Helitron transposon mediated exon shuffling

5.1.1 Crossover during sexual recombination of parental genomes

Evolution of eukaryotes is mediated by sexual recombination of parental genomes and since introns are longer than exons most of the crossovers occur in noncoding regions. In these introns there are large numbers of transposable elements and repeated sequences which promote recombination of nonhomologous genes. In addition it has also been shown that mosaic proteins are composed of mobile domains which have spread to different genes during evolution and which are capable of folding themselves. There is a mechanism for the formation and shuffling of said domains, this is the modularization hypothesis. This mechanism is divided into three stages. The first stage is the insertion of introns at positions that correspond to the boundaries of a protein domain. The second stage is when the "protomodule" undergoes tandem duplications by recombination within the inserted introns. The third stage is when one or more protomodules are transferred to a different nonhomologous gene by intronic recombination.

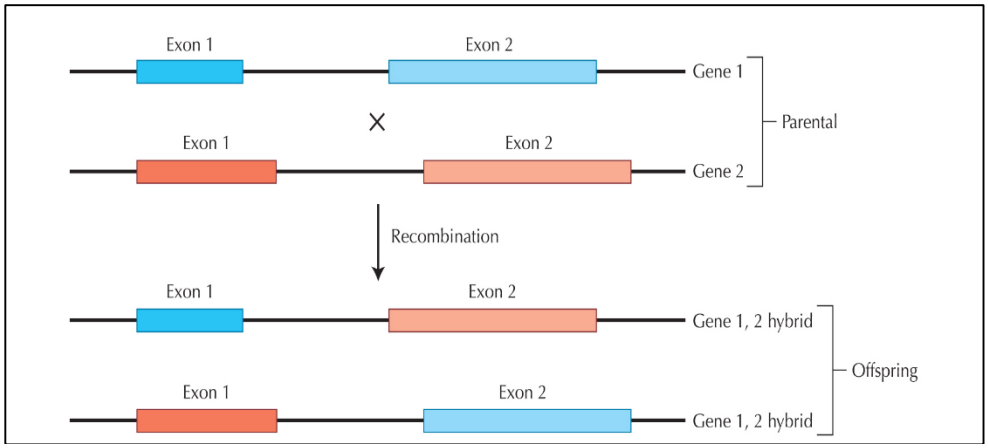

Fig 8: Exon shuffling occurs when exons from different genes are mixed and matched by recombination in the region between the exons.

5.1.2 Illegitimate recombination

Illegitimate recombination (IR) is another of the mechanisms through which exon shuffling occurs. IR is the recombination between short homologous sequences or nonhomologous sequences. There are two classes of IR: The first corresponds to errors of enzymes which cut and join DNA (i.e., DNases.) This process is initiated by a replication protein which helps generate a primer for DNA synthesis. While one DNA strand is being synthesized the other is being displaced. This process ends when the displaced strand is joined by its ends by the same replication protein. The second class of IR corresponds to the recombination of short homologous sequences which are not recognized by the previously mentioned enzymes. However, they can be recognized by non-specific enzymes which introduce cuts between the repeats. The ends are then removed by exonuclease to expose the repeats. Then the repeats anneal and the resulting molecule is repaired using polymerase and ligase.

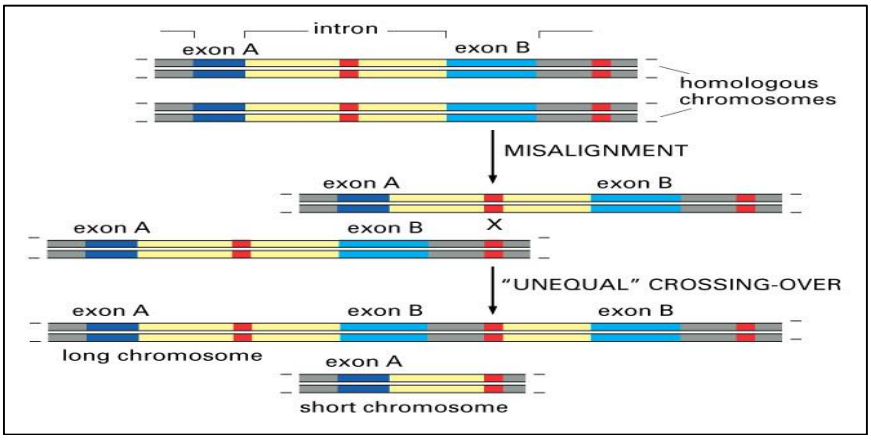

Fig 9: Mechanism of illegitimate recombination mediated exon shuffling.

5.1.3 Long interspersed element (LINE)-1 mediated exon shuffling

A potential mechanism for exon shuffling is the long interspersed element (LINE) -1 mediated 3' transduction. However it is important first to understand what LINEs are. LINEs are a group of genetic elements that are found in abundant quantities in eukaryotic genomes. LINE-1 is the most common LINE found in humans. It is transcribed by RNA polymerase II to give an mRNA that codes for two proteins: ORF1 and ORF2, which are necessary for transposition. Upon transposition, L1 associates with 3' flanking DNA and carries the non-L1 sequence to a new genomic location. This new location does not have to be in a homologous sequence or in close proximity to the donor DNA sequence. The donor DNA sequence remains unchanged throughout this process because it functions in a copy-paste manner via RNA intermediates; however, only those regions located in the 3' region of the L1 have been proven to be targeted for duplication.

