How Genes Evolve

Molecular Techniques Give Us a Closer Look at Evolution at the DNA Level

Ravi Parkash

The author is Professor of Biosciences at Maharshi Dayanand University, Rohtak. In addition to teaching post-graduate courses in genetics and evolutionary biology, his research interests mainly focus on ecological and evolutionary genetics of Indian *Drosophila* species.

The increasing power of molecular techniques of gene cloning, DNA sequencing and use of restriction enzymes has revolutionised our ability to understand mechanisms of evolution at the level of genes. Exciting new findings on pseudogenes and gene elongation have led to an increased understanding of how our genes themselves change through evolutionary time, often in bizzarre ways.

Molecular evolutionary studies have experienced two periods of exciting developments. The first period started in the 1960s with the introduction of techniques of amino acid sequencing and gel electrophoresis to evolutionary studies, which led to construction of phylogenetic trees from molecular data. More recently, use of the ever increasing repertoire of molecular biology techniques has helped in probing the detailed structure and organisation of genes which allows us to study their evolution with much greater resolution than before. The process of gene evolution includes changes in gene structure, function and regulation over evolutionary time. The course of gene evolution can be partly reconstructed by examining the structure and activity of specific proteins/enzymes in a wide spectrum of organisms belonging to different taxonomic groups.

Gene evolution involves changes in gene structure, function and regulation over evolutionary time.

The Evolution of Globin Genes

Haemoglobin is a well known respiratory pigment which transports oxygen to various tissues of the body. Tetrameric haemoglobin (Hb) molecules consist of two pairs of polypeptides e.g. human foetal Hb consists of two alpha chains and two gamma chains, while adult Hb consists of two

Box 1. Orthologous and Paralogous Genes

The analysis of extent of genetic changes from protein sequence data are based on the assumption that the genes coding for the proteins are homologous i.e. descended from a common ancestor. There are two types of homologous relationships among genes i.e. orthologous and paralogous. Orthologous genes are descendants of a single ancestral gene present in the ancestral species. Evolutionary changes in orthologous genes, therefore, reflect the evolution of these genes in different species in which they are found e.g. the evolution of the gene coding for the protein cytochrome-C which is part of the respiratory apparatus. On the contrary, paralogous genes are descendants of a duplicated gene. Paralogous genes therefore evolve within the same species as well as in different species e.g. genes coding for alpha, beta, gamma and delta haemoglobin chains in man are paralogous genes. The evolution of paralogous genes depicts differences that have accumulated from the time when that gene got duplicated. Homologies between paralogous genes serve to establish gene phylogenies i.e. the evolutionary history of duplicated genes within a given lineage. The timings of gene duplication events can be inferred based on which of the species have the duplicate gene product and which do not.

alpha and two beta polypeptides. We now know that two distinct families of human globin genes have evolved from a single ancestral gene through repeated duplications. Such genes are termed paralogous and are very useful in evolutionary studies (*Box 1*). From the degree of amino acid sequence similarity among the various globin genes, and the evolutionary history of the creatures in which various globin genes are found, the timings of the duplication events have been reconstructed (*Figure 1*). The evolution of the genes

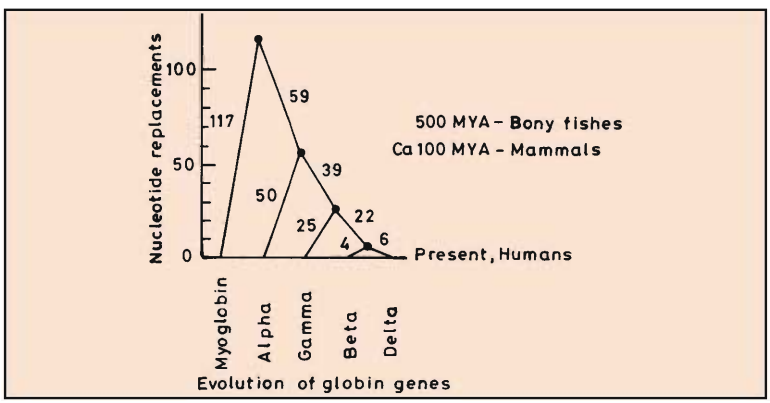


Figure 1. Evolutionary history of haemoglobin genes. The dots indicate where the ancestral genes were duplicated giving rise to a new gene line. The minimum number of nucleotide substitutions required to account for the amino acid differences between the proteins is shown for each branch. The time estimates in millions of years ago (MYA) are based on paleontological and morphological studies of vertebrates.

