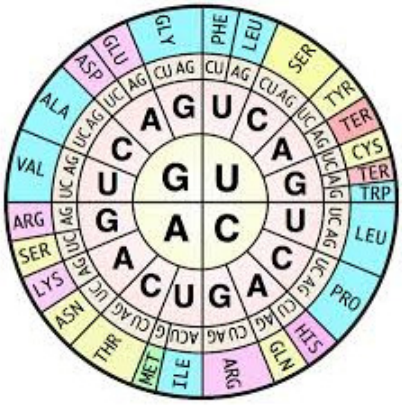
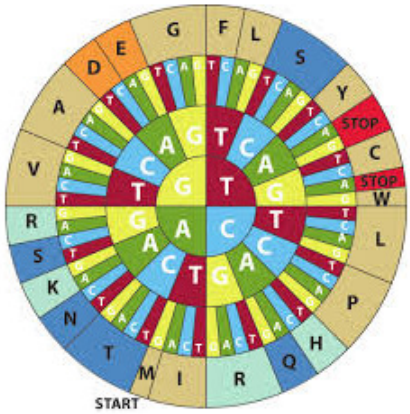
CHAPTER-2

REVIEW OF LITERATURE





2. Review of Literature

Comparisons of mitochondrial systems are useful for modeling genome evolution and phylogenetic inference. These include gene content and gene arrangement, base composition, modes of replication and transcription, protein, tRNA and rRNA gene secondary structures, and genetic codon variations. During the last 10 years, mitochondrial genome sequence and gene arrangement comparisons were employed as powerful new tools for resolving ancient phylogenetic relationships.

Chaney *et al.* (2015) reported the functional effects of codon usage and its general roles. Synonymous codon usage was once considered to be functionally neutral, but evidence later on suggested that it was formed by evolutionary selection and that it affects other features of protein biogenesis ahead of specifying the amino acid sequence of the protein. Rare codons, once believed to have only negative effect on the speed and accuracy of translation, are currently identified to play an essential role in various functions, such as expression level, regulation of co-translational folding, covalent modifications and secretion. Mutations changing synonymous codon usage are associated with human diseases (Chaney *et al*. 2015).

Uddin *et.al* (2015) studied codon bias and gene expression of mitochondrial ND2 gene in chordates. Mitochondrial ND gene, which encodes NADH dehydrogenase, is the foremost enzyme of the mitochondrial electron transport chain. The study emphasised on the analysis of compositional properties and selection pressure on the codon usage patterns in the coding sequence of MT-ND2 gene across pisces, aves and mammals by using bioinformatics approach like relative synonymous codon usage (RSCU), effective number of codons (ENC) and codon adaptation index (CAI). Codon usage bias is low as reflected by high ENC values in MT-ND2 gene among pisces, aves and mammals in this study. The most frequently used codons end with A/C and the gene was AT rich in all the three classes. The over-represented codons were TCA, CTA, CGA and TGA in all three classes. The F1 of correspondence analysis showed significant positive correlation with G, T3 and CAI whereas the F2 of correspondence analysis showed significant negative correlation with A and T but significant positive correlation with G, C, G3, C3, ENC, GC, GC1, GC2 and GC3. Codon usage bias in MTND2 gene is not associated with the expression level. Both mutation pressure and natural selection affect the synonymous codon usage pattern in MT-ND 2 gene (Uddin *et al*.).

Duan *et al.* (2015) studied comprehensive analysis of codon usage patterns in Blunt Snout Bream (*Megalobrama amblycephala*) based on RNA-Seq Data**.** In *M. amblycephala*, high-frequency codons (CUG, AGA, GUG, CAG and GAG), as well as low-frequency ones (NUA and NCG codons) were identified based on RNA seq data. They reported codon usage bias close to start and stop codons showed apparent heterogeneity, which still occurs along the nucleotide sequence. Codon usage bias (RSCU and SCUO) was correlated to GC3 (GC content of 3rd nucleotide in codon) bias. Their studies indicated species specificities by using GC contents, codon usage and codon context analysis (Duan *et al.* 2015).

Subramanian and Sarkar (2015) performed analysis of codon usage patterns between *Leishmania* and other known Trypanosomatid species to know the variations in gene organization and its effect on the phenotype across different *Leishmania* species. They reported the causes and consequences of codon usage bias in *Leishmania* genomes and established GC bias at 3rd codon position that directs codon usage bias in different *Leishmania* species, rather than composition of amino acid bias. They reported that, inside *Leishmania*, homogenous codon contexts coding for fewer common amino acid pairs and codons avoiding the creation of folding structures in mRNA are effectively selected. This study explains the function of evolution in shaping the otherwise conserved genome to reveal species-specific differences in function-level for efficient survival (Subramanian and Sarkar 2015).