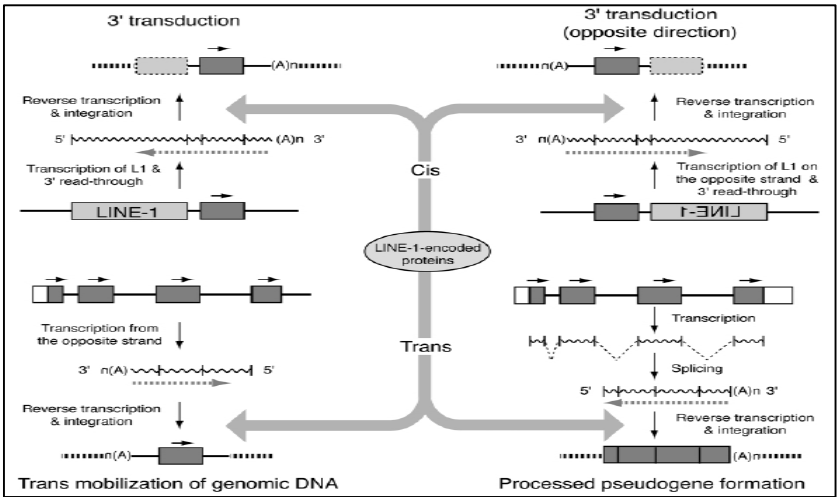

Fig 10: A model of how L1 Retrotransposition can mobilize sequences. At the top is the model for the cis pathway. At the bottom is the trans pathway.

5.1.4 Helitron transposons mediated exon shuffling

Another mechanism through which exon shuffling occurs is by the usage of helitrons. Helitron transposons were first discovered during studies of repetitive DNA segments of rice, worm and the thale crest genomes. Helitrons have been identified in all eukaryotic kingdoms, but the number of copies varies from species to species. Helitron encoded proteins are composed of a rolling-circle (RC) replication initiator (Rep) and a DNA helicase (Hel) domain. The Rep domain is involved in the catalytic reactions for endonuclelytic cleavage, DNA transfer and ligation. In addition this domain contains three motifs. The first motif is necessary for DNA binding. The second motif has two histidines and is involved in metal ion binding. Lastly the third motif has two tyrosines and catalyzes DNA cleavage and ligation. There are three models of gene capture by Helitrons: the 'read-through" model 1 (RTM1), the 'read-through" model 2 (RTM2) and a filler DNA model (FDNA). According to the RTM1 model an accidental "malfunction" of the replication terminator at the 3' end of the Helitron leads to transposition of genomic DNA. It is composed of the read-through Helitron element and its downstream genomic regions, flanked by a random DNA site, serving as a "de novo" RC terminator. According to the RTM2 model the 3' terminus of another Helitron serves as an RC terminator of transposition. This occurs after a malfunction of the RC terminator. Lastly in the FDNA model portions of genes or non-coding regions can accidentally serve as templates during repair of ds DNA breaks occurring in helitrons. Even though helitrons have been proven to be a very important evolutionary tool, the specific details for their mechanisms of transposition are yet to be defined. An example of evolution by using helitrons is the diversity commonly found in maize. Helitrons in maize cause a constant change of genic and nongenic regions by using transposable elements, leading to diversity among different maize lines.

