E-content

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

Unit: 3.5b

Evolution of multigene family

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

Introduction

A multi-gene family is a group of genes that have descended from a common ancestral gene and therefore have similar functions and similar DNA sequences.

Gene families are a direct consequence of the fact that essentially all new genes arise by gene duplication, either by wholesale duplication of entire genes or by duplication and shuffling of exons from different genes

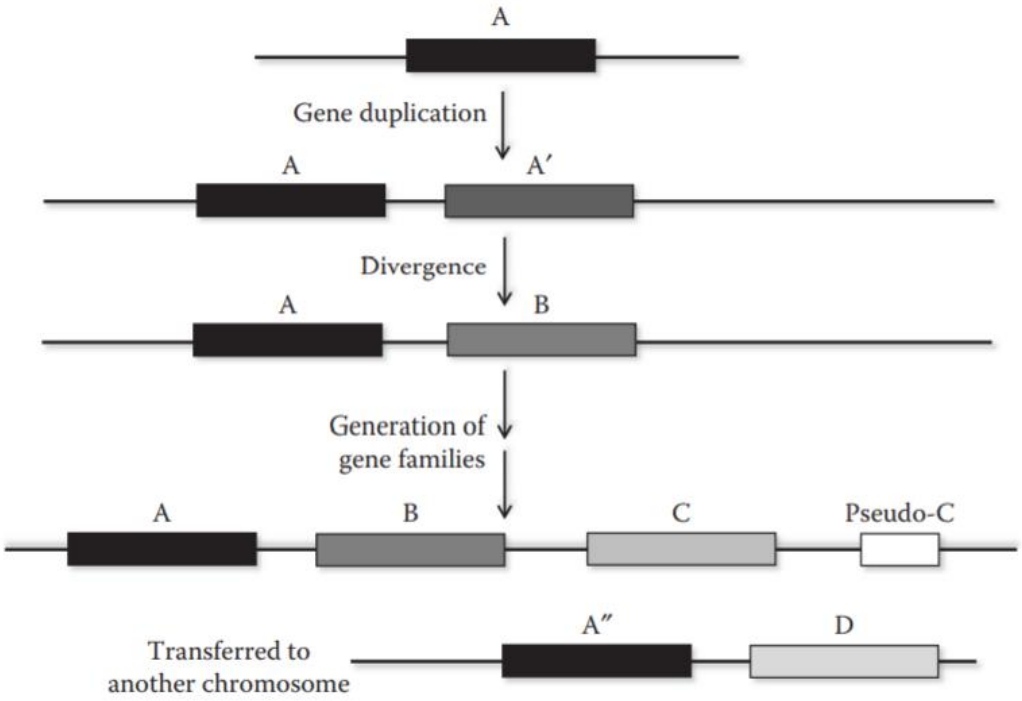
The clustering of functionally related members of a gene family reflects their common ancestry and subsequent duplication and divergence.

DNA duplications that involve one or more genes generate gene pairs. If both copies are maintained in subsequent generations then a multigene family will exist in the genome.

Chromosomal rearrangements disperse the multigene families throughout the genome.

Dispersed members of the multigene family can still be recognized by sequence comparison.