Choudhury and Chakraborty (2015) studied SPANX (sperm protein coupled with the nucleus in the X chromosome) genes in human Y chromosome. Their results illustrate that codon usage bias is low and most of the GC ending codons were positively correlated with GC3 bias. Both mutation pressure and natural selection affect the codon usage pattern in SPANX genes, and the role of natural selection is higher than mutation pressure as revealed from the neutrality plot (Choudhury and Chakraborty 2015).

Dohra *et al.* (2015) reported amino acid and codon usage in *Paramecium bursaria.* Unequal usage of synonymous codon and amino acid in highly expressed genes can reveal a balance between translational selection and other factors. Gene expression level with synonymous codon or amino acid usage correlation is emphasized in genes down-regulated in symbiont-bearing cells compared to symbiont-free cells. Their results indicate that the selection is related with *P. bursaria–Chlorella* symbiosis (Dohra *et al*. 2015).

Doupa *et al.* (2015) determined the implication of the mutations of Cytochrome b in the evolution of breast benign tumors among Senegalese women to find out the effect of the

natural selection on the observed variability. Their result showed that the population of benign cells was growing rapidly from an ancestral population sparse. These findings have allowed to believe the hypothesis that Cytochrome b mutations would be involved in the evolution of breast benign tumors among Senegalese women (Mbaye *et al.* 2012).

Uddin and Chakraborty (2015) studied synonymous codon usage pattern in mitochondrial CYB gene in pisces, aves, and mammals using various codon usage parameters. They reported unequal distribution of nucleotide composition such as A, T, G and C in pisces, aves and mammals in CYB gene.. Heat map showed that AT-ending codons were mostly negative and GC-ending codons were mostly positive. The codons absent in pisces were AGT (except *Toxotes chatareus*), TGT, and CAG (except *Elasma zonatum*). The codons such as AGT (except *Falco peregrinus*), CGT (except *Vidua chalybeata*), and ACG (except *Aythya americana*) were absent in aves whereas, in mammals, the absent codons were namely CAG (except *Canis familiaris*) and ACG (except *Rattus norvegicus*). Codon usage bias was weak in pisces, aves, and mammals. Correlation analysis further suggests that mutation pressure is mainly responsible for codon usage pattern suggested from correlation analysis (Uddin and Chakraborty 2015).

Supek (2015) reported that mutations in coding region substitute one synonymous codon to another but do not affect the amino acid sequence for encoding the protein which may exhibit a plethora of effects on the living cell. Therefore, they are regularly selected during evolution, and originating from synonymous codon usage biases in genomes. The author performed a comparative study of bacterial, archaeal, fungal, and human cancer genomes and have found many associates between the accrual of synonymous mutations and a gene's biological role during evolution. In highly expressed genes, certain functional groups are enriched with optimal codons, which are decoded by the abundant tRNAs, thus enhancing the accuracy and speed of the translating ribosome. The codon adaptation differs in the set of genes between genomes, and these differences illustrate vigorous associations to organismal phenotypes. They also found other distinct codon bias patterns in cyclically expressed genes, amino acid starvation genes, oxidative stress response genes, tissue-specific genes in animals and plants, cellular differentiation genes, and oncogenes. His work also highlights the evolutionary trace of codon bias patterns across orthologous genes which may be studied to learn about a gene's importance to different phenotypes, or, more generally, its function in the cell (Supek 2015).

Dilucca *et al.* (2015) studied codon bias patterns of *E. coli's* interacting proteins. They proposed a new parameter for codon bias index, CompAI, which is based on the

competition between cognate and near-cognate tRNAs during translation, exclusive of codon usage bias of highly expressed genes in order to shed light on codon usage pattern on *E. coli's* interacting proteins. They performed a comparison of CompAI with other widely used indices: tAI, CAI, and Nc in genome-wide evaluation of codon bias for *E.coli*. They observed that CompAI and tAI capture related information by being positively correlated with gene conservation, measured by the Evolutionary Retention Index (ERI), and essentiality, whereas, CAI and Nc appeared to be less responsive to evolutionary-functional parameters. Their work also targeted to study the correlation of codon bias at the genomic level with the network character of protein-protein interactions in *E.coli*. They found that the most closely connected communities of the network share a similar level of codon bias (as measured by CompAI and tAI). Essentially, among all the parameters of codon bias, CompAI turns out to have the most coherent distribution over the communities of the interactome, pointing to the importance of competition among cognate and near-cognate tRNAs for elucidating codon usage adaptation. Especially, CompAI may potentially correlate with translation speed measurements, by accounting for the specific hold-up induced by wobble-pairing between codons and anticodons (Dilucca *et al*. 2015).

Li *et al.* (2015) reported that GC-content of synonymous codons strongly persuades amino acid usage. They found the overall usages of the four amino acids with the most GC-synonimous (GCsyn) and the five amino acids with the lowest GCsyn and both differ with the regional GC-content, while the usage of the remaining 11 amino acids with intermediate GCsyn is less changeable. Their study reveals that codon usage frequencies are almost constant in regions with similar GC-content. Their results imply that GCsyn correlates with GC content and has an effect on codon/amino acid usage, which suggests a novel way to know the function of codon and amino acid usage in determining genomic architecture and organism's evolutionary patterns (Li *et al*. 2015).