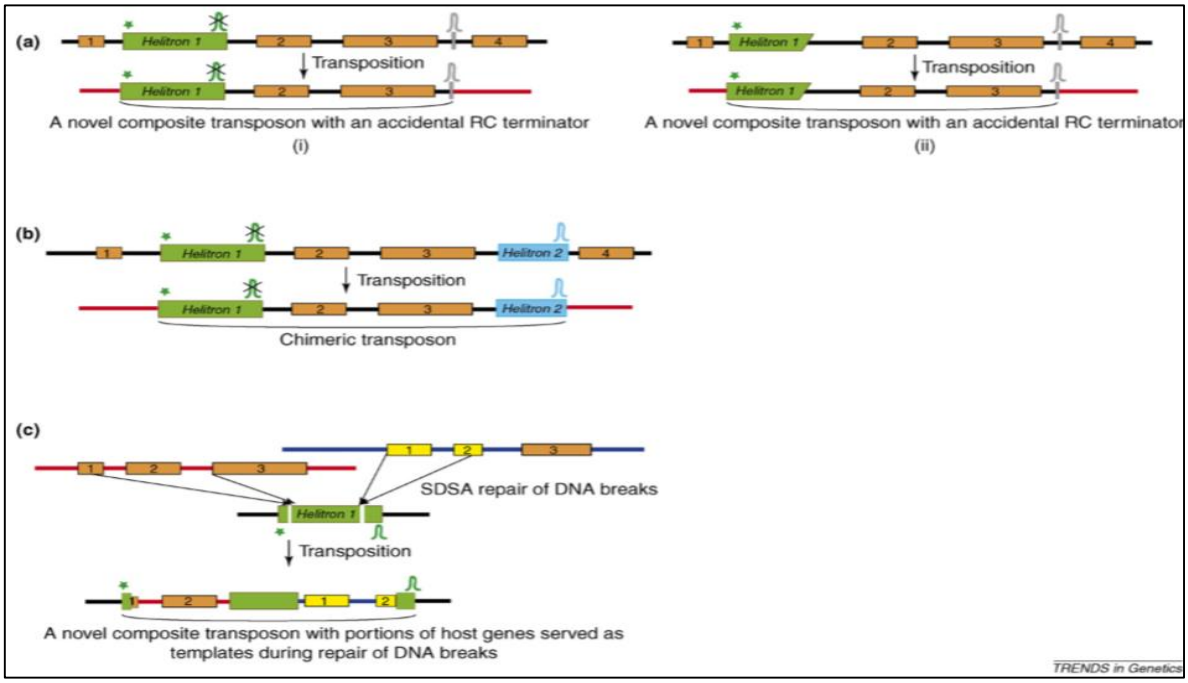

Fig 11: The RTM1, RTM2 and FDNA models of gene capture by Helitrons. (a) In the RTM1 model, the same portion of the host gene can be copied to a composite transposon. The RC terminator in the new transposon is formed de novo by a terminator-like signal present in the intron following exon 3. (b) In the RTM2 variant model, a portion of a host gene is copied to a novel chimeric transposon. (c) In the FDNA model, two genes residing in different chromosomes serve as templates during repair of DSBs that have occurred in the transposed helitron.

6. Horizontal gene transfer

Horizontal gene transfer (HGT; also known as lateral gene transfer) is the process by which an organism incorporates genetic material from another organism without being a direct descendant of that organism. HGT has also been frequently documented in phagocytic and parasitic unicellular eukaryotes.

HGT has also been frequently documented in phagocytic and parasitic unicellular eukaryotes (Keeling and Palmer 2008). However, until recently, HGT involving animals and plants appeared to be confined to events associated with endosymbiosis (e.g., transfer of mitochondrial or plastid genes tothenucleargenome) or parasitism (e.g., transferofgenes from the intracellular Wolbachia bacteria to their Drosophila hosts). It is thought that HGT is limited in animals because of a highly segregated and sheltered germline (Keeling and Palmer 2008). Interestingly, however, a recent study revealed that a species of rotifers (wheel animals) has acquired numerous genes from various other organisms (i.e., bacteria, fungi, and plants), potentially associated with the extreme environmental stress (repeated desiccation) to which this organism is subjected.

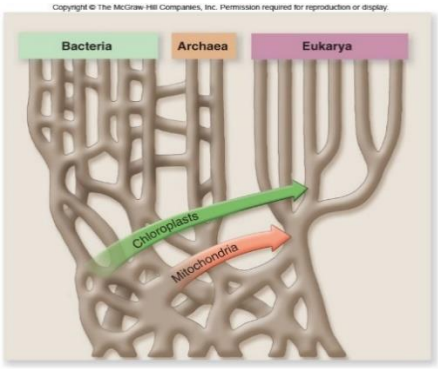

Fig 12: Horizontal Transfer of chloroplast and mitochondrial genes from prokaryote to eukaryote

However, although several acquired genes seem to have remained intact, the functional relevance of this curious case of HGT still needs to be established. Another intriguing example is a recently discovered new gene in rodents, which stems from a copy of SPIN, a family of transposable elements that was acquired horizontally. This domesticated DNA transposon, whose functionality is strongly supported by selection tests, apparently became transcribed (together with flanking exons) from a preexisting promoter located far upstream of its insertion site.

7. Gene fusion the origin of new chimeric genes

The process of gene fusion is defined as the fusion of two previously separate source genes into a single transcription unit the so called fusion or chimeric gene. Gene fusion is a fascinating mechanism of new gene origination that is almost bound to give rise to new functions given its combinatorial nature (assuming that the fusion gene is beneficial and selectively preserved).

7.1 DNA-mediated gene fusions

A common theme underlying several of the different gene fusion mechanisms is gene duplication, which provides the necessary raw material for the emergence of new fusion genes, allowing ancestral gene functions to be preserved. Thus, chimeric genes often arise from For example, the dispersion and shuffling of numerous segmental gene copies in hominoids through various recombination and translocation events has led to the formation of many mosaic gene structures, some of which have become transcribed. Among these transcribed chimeras, there are several genes with known functions (e.g., USP6, also known as Tre2, oncogene with testis expression). Juxtaposed pieces of duplicate gene copies through fission and fusion processes.