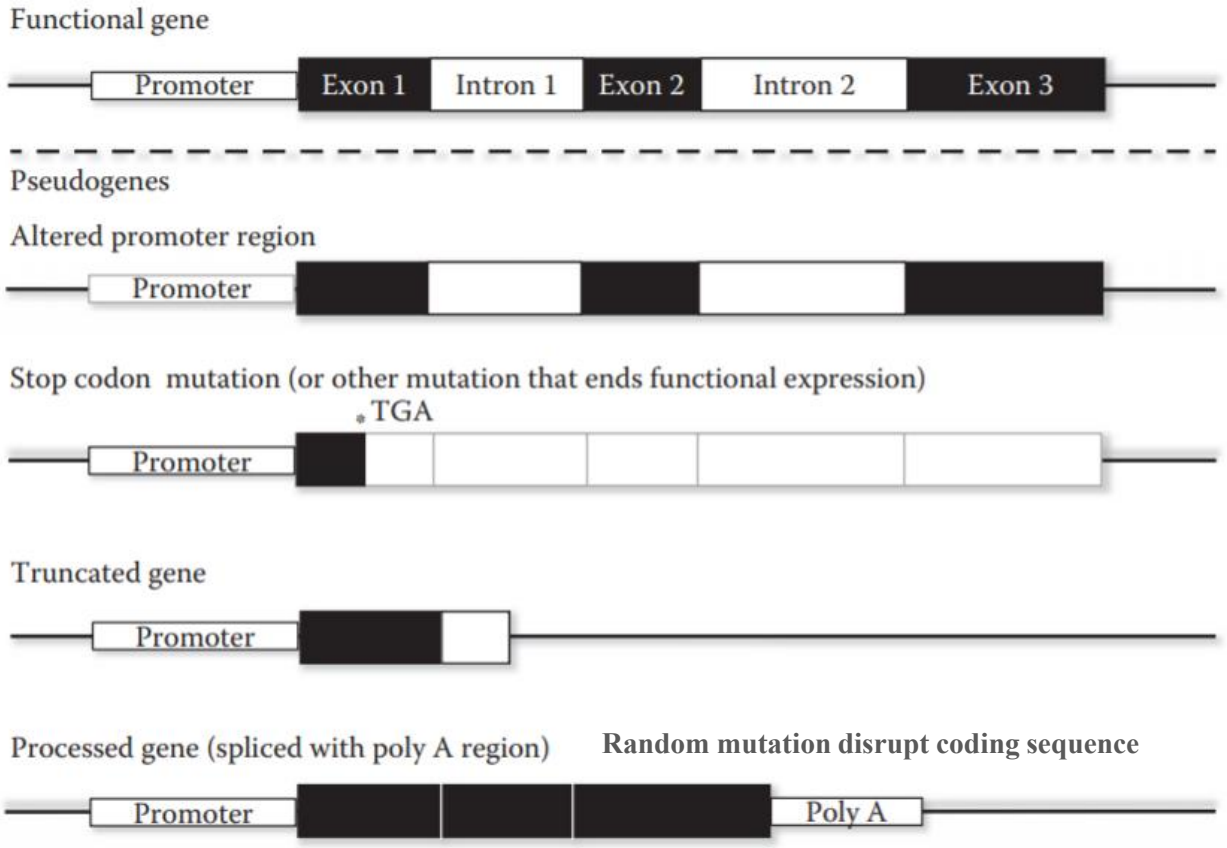
Generation of members of a multigene family



The first step in the formation of a multigene family is the duplication of a single-copy gene. Normally, this occurs via recombination. The copy has more freedom to mutate, because the original copy still functions normally. Therefore, the duplicate gene can diverge into a gene whose product has altered functions. In this process, some of the duplicated genes may change to a degree where they are no longer functional (i.e., either not expressed, or the protein produced is nonfunctional), thus becoming pseudogenes. Eventually, some of the gene family may be transferred to other chromosomes, by transposition, chromosomal translocation, or chromosome fragmentation

Types of Pseudogenes

Several different types of pseudogenes have been identified in various multigene families. The gene illustrated at the top is functional. It has a functional promoter, as well as several exons and introns. The simplest pseudogene is one in which the coding region is unaltered, but the promoter has mutated such that transcription is never initiated.



Another common mutation is one in which a codon has changed to a stop. The RNA is transcribed, but only part of the protein is produced during translation. A third mutation is where the gene itself has been truncated, usually by a recombination event. The fourth type appears to be an inserted DNA copy of the processed messenger RNA (mRNA) for the gene. The exons are present, but all of the introns are missing, as if they have been spliced out. Additionally, a number of adenines are present on the 5′ end of the gene, just as would be present as a poly A tail in the mature mRNA. This last type of pseudogene is generated by transposition or reverse transcription, rather than by recombination.

Examples of Multigene Families

Family	Number of Versions
Actins	61
Cell surface antigens	50–100
Egg shell proteins (insect)	50
α Globins	1–5
β Globins	5
Heat-shock proteins	>10
Histones (H2A, H2B, H3, and H4)	75
Immunoglobins (variable region)	381
Keratins	>20
Laccases (fungi)	3–5
Major histocompatability genes	30
Myosin (heavy chain)	5–10
Protein kinases	10 to >100
Ribosomal RNA genes	1 to dozens (1–23,000 copies)
T-cell receptors	33
Transfer RNA genes	Dozens (1–100 copies each)
Transcription factors	>100
Tubulins (α and β)	3–15
Visual pigment protein (human)	4
Vitellogenin (frog and chicken)	5
Zinc fingers	700

1. The Globin gene family

One of the best examples of a multigene family is the globin gene family, the genes that encode hemoglobins (carry oxygen in the blood) and myoglobin (carries oxygen in other tissues).

