OPEN ACCESS Freely available online

Immunogenetics: Open Access

Research Article

# Gene Mutation in Mammals

**Sudha Bansode***

*Department of Immunogenetics, Satyabama University, S.PG.C. Nagar, K.Vellakulam Post- 625 701, Tamil Nadu, India*

## ABSTRACT

In most organisms genes are segments of DNA molecules. In the broad sense the term 'Mutation' refers to all or any the heritable changes within the genome, excluding those resulting from incorporation of genetic material from other organisms. A mutation is an abrupt qualitative or quantitative change within the genetic material of an organism. Mutations may be intragenic or intergenic. Intragenic mutations or point mutations include alterations within the structure of the DNA molecule within a gene. In a point mutation there is a change in the normal base sequence of the DNA molecule. This change results in a modification of the structural characteristics or enzymatic capacities of the individual. The unit of gene mutation is the muton. This may consist of one or many nucleotide pairs. Intergenic mulations, of which chromosomal changes in structure are examples, involve long regions of DNA, i.e. many genes. These include deletion or addition of segments of chromosomes, resulting in deficiency and duplication respectively. In large deletions a base sequence corresponding to an entire polypeptide chain is sometimes lost. Such mutations are very useful in genetic mapping.

**Keywords:** Intergenic mulations; DNA; Genetic mapping

## INTRODUCTION

Rates of gene mutation are often determined indirectly by estimating the speed at which the neutral substitutions accumulate in protein-coding genes [1]. Synonymous substitutions in protein-coding genes generally are free from natural selection and are used frequently for inferring neutral substitution rates [1, 2]. In particular, the fourfold-degenerate sites are expected to harbor only the neutral substitutions, because all mutations at these sites are synonymous at the amino alkanoic acid sequence level. By using estimates of evolutionary distances based on neutral substitutions, many studies have examined the null hypotheses of uniformity of neutral mutation rates among genes within a genome and among mammalian lineages and have come to conflicting conclusions [3]. For example, significant differences in mutation rates among mammalian lineages reported over the last two decades led to the proposal of the generation time effect hypothesis [4]. However, Easteal et al. have argued that previous results of substantial differences among lineages observed may have been caused by the use of incorrect fossil dates or inappropriate out groups. Similarly, there's significant controversy regarding differences in mutation rate among genes within a genome [5] and over 10-fold differences within the estimates of the mutation rates among studies [6]. One common feature of many of these studies is that they have either analyzed a little number of genes or only a couple of species. Analysis of an outsized sample of genes from a genome and diverse phylogenetic lineages is that the key to testing the null hypothesis of equal mutation rates within and among genomes. A large number of genes are necessary, because only a fraction (15%) of codon positions during a sequence are fourfold-degenerate and that we got to sample genomic regions extensively. Furthermore, mutation rate information from many inter as well as intraordinal mammalian species pairs is important to check whether the observed differences, if any, among mammalian orders are likely to be tied significantly to differences in generation times and physiological attributes among groups. Therefore, we have assembled a data set of 17,208 protein-coding DNA sequences belonging to five, 669 different nuclear genes from a total of 326 placental mammalian species to characterize the extent of difference in mutation rates among genes during a genome and among lineages.

**Correspondence to**: Sudha Bansode, Department of Immunogenetics, Satyabama University, S.PG.C. Nagar, K.Vellakulam Post- 625 701, Tamil Nadu, India, Email: b.sudha@yahoo.com

**Received**: February 10, 2021; **Accepted**: February 24, 2021; **Published**: March 04, 2021

**Citation:** Bansode S (2021) Gene Mutation in Mammals. Immunogenet Open Access. 6:139.

**Copyright:** © 2021 Bansode S. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## MATERIALS AND METHODS

### Data mining and assembly

Phylogenetic trees of 8,627 gene families within the HOVERGEN database release 36 were constructed from amino alkanoic acid sequence alignments by using the neighbor-joining method in MEGA2. The cDNA sequence alignments for orthologous sequence sets then were generated using amino acid sequence alignments as guides. Neighbor joining trees were scanned automatically followed by manual inspection to identify orthologous sequence sets. We enforced strict orthology definitions by considering sequences to be orthologous as long as no gene duplication events were detected since their divergence from the foremost recent common ancestor. All gene families containing fewer than three sequences were excluded, which produced a group of three, 132 gene families. There was a complete of 326 species, with 113 species represented by 2 genes. The ultimate data set assembled 17,208 protein-coding DNA sequences belonging to five, 669 different nuclear genes from a complete of 326 placental mammalian species available within the databanks. The number of sequences available from different species and groups varied extensively: primates, 5,618 sequences; sciurognath rodents, 8,142 sequences; artiodactyls, 2,042 sequences; lagomorphs, 573 sequences; carnivores, 483 sequences; hystricognath rodents, 205 sequence and perissodactyls, 145 sequences.

All computations were done by using only the fourfold degenerate site for sequence pairs. We took a stringent approach in identifying fourfold-degenerate sites by selecting only those sites that have remained fourfold-degenerate throughout the evolutionary history of the pair of species compared. This task was accomplished by designating a site as four fold degenerate only if it was so in both the sequences compared. It shows the distribution of the amount of fourfold-degenerate sites in 3,722 genes in the human-mouse comparison.

### Estimation of evolutionary distance

Evolutionary divergence (d4) between sequences at fourfold-degenerate sites was estimated by using the Tamura–Nei method to correct for multiple hits by accounting for transition transversion rate and base frequency biases. For a given species pair, multigene evolutionary distance was computed by taking the typical of evolutionary distance over all genes. For two groups of species, the evolutionary distance was estimated by first computing average gene distances between species belonging to the 2 groups then taking a mean of those distances over all genes.