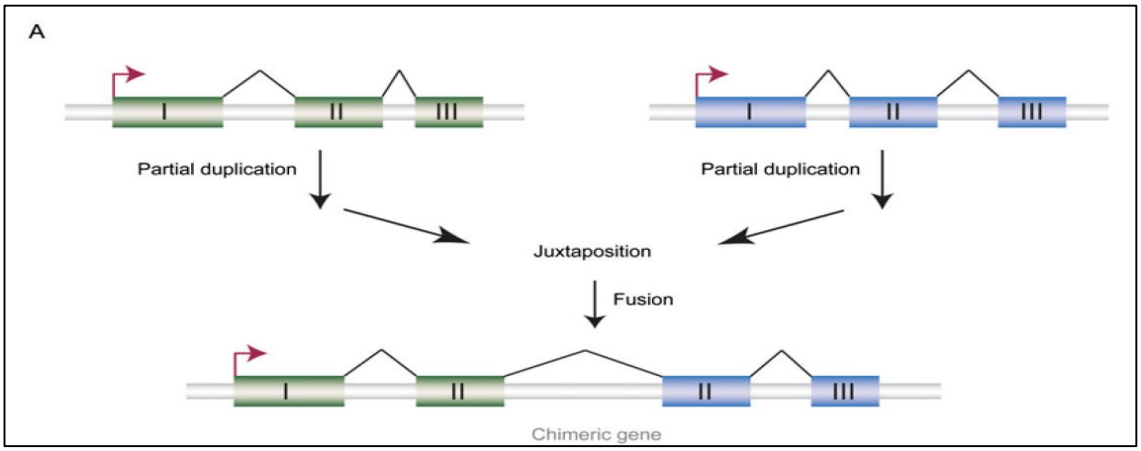

Fig 13: Origin of new chimeric gene or transcript structures. DNA-based (genomic) gene fusion. Partial duplication (and hence fission) of ancestral source genes precedes juxtaposition of partial duplicates and subsequent fusion (presumably mediated by the evolution of novel splicing signals and/ or transcription termination/poly adenylation sites).

7.2 Transcription-mediated gene fusions

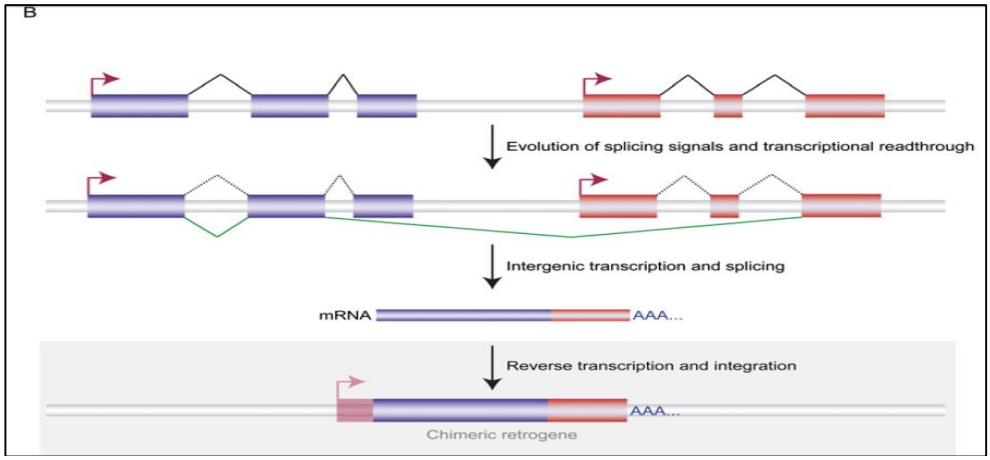

Fig 14: Transcription-mediated gene fusion. Novel transcript structures may arise from intergenic splicing after evolution of novel splicing signals and transcriptional read-through from the upstream gene. New chimeric mRNAs may sometimes be reversed transcribed to yield new chimeric retrogenes.

Transcription-mediated gene fusion an alternative gene fusion mechanism that combines exons from independent consecutive genes in the genome at the transcription level by intergenic splicing. Given that this mechanism draws from exons of preexisting genes, it does not represent a true process of new gene formation. It gives rise to new transcription units with potentially novel functions that may sometimes be fixed as new genes in the genome through secondary events.

8. Gene origination from scratch

New genes arise from previously nonfunctional genomic sequence, unrelated to any preexisting genetic material.

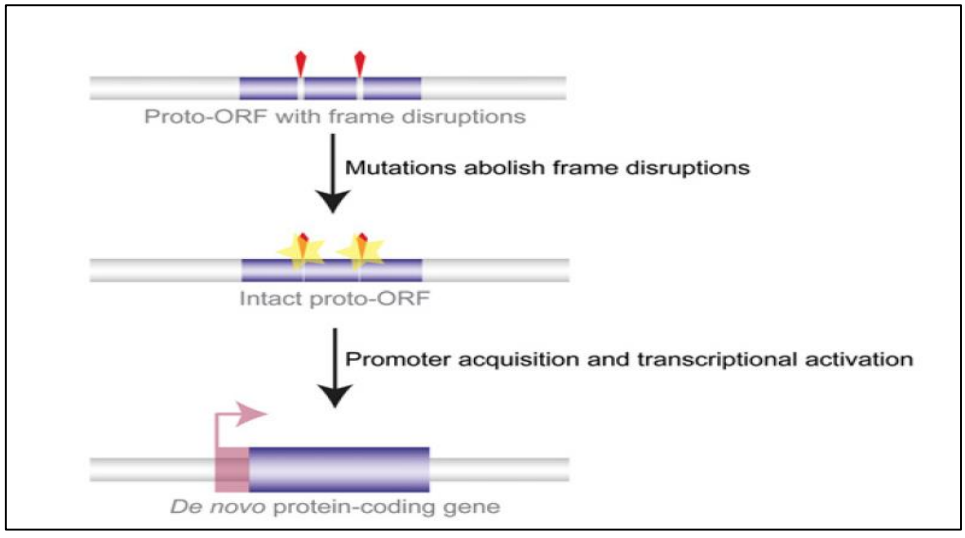

Fig 15: Origin of protein-coding genes from scratch.