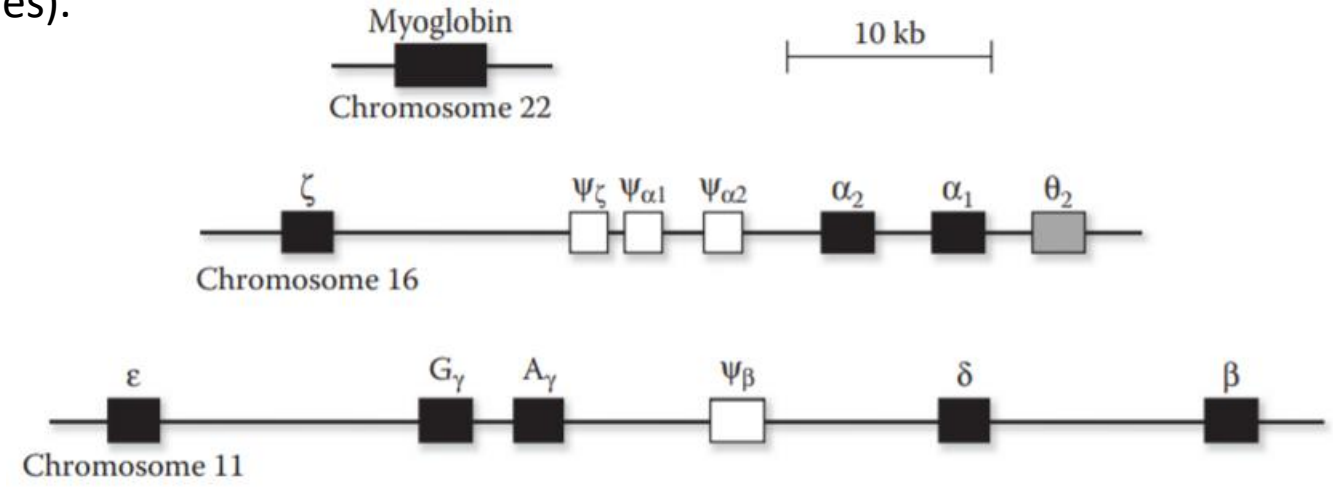


Figure: Arrangement of genes in the human globin gene family.

The myoglobin gene is on chromosome 22 and is isolated from all other members of the gene family.

The α-type globin genes are all located on chromosome 16.

This cluster consists of one gene (ζ) that is expressed only in the embryo, as well as two genes (α1 and α2) that are expressed in the fetus and adult.

Four pseudogenes are also present, including three (Ψζ, Ψα1, and Ψα1) that are not expressed and one (θ2) that is expressed, but the protein is nonfunctional.

Figure: Arrangement of human β-like globins

The β-like globins exist as a cluster of five genes and one pseudogene (Ψβ) located on chromosome 11.

One gene (ε) is expressed only in the embryo, whereas two genes (Gγ and Aγ) are primarily expressed in the fetus.

The two fetal genes continue to be expressed for a few months after birth, but are gradually replaced primarily by adult β globin and a small percentage of δ globin.

The genes are expressed developmentally in the order that they appear on the chromosomes. Black boxes indicate the expressed gene. Gray and white boxes indicate pseudogenes that are expressed (but nonfunctional) or not expressed, respectively.

Evolution of globin genes

When globin genes have been examined in other organisms, a phylogenetic reconstruction can be produced from the sequence information. All globins appear to have evolved from a heme-carrying protein that existed at least 1.1 billion years ago. At about that time, there was a gene duplication, and the two versions began to diverge such that the transfer of oxygen from one globin to the other version would be possible at specific concentrations (partial pressures) of oxygen.

This coincided with the appearance of the first multicellular eukaryotes. Thus, the cells on the outside of the organisms could pick up oxygen from the outside environment (water, in this case) and transfer it to the cells on the inside of the organism.