In mammals, the composition of haemoglobin changes during the course of development to suit the different biochemical environments in which the molecule has to work.

coding for the various haemoglobin chains can be summed up as follows. The amino acid sequence differences between myoglobin and haemoglobin are large enough to suggest that the genes coding for these two proteins diverged long ago in the history of vertebrates. It has been suggested that the divergence between genes coding for the ancestral haemoglobin (oxygen carrying) molecule and an ancestral myoglobin (oxygen storing) molecule occurred about 900 MYA (million years ago). This hypothesis is supported by the fact that the haemoglobin of the primitive chordates (Agnatha or jawless pre-vertebrates) consists of only one polypeptide chain. The great number of amino acid differences between the alpha and beta haemoglobin chains can be accounted for by hypothesising an early divergence (5000 MYA) of these chains from the ancestral haemoglobin. The fewer differences between beta and gamma polypeptide chains, as well as the fact that gamma chains are found only in mammals, suggest that gamma gene diverged from beta more recently (100 MYA), during the evolution of mammals. In more recent times (40 MYA), the beta lineage underwent gene duplication again during the evolution of primates, yielding the delta polypeptide chain (*Figure 1*).

The evolution of multiple globin genes through gene duplication is likely to have had great adaptive significance. The early haemoglobin gene duplication leading to the evolution of alpha and beta type haemoglobins allowed the production of a tetramer that was more effective in oxygen transport and release as compared to the ancestral monomeric form of haemoglobin. Further, gene duplications in mammals allowed new types of haemoglobin tetramer (two alpha and two gamma chains) that specialised in supplying the intra-uterine and relatively anoxic embryo and foetus with oxygen. Thus, in mammals, the composition of the haemoglobin tetramers changes during the course of development to suit the different biochemical environments in which the molecule has to work (*Figure 2*).

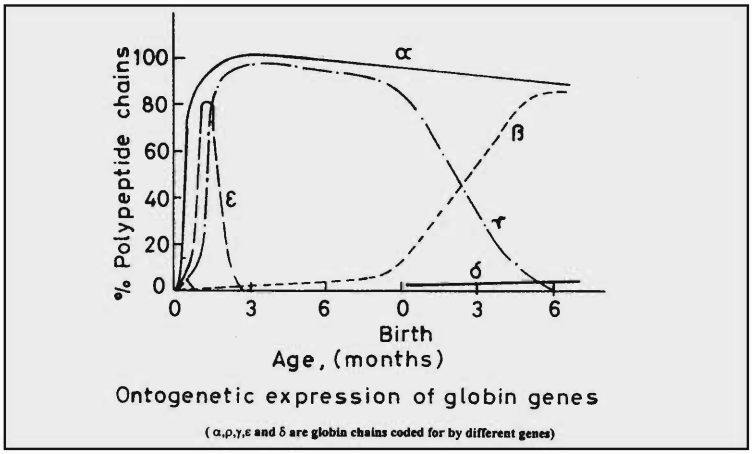


Fig. 2 The pattern of developmental expression (ontogeny) of haemoglobin genes in humans. Ontogenetic regulation of different haemoglobin genes has been depicted on the basis of temporal expression of various haemoglobin polypeptide chains i.e. alpha gene expresses throughout pre-natal and post-natal development while gamma gene switches off and beta gene switches on post-natally. In adults, haemoglobin is mainly (97.5%) composed of two alpha and two beta chains while there is limited expression (2.5%) of two alpha and two delta chains.

How Do We Understand Gene Evolution at the Molecular Level?

More recently, gene cloning and nucleotide sequencing techniques have been used to analyse the evolution of globin genes at an even finer level of resolution. Comparison of nucleotide sequences in the coding and non-coding regions of the globin genes indicates that these regions have continued to diverge by base substitution as well as by duplication and deletion of short regions. It has been found that the rate of nucleotide substitutions leading to synonymous codons is considerably greater than the rate leading to amino acid substitutions. Moreover, the rate of nucleotide substitution in non-coding regions of the genes is much higher than that of coding sequences. The rates of amino acid substitution have been found to vary in different haemoglobin genes.

One unexpected outcome of the sequencing of globin genes has been the detection of pseudogenes in each gene family. Pseudogenes are stretches of DNA having some sequence homology with protein coding regions but they are transcriptionally inactive. These extra gene copies have resulted due to gene duplication and accumulated mutations which made them inactive. Thus, pseudogenes are free from the constraints of natural selection because they have no phenotypic effect. It has

Pseudogenes are stretches of DNA having some sequence homology with protein coding regions but which are transcriptionally inactive.

In addition to the duplication of complete genes, partial gene duplication is another mechanism by which gene evolution takes place.

been suggested that the genomes of higher eukaryotes may be mosaics of dispersed pseudogenes which serve as a reservior of DNA sequences and may evolve to confer new functions. Pseudogenes have now been found in many other gene families adding support to the view that gene duplications leading to formation and dispersal of pseudogenes might be a critical event during gene evolution.

How Have Lactate Dehydrogenase (LDH) Genes Evolved?