New coding regions may emerge de novo from noncoding genomic sequences. First, proto-open reading frames (proto-ORFs; thin blue bars) acquire mutations (point substitutions, insertions/deletions; yellow stars) that remove, bit by bit, frame-disrupting nucleotides (red wedges). Transcriptional activation of ORFs (through acquisition of promoters located in the 5' flanking region) encoding proteins with potentially useful functions may allow for the evolution of novel protein-coding genes. (Large blue box) Functional exon, (pink right-angled arrow) transcription start site, (transparent pink box) untranslated 5' sequence. Note that the transcriptional activation step may, alternatively, also precede the formation of complete functionally relevant ORFs.

9. De novo emergence of protein-coding genes

Apparently arose from previously noncoding (and non-repetitive) DNA sequences Details regarding the emergence of the original coding sequence remain unclear. Lack of any corresponding orthologous sequences suggest a de novo origin for this gene family. 14 de novo-originated genes have been identified in Drosophila.

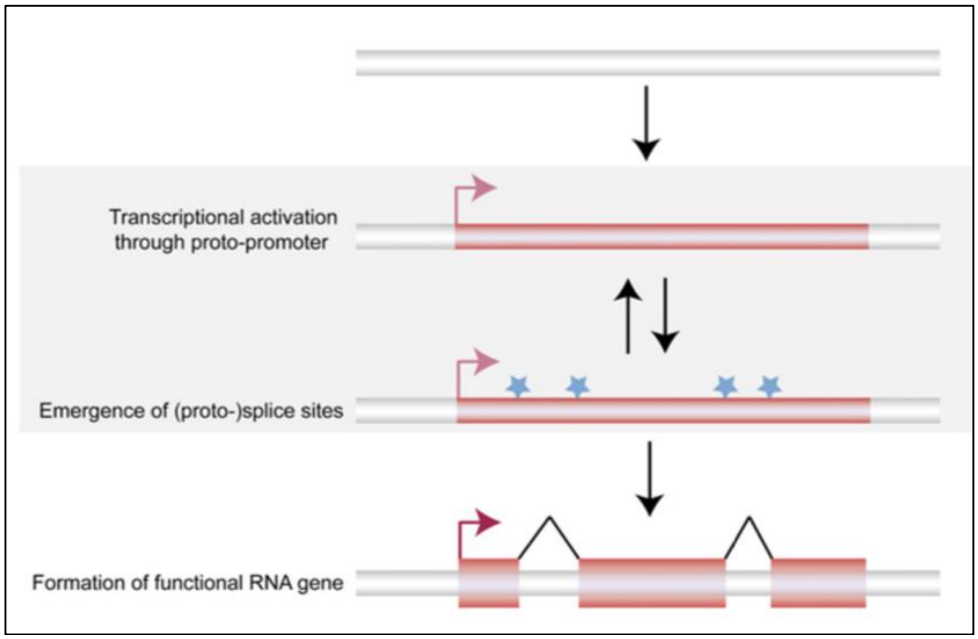

Fig 16: Mechanism of de novo emergence of protein coding genes

In this scenario, previously nonfunctional genomic sequence becomes transcribed (thin red box) through the acquisition/activation of a proto-promoter sequence (right-angled arrows). The transcriptional activation may be followed or preceded by the evolution of (proto-) splice sites (light blue stars). Together, these events allow for the formation of potentially functional and selectively beneficial multi exonic noncoding RNA genes. (Large red boxes) Exons, (thin black lines) splicing, (red right-angled arrows) TSSs.

10. Combined mechanisms- The Sdic Gene Cluster an example (Nurminsky, et, al. 1998)

The Sdic gene is a recently evolved chimeric gene in *D. melanogaster*, discovered and described by Nurminsky and colleagues in 1998. This gene possesses several unique features that provide an exceptional opportunity for the study of new gene functions, the fateofgene duplications, and the evolution of male reproductive traits. Sequence analysis of the Sdic gene revealed that Sdic is a chimera of two genes that exist intact in the genome. Sdic is composed of parts of Ann X, which encodes an annexin protein, and Cdic (also referred to in the literature and Fly Base as sw), which encodes an intermediate polypeptide chain for the cytoplasmic dyneins. The structure of Sdic along with the fact that Ann X and Cdic exist intact in the genome indicates that Sdic originated as a duplication and fusion of Ann X and Cdic, followed by small deletions and rearrangements. Its formation involved the creation of novel promoter elements (which provided testis-specific expres- sion) from the fusion of portions of an Ann X exon and a Cdic intron. Its coding region, however, derived solely from Cdic. The comparison of the coding region of Sdic with Cdic shows that Sdic lacks the 3' region of Cdic (which corresponds to 100 amino acids residues at the C-terminal part of the Cdic protein) and at its 5end underwent extensive refashioning by the occurrence of multiple mutations, deletions (including frameshift deletions), and insertions, culminating in a new 5' exon that encodes a totally novel N-terminus for the protein (Figure 17).