This process continued for approximately 500 million years until the next change, which coincided with the appearance of the first fish.

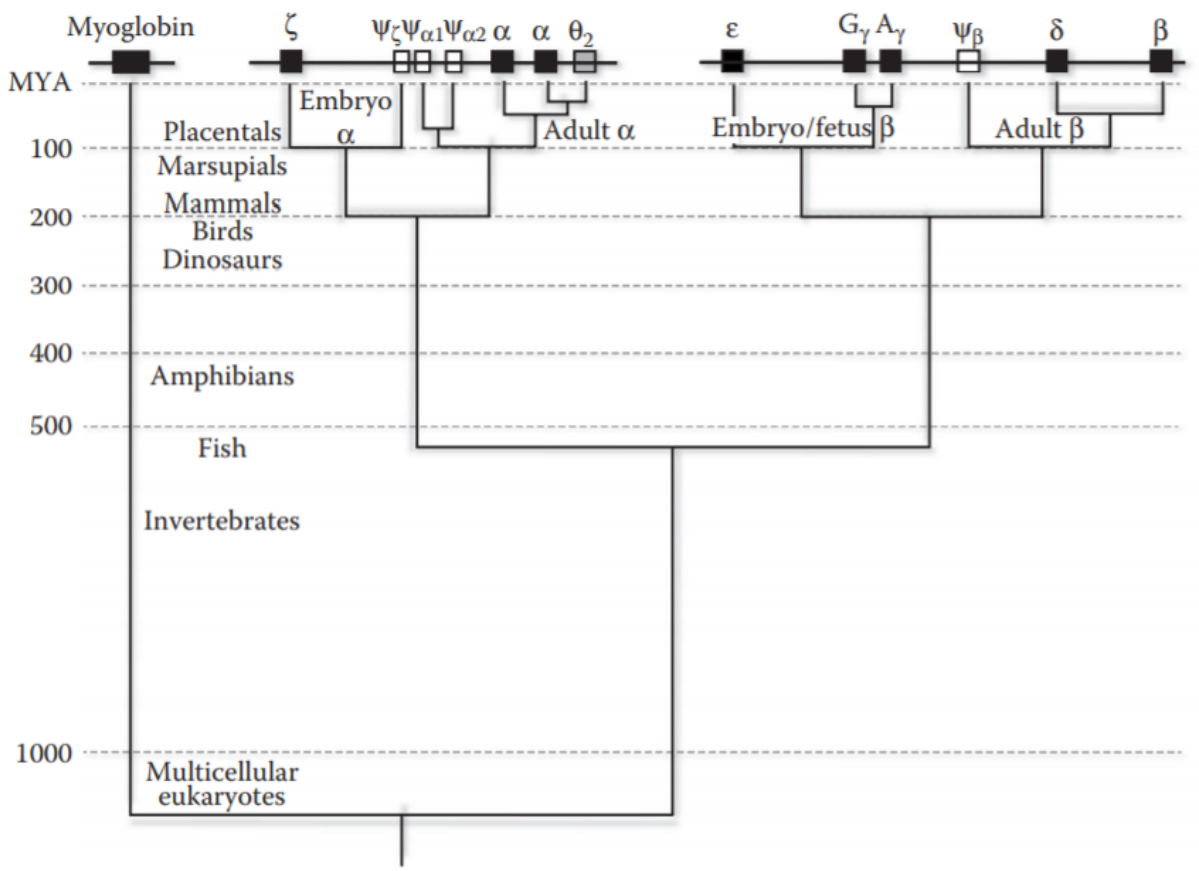


Figure: Evolution of globin genes

The first fish were small, but eventually larger fish evolved with higher demands for energy and oxygen. There was another round of duplications of the globin genes at this time, eventually producing several α- and β-type genes, and these formed tetramers to produce hemoglobin. Therefore, each hemoglobin unit could carry four oxygen molecules. This change occurred in the amphibians. Also, it was around this time the α- and β-type globin genes became unlinked, that is, they were on separate chromosomes.

The change in the globins and their ability to bind and release oxygen in different concentrations of oxygen also meant that the organisms could live outside of an aqueous environment. In other words, they could live on land, at least part of the time, extracting oxygen from the atmosphere instead of the water.