### Estimation of expected variance for a multigene distribution

The expected amount of variance within the distribution of multigene distances for a given pair of species is that the sum of the estimation variation (Ve) contributed by the utilization of distance methods to correct for multiple hits and therefore the variance contributed by the stochastic nature of the evolutionary process (Vs). For a set of N independent genes, Ve i(Ve,i)N, where Ve,i is the estimation variance for gene i. It is computed by using Tamura–Nei's variance formula, VTN, as Ve,i VTN diLi, where Li is the number of sites, and di is the Tamura–Nei distance. Under the null hypothesis of equal mutation rate per site, Vs is obtained by considering a Poisson process governing the arrival of mutations at a finite number of fourfold-degenerate sites in a given gene. For Ngenes, it is given by Vs i ( Li)N.

### Determination of physical location of genes in the human genome map

Currently the human genome map consists of relative positions of a large number of contigs on each chromosome. The gene content of every contig was obtained from the NCBI ftp site. We first mapped the GenBank accession numbers of human genes in our data set to their corresponding unique LocusIDs by using the LocusLink public resource and then constructed an entire map for all human sequences included in this study. Chromosomal locations of mouse genes also were obtained from LocusLink. For the analysis of gene proximity and mutation rate, the physical distance between a gene pair on human chromosome was estimated by subtracting the ending nucleotide position of the primary gene from the starting nucleotide position of the second gene (Table 1).

| **Database** | **HOVERGEN database (20) release 36** |
|---|---|
| Placental mammalian species | 326 |
| No. gene families | 8,627 |
| No. protein-coding DNA sequences | 17,208 |
| Method | neighbor-joining |
| Relative rate difference (primates vs. rodents) | 34% |
| **Species** | **No. of sequences** |
| Primates | 5,618 |
| Sciurognath rodents | 8,142 |
| Artiodactyls | 2,042 |
| Agomorphs | 573 |
| Carnivores | 483 |
| Hystricognath rodents | 205 |
| Perissodactyls | 145 |

**Table 1:** Phylogenetic tree construction and relevant parametric data.

## RESULTS AND DISCUSSION

### Homogeneity of substitution patterns between lineages

Although the fourfold-degenerate sites are expected to accumulate only synonymous substitutions, the evolutionary distances estimated by using these sites are useful in estimating the underlying mutation rate as long as the nucleotide substitutions have accrued with the same substitution pattern in the two species compared. That is, the homologous sites within the two sequences compared during a given gene must have evolved with an equivalent instantaneous substitution matrices. Substitution patterns during a given gene may shift in one lineage as compared with its orthologous counterpart for variety of reasons including chromosomal rearrangements, gene transfer, or centromere movement (e.g., mouse genome). In these cases, substitution patterns in genes could also be affected to repair mutations that make the bottom composition of the gene to be more almost like its chromosomal location amelioration effect, and this will be more pronounced at the sites that are selectively neutral. Therefore, the substitution rate at neutral sites in those genes is going to be above the particular mutation rate, rendering such genes unsuitable for inferring mutation rates. Therefore, we conducted the disparity index test for every pair of orthologous sequences to spot genes during which fourfold-degenerate sites aren't evolving with homogeneous substitution patterns among the lineages compared. The disparity index test directly examines the null hypothesis of homogeneity of the evolutionary pattern between two lineages by testing whether the observed difference in nucleotide frequencies between sequences is quite that expected accidentally alone, given the amount of differences observed between sequences.

Such results seem to be caused by the inclusion of genes evolving with heterogeneous substitution patterns, because we discover that these genes show a way larger relative rate difference (34%) between primates and rodents than those from genes that pass the disparity index test for homogeneity of substitution patterns (9%). Therefore, fixation of mutations under heterogeneous substitution pattern within the orthologous sequences instead of the difference in mutation rates is probably going to be the cause for the results reported previously. The absence of serious correlation between the mutation rate and generation time is probably going to prompt a reassessment of the generation time effect hypothesis.

## CONCLUSION

In conclusion, our results argue against the widely held notion about large differences in mutation rates among genes during a genome and among major mammalian lineages. This approximate similarity of mutation rates among genes and among lineages is probably going to be important for estimating divergence time for closely related species, testing for selection by comparative sequence analysis, inferring coalescent times, and understanding the mutational processes that govern evolution of mammalian genomes.

## REFERENCES

1. World Health Organization. Framework for verifying elimination of measles and rubella. Relev Epidemiol Hebd. 2013;88:89-99.
2. Best JM, Banatvala JE. Rubella. In: Zuckerman AJ, Banatvala JE, Pattison JR, Griffiths PD, Schoub BD, editors. Principles and practice of clinical virology. 5th ed. West Susex Wil. 2004;8:427-457.
3. Ethiopia Health Topics, Ethiopia Immunization, WHO Country Office Ethiopia. 2015.
4. Robertson SE, Featherstone DA, Gacic-Dobo M, Hersh BS. Rubella and congenital rubella syndrome: global update. Rev Panam Salud Publica. 2003;14:306–315.
5. Getahun M, Beyene B, Gallagher K, Ademe A, Teshome B, Tefera M, et al. Epidemiology of rubella virus cases in the pre-vaccination era of Ethiopia. BMC Public Health. 2016;16:11-68.
6. Getahun M, Beyene B, Gallagher K, Ademe A, Teshome B, Tefera M, et al. Epidemiology of rubella virus cases in the pre-vaccination era of Ethiopia. BMC Public Health. 2016;16:11-68.