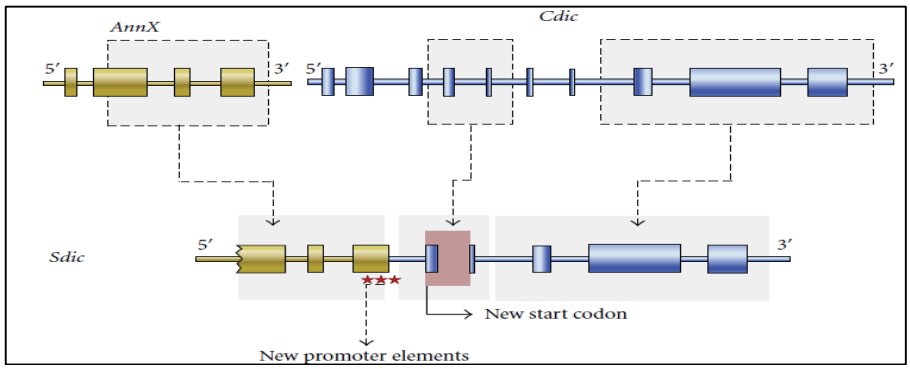

Fig 17: Formation of the Sdic gene from parts of the genes Ann X and Cdic. Introns are represented as thin cylinders and exons as thick cylinders. The stars represent Sdic promoter elements.

There are several copies of Sdic located in tandem at the base of the X chromosome, in region 19 of the larval salivary gland polytene chromosomes, forming a gene cluster. This repeated region is flanked by the parental genes, on the 5' side by Cdic and on the 3' side by Ann X. According to the available genomic sequence of D. melanogaster, the Sdic gene is repeated four times in tandem between the genes Cdic and Ann X genes. Within this cluster there are also four dead-on-arrival retro transposable elements of the RT1C family, one RT1C copy located upstream of each Sdic gene copy (Figure 18).

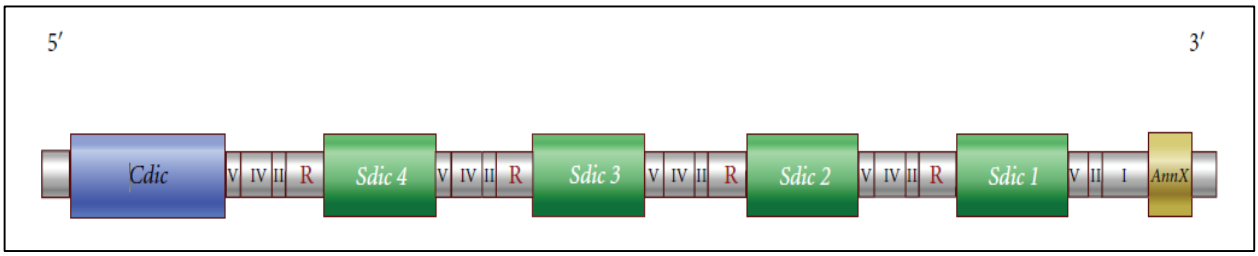

Fig 18: The Sdic gene cluster. The cluster is composed by four Sdic genes, with an RT1C retro transposable element upstream of each Sdic gene. The cluster is located between the parental genes Cdic and Ann X. Cdic is represented in blue, Sdic genes in green and Ann X in yellow; intergenic regions are grey; R represents RT1C elements. These genes are located in the minus strand, so the order of genes in this figure is the opposite order of these genes in Flybase.

11. Origin of noncoding RNA genes

Recent transcriptome studies have unveiled an unexpectedly rich repertoire of noncoding RNA species, which, in mammals, are derived from hundreds of small and thousands of lnc RNA loci. It is known that at least mi RNA and pi RNA genes proliferated and diversified via gene duplication (for lnc RNAs, there is so far little evidence).

1. Noncoding RNA genes transformed from Protein-coding genes
2. Noncoding RNAs from transposable elements

11.1 Noncoding RNA genes transformed from Protein-coding genes

Formation of new lnc RNA genes may directly draw from previous gene structure information and regulatory capacity. Given the constant generation of new protein-coding gene copies through gene du- plication and the frequent (often associated) gene death processes during evolution, example the origin of Xist and spx genes in, *Drosophila.*

In this process, the original (functionally redundant) protein-coding gene loses its function and becomes a pseudogene. After or during loss of protein function and coding exon decay, a new functional noncoding RNA gene may arise, a process that may draw from regulatory elements and other sequences (splicing signals, exon sequences, polyadenylation sequences, etc.) from the ancestral protein-coding gene. (Blue boxes) Protein-coding exons, (red boxes) RNA exons, (transparent boxes) pseudogenized exons, (thin black lines) splicing, (dotted lines) lost ancestral splicing capacity, (red right-angled arrows) transcription start site.