Over the next 100 million years, land animals diversified, and eventually, dinosaurs and birds appeared. As with amphibians, they laid eggs, but they laid them on land. This presented yet another environment where oxygen concentrations differed from the previous environments. By this time, there were at least two versions of each α- and β-type globin gene. This allowed the uptake of oxygen from the atmosphere into the egg, as well as from the atmosphere into the adult lungs and blood, while assuring adequate transfer of oxygen into the tissues of the fetuses and the adults.

Next, the mammals appeared as another part of the trend toward diversification in development and globin gene family evolution.

The earliest mammals laid eggs, but part of embryo development occurred within the mother. Therefore, oxygen had to be delivered from the mother's blood to the blood of the embryo. At this time, a concomitant diversification of both α- and β-type globins is also seen. The embryo versions of these genes appear at this time (200 million years).

Further duplication and diversification led to the ability of certain mammal mothers to carry their offspring longer during development. In marsupial mammals, the embryo develops into a fetus before exiting the mother and moving into the pouch for additional growth and development.

Finally, the placental mammals, such as humans, represent another stage in this expansion and diversification in globin genes. The most recent globin versions (Gγ and Aγ globins) were necessary to allow an increased transfer of oxygen into a large fetus, although it was still inside of its mother.

In this way the evolutionary events correlates with the evolution of globin genes.

2. Histone Gene Family

Another example of a multigene family is the histone gene family. Histones are basic (positively charged) proteins that are one of the major proteins that interact with DNA.

All of the histone genes in eukaryotes evolved from histone-like proteins that are found in members of the archaea, lower triangle).

Initially, there was a duplication and diversification event that led to a set of genes related to H3 and H4, and another related to H1 and both H2 types. Further duplication and diversification occurred, probably over a period of at least one billion years that resulted in the five types of histones (H1 through H4).

The triangles indicate a summary of the diversity of each of the histone gene types. The lengths of the triangles are indications of the diversity in each group, which shows that H1 is the most variable and H4 is the most conserved of the histone types.

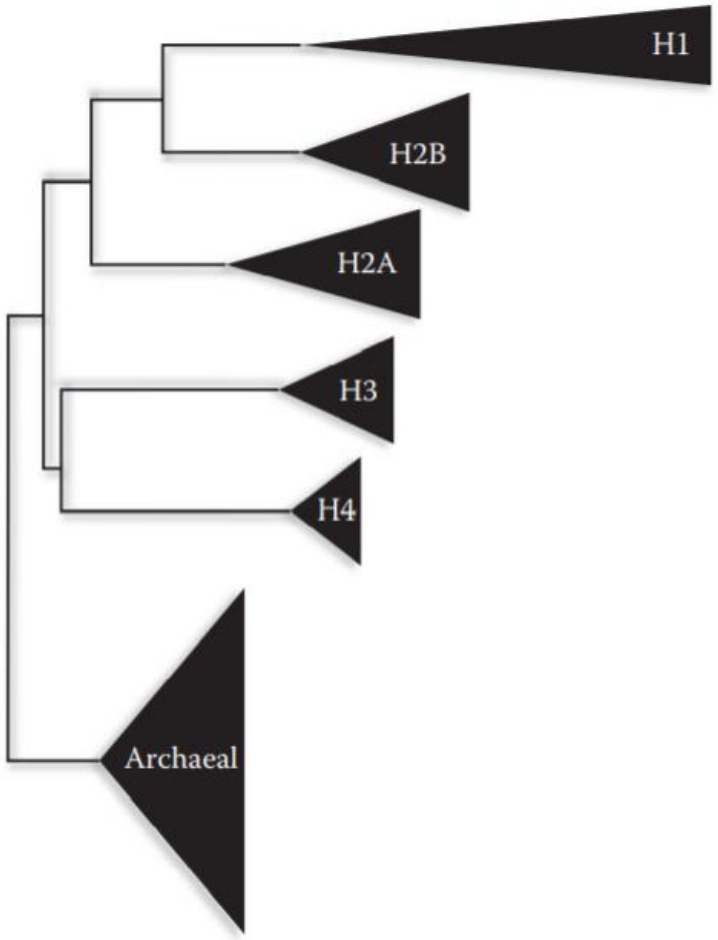


Figure: Evolution of histone genes.

References

Evolution by Douglas J. Futuyma

Integrated Molecular Evolution by Scott Orland Rogers

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