The enzyme lactate dehydrogenase regulates interconversion of pyruvate to lactate in different body tissues. Vertebrate LDH is a tetramer made up of either four identical polypeptides (A4 and B4) or combinations of A and B polypeptides such as A3 B1, A2 B2 and A1 B3; and all such five molecular variants of LDH are called *isoenzymes* or isozymes. Only one LDH gene codes for A4 isozyme in the pre-vertebrates while all vertebrate species express three paralogous genes (A, B and C). Thus, early divergence of LDH A and B genes is likely to have occurred approximately 500 MYA. The third LDH gene of vertebrates arose due to gene duplication of LDH B. During LDH gene evolution, the original and identically duplicated genes became independent in both function as well as regulation e.g. LDH-A gene codes for the major isozyme of white skeletal muscles while LDH-B gene codes for heart muscle isozyme; this tissue specific expression has been adaptively conserved in all vertebrates. On the contrary, the third gene LDH-C is expressed only in the retina of most diploid bony fishes and in the liver of a few other species, while in birds and mammals LDH-C is expressed only in primary spermatocytes.

How Genes Evolve Through Elongation

In addition to the duplication of complete genes, partial gene duplication is another mechanism by which gene evolution takes place. Many proteins of present-day organisms show internal repeats of amino acids which often correspond to the

functional or structural domain of the proteins e.g. tropomysine chain, plasminogen, ferredoxin, serum albumin, protease inhibitor, immunoglobulin E chain C-region etc. It has been suggested that the genes coding for these proteins were formed by internal duplication of some parts of a pre-existing gene and the function of these genes was improved by increasing the number of active sites or, perhaps, by enhancing the stability of the protein produced.

The evolutionary history of some genes has now been constructed from the gene's internal structure. The analysis of type-1 collagen gene of chicken has revealed the occurrence of tandem multiplication in its structure. This gene contains more than 50 exons (see the article ' Immune System and Bodily Defence', *Resonance*, September, 1997) and is about 38 Kb (= 38000 nucleotides long). The exons consist of a repeated sequence, 9 nucleotides along, that codes for the triplet Gly-X-Y where X and Y are often prolines. Data on the sequence homology of 90 exons of collagen type-1 gene have revealed that 12 exons contain 54 bp (base-pairs); four 99 bp and three 108 bp. It is possible that all these exons were derived from an ancestral exon of 54 bp by multiple duplication and recombination. The origin and evolution of some immunoglobulin (antibody) genes has also been found to involve a tandem triplication of a small building block. Thus, increase in gene size or gene elongation often constitutes one of the most important steps in the evolution of complex genes from primitive ones.

Increase in gene size or gene elongation often constitutes one of the most important steps in evolution of complex genes from primitive ones.

Summary

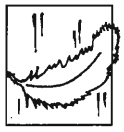
The use of molecular techniques has revealed at least two major classes of duplicate genes. Some genes, detected through analysis of their products, occur in a few copies and the effects of natural selection on their evolution are much more evident (e.g. LDH and Hb genes). Other genes are found to contain highly repeated nucleotide sequences in the genome (e.g. gene families coding for t-RNAs, r-RNAs and satellite DNA sequences) which are

Suggested Reading

- S Ohno. *Evolution of Gene Duplication.* Springer-Verlag. 1970
- F J Ayala. *Molecular Evolution.* Sinauer Associates, Sunderland, Massachussets. 1976.
- M Nei and R K Koehn. *Evolution of Genes and Proteins.* Sinauer. 1983.
- W H Li and D Graur. *Fundamentals of Molecular Evolution,* Sinauer. 1991.

probably maintained through concerted evolution or molecular drive. Our understanding of gene evolution through gene duplication has been revolutionised by the application of powerful new molecular techniques. The data on the extent of occurrence of gene duplication is, however, limited and a final picture is still awaited. Nevertheless, it is now firmly established that tandem gene duplications have occurred quite frequently and seem to constitute one of the major evolutionary mechanisms for the origin and evolution of new types of genes.

Address for Correspondence
Ravi Parkash
U H - 9, Medical Campus
Rohtak 124 001
Fax No. (01262) 43791
Res. Tel. No. (01262) 41577



Acharya J C Bose on
Patenting His Invention

Messrs Muirhead & Co, patentees of the well-known system of wireless telegraphy wrote on 13th November, 1900, that they had discussed with Acharya J C Bose some of his results bearing on "Certain practical points in the manufacture of wireless telegraphic apparatus", adding "We have already benefited by your work in the construction of the most important part of such apparatus".

Regarding this Acharya J C Bose wrote in his diary "As a practical outcome of my theory, the head of a great firm working on wireless telegraphy told me that the advantage he derived from the suggestions contained in that paper was beyond anything he could have dreamt of. About my further ideas on the subject he begged me not to make things public but allow him to take out patents. He told me he could make great things out of my ideas. But I cannot find heart to give any part of my life for money-making purposes ...."

In 1901, one of the large manufacturers of wireless apparatus proposed to Bose to sign a remunerative agreement, as to his new type of receiver, but Bose refused.

An American friend surprised with what seemed such unpractical quixotism, forthwith patented the invention in Bose's name in America (1901–1904). But J C Bose would not use his rights and allowed the patent to lapse.

From: *J C Bose and Microwaves – A Collection*