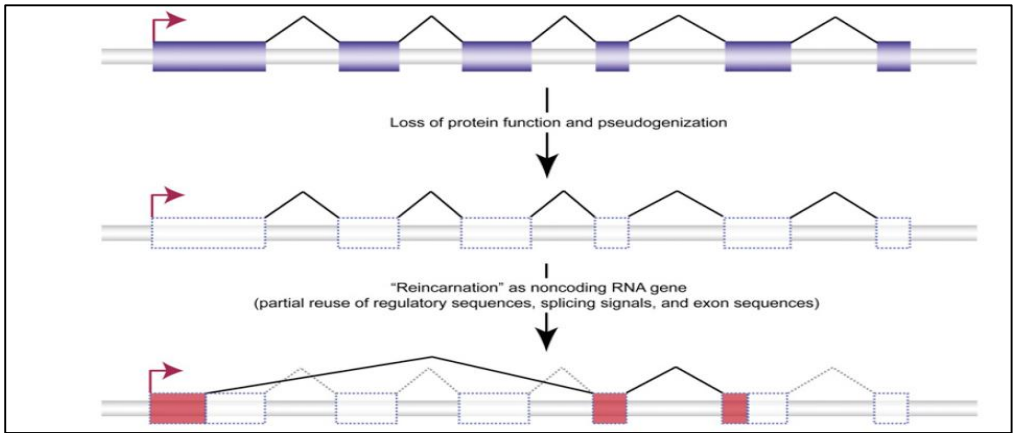

Fig 19: Origin of noncoding RNA gene from ancestral protein coding gene.

11.1.1 Drosophila Xist gene (Duret *et al*. 2006) [2]
The Xist gene, well known for its crucial role in X chromosome dosage compensation in eutherian mammals (where it triggers transcriptional inactivation of one female X chromosome), emanated from the remnants of a former protein-coding gene. This metamorphosis involved the loss of protein-coding capacity of the precursor gene's exons and subsequent reuse of several of these exons and original promoter elements in the newly minted Xist RNA gene. But the origin of lncRNA genes from protein coding antecedents is not confined to mammals.

11.2 Noncoding RNAs from transposable elements
In addition to various other protein-coding genes that arose on the basis of transpos- able element sequences in diverse taxa (i.e., vertebrates, fruit flies, and plants), several long and small RNA genes were shown to represent ''reincarnated'' retro transposons. This process is exemplified by the origin of the brain cytoplasmic lnc RNA genes (BC1 and BC200). Although these genes evolved independently from retro transposons in rodents and anthropoid primates (Brosius 1999) [1], they adapted to similar roles in translational regulation in the brain. While cases of lncRNAs that were derived from transposon ancestors are so far scarce, newsmallRNA genes seemto rather frequently have emerged from transposable elements. Germline expressed piRNAs and endo siRNAs should also be mentioned in this context, because they are frequently derived from the various lineage-specific transposable elements (Malone and Hannon 2009) [7].

12. The testis: A catalyst for the birth and evolution of new genes in animals
Studies of new genes in animals have ascribed one specific organ an intriguing and potentially central role in the process of gene birth and evolution. Probably not fortuitously, already the first detailed investigations of recent gene origination in mammals Pgk2 and Drosophila (jingwei) revealed the newly formed genes to be specifically expressed in one tissue: the testis. Global studies of retroduplication later showed an overall propensity of young retro genes to be expressed in this organ in these species. These observations, suggested that the testis may represent a crucible for new gene evolution, allowing novel genes to form and evolve, and potentially adopt functions in other (somatic) tissues with time.

12.1 The ''out of the testis'' hypothesis for the emergence of new genes
Several factors likely contributed to the ''out of the testis'' emergence of new genes. It is well established that, at the genomic and molecular level, the testis constitutes the most rapidly evolving organ, owing to the intense selective pressures to which it is subjected and that are associated with sperm competition, sexual conflict, reproductive isolation, germline pathogens, and mutations causing segregation distortion in the male germline (Nielsen *et al*. 2005) [8]. Thus, the testis may represent an evolutionarily ''greedy'' tissue, highly receptive for the accommodation of evolutionary genomic innovations such as new genes. This hypothesis suggests that the transcription of new gene copies/ structures (green boxes) is facilitated in certain testis germ cells meiotic spermatocytes and post meiotic round spermatids (which are found in the seminiferous tubules, where spermatogenesis takes place) because of the potentially overall permissive chromatin state and overexpression of key components of the transcriptional machinery in these cells.

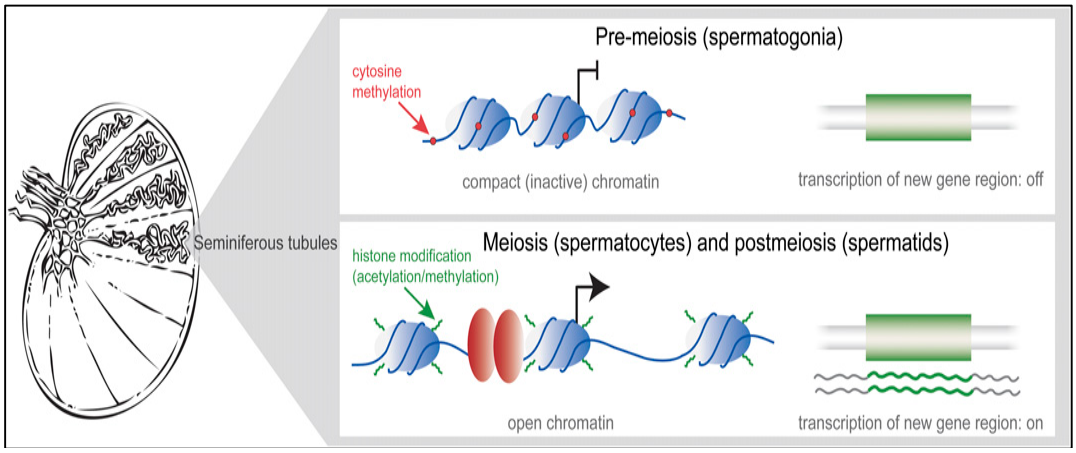

Fig 20: The ''out of the testis'' hypothesis for the emergence of new genes

The transcriptionally active chromatin state in spermatocytes and spermatids is thought to be a result of a potentially widespread demethylation of CpG dinucleotide-enriched promoter sequences and modifications (acetylation and methylation) of histones (blue ovals), which facilitate access of the transcriptional machinery (red ovals). Once transcribed, new functional genes (transcripts shown as green wavy lines) with beneficial products may be selectively preserved and evolve more efficient promoters (a process that might be facilitated by the fact that spermatocyte/spermatid-specific expression requires only relatively simple promoters). Eventually, such new genes may also evolve more diverse expression patterns and thus also obtain functions in other (somatic) tissues.

13. Evolution of new genes
Darwin's original contributions were the mechanism of natural selection and copious amounts of evidence for evolutionary change from many sources. He also provided thoughtful explanations of the consequences of evolution for our understanding of the history of life and modern biological diversity.
The primary mechanism of change over time is natural selection, this mechanism causes changes in the properties (traits) of organisms within lineages from generation to generation. Darwin's process of natural selection has four components.

1. Variation. Organisms (within populations) exhibit individual variation in appearance and behavior. These

variations may involve body size, hair color, facial markings, voice properties, or number of offspring. On the other hand, some traits show little to no variation among individuals for example, number of eyes in vertebrates.
2. Inheritance. Some traits are consistently passed on from parent to offspring. Such traits are heritable, whereas other traits are strongly influenced by environmental conditions and show weak heritability.
3. High rate of population growth. Most populations have more offspring each year than local resources can support leading to a struggle for resources. Each generation experiences substantial mortality.
4. Differential survival and reproduction. Individuals possessing traits well suited for the struggle for local resources will contribute more offspring to the next generation.

14. Neutral Theory of Molecular Evolution

Neutral theory of evolution has become central to the study of evolution at the molecular level, in part because it provides a way to make strong predictions that can be tested against actual data. The neutral theory holds that most variation at the molecular level does not affect fitness and, therefore, the evolutionary fate of genetic variation is best explained by stochastic processes. This theory also presents a framework for ongoing exploration of two areas of research: biased gene conversion, and the impact of effective population size on the effective neutrality of genetic variants.

The evolution of living organisms is the consequence of two processes. First, evolution depends on the genetic variability generated by mutations, which continuously arise within populations. Second, it also relies on changes in the frequency of alleles within populations over time. The fate of those mutations that affect the fitness of their carrier is partly determined by natural selection. On one hand, new alleles that confer a higher fitness tend to increase in frequency over time until they reach fixation, thus replacing the ancestral allele in the population. This evolutionary process is called positive or directional selection. Conversely, new mutations that decrease the carrier's fitness tend to disappear from populations through a process known as negative or purifying selection. Finally, it may happen that a mutation is advantageous only in heterozygotes but not in homozygotes. Such alleles tend to be maintained at an intermediate frequency in populations by way of the process known as balancing selection.

15. Conclusions

Genes are originated from preexisting old genes. Duplication mediated through mobile elements play a major role in origin of new genes. Advantageous genes fixed through natural selection. Null alleles and deleterious alleles fixed through genetic drift. Origin and evolution of new genes gives concept of sharing a gene between the species. Study of new genes in related species helps for selection of parents for inter specific hybridization.

16. References

1. Brosius J. RNAs from all categories generate retro sequences that may be exapted as novel genes or regulatory elements. Gene. 1999; 238:115-134.
2. Duret L, Chureau C, Samain S, Weissenbach J, Avner P. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Sci. 2006; 312:1653-1655.
3. Fablet M, Bueno M, Potrzebowski L, Kaessmann H, Evolutionary origin and functions of retrogene introns. Mol. Biol. Evol. 2009; 26:2147-2156.
4. Kaessmann H, Vinckenbosch N, Long M. RNA-based gene duplication: mechanistic and evolutionary insights. Nat. Rev. Genet. 2009; 10:19-31.
5. Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. 2008; 9:605-618.
6. Makalowski W, Mitchell GA, Labuda D. Alu sequences in the coding regions of mRNA: a source of protein variability. Trends Genet. 1994; 10:188-193.
7. Malone CD, Hannon GJ. Small RNAs as guardians of the genome. Cell. 2009; 136:656-668.
8. Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS. Biol**.** 2005; 3:170-171.
9. Nurminsky DI, Nurminskaya MV, De Aguiar D, Hartl D. L. Selective sweep of a newly evolved sperm specific gene in Drosophila. Nat., 1998; 396:572-575.