Hussain *et al.* (2015) investigated molecular diversity and phylogenetic analysis in Punjab Urial (*Ovisvignei punjabiensis*) in mitochondrial ATP6 and ATP8 genes. In Pakistan the Punjab Urial (*O. vigneipunjabiensis*), is an important sub-species of *O. vignei,* but a limited molecular data is reported that urged them to explore its genetic diversity and phylogeny using mitochondrial DNA, ATP6 and ATP8 genes. Their study provided valuable genomic information regarding genetic diversity in Punjab Urial and its phylogenetic relationships with related taxa, emphasizing on the need of execution of conservation strategies to protect this unique genetic resource of Pakistan (Hussain *et al.*).

Jia *et al.* (2015) reported codon usage pattern and factors in *Bombyx mori* and it could be helpful to improve the bioreactor based on *B. mori*. The codon bias in *B. mori* is weak based on a total of 1,097 annotated mRNA sequences. It also shows that the gene expression level is related to the GC content, and the amino acids with higher general average hydropathicity (GRAVY) and aromaticity (Aromo) and effective number of codon (ENC). The codon usage bias is mainly influenced by nucleotide composition, mutation pressure, natural selection, and expression level. Furthermore, the "optimal codons" of *B. mori* are all encoded by G and C, which gives useful information for increasing the gene expression in *B. mori* through codon optimization (Jia Xian *et al.* 2015).

Lakshmanan *et al.* (2015) highlighted that the non D-loop direct repeats (DRs) in mitochondrial DNA (mtDNA) have been usually implicated in the mutagenesis of mtDNA deletions associated with neuromuscular disease and ageing. In addition, these DRs have been hypothesized to put a restraint on the lifespan of mammals and are under a negative selection pressure. They examined the relationship between the mutagenicity of such DRs and the lifespan of species using 294 mammalian mtDNA, and found no evidence why long lived mammals possess fewer mutagenic DRs than short-lived mammals. Further they compared DR counts in human mtDNA and in selectively randomized sequences, and showed that the number of DRs in human mtDNA is mainly determined by global mtDNA properties, such as nucleotide composition and the biasness in synonymous codon usage (Lakshmanan *et al*. 2015).

Shen *et al.* (2015) compared the strength of GC3 bias of genes in human and mouse using relative GC3 bias values. They reported that GC3-rich and GC3-poor gene products might have distinct subcellular spatial distributions. Their results indicated that similar GC3 biased genes might be co-translated in specific spatial regions to share local translational machineries, and that GC3 could be involved in the organization of genome architecture (Shen *et al*. 2015).

Kelkar *et al.* (2015) investigated that in *C. crescentus*, the expression of the more A+T-rich gene variants slow down growth, signifying that selection on genic base composition is, partly responsible for the high G+C content of this genome. In comparison, no similar effect was observed in *P. aeruginosa*, which has similarly high genomic G+ C contents. Selection for improved genic G+C-contents in *C. crescentus* acts alone of the species-specific codon usage pattern and signifies an additional selective force acting in bacterial genomes (Kelkar *et al*. 2015).

Sharma *et al.* (2014) analyzed codon usage in two Hemipteran insect species namely *Bemisia tabaci* and *Homalodisca coagulate* and found ENC (a measure of codon bias) value ranges from 43 to 60 (52.80) in *B. tabaci* but from 49 to 60 (56.69) in *H. coagulata*. In both insect species, a significant positive correlation was observed between A and A3%, C and C3%, and GC and GC3%, respectively, which suggest that mutation pressure causes the codon usage pattern in two Hemipteran insect species (Sharma *et al*. 2014).

Mirsafian *et al.* (2014) reported synonymous codon usage bias patterns for the evolutionarily close proteins of albumin superfamily, namely, albumin, α-fetoprotein, afamin, and vitamin D-binding protein. They demonstrated that the genes of the four albumin superfamily members have low GC content and high values of the effective number of codons (ENC) suggesting high expressivity of these genes and less bias in codon usage preferences. Their study also highlighted the evidence that the albumin superfamily members are not subjected to mutational selection pressure (Mirsafian *et al.* 2014).

Uddin and Chakraborty (2014) found that GC content varies from 27.5 to 46.4% with a mean of 37.49±5.12 and GC content at 3rd codon position varies from 16.2 to 47.8% with a mean of 34.11±9.76 in mitochondrial ATP 8 gene. ENC values are much higher and range from 42 to 60 with a mean of 54.1±5.93, which suggest that low codon usage bias. A significant positive correlation was observed between A% and A3% ($r=0.6502$, $p<0.05$), C% and C3% ($r=0.9804$, $p<0.001$), GC% and GC3% ($r=0.9732$, $p<0.001$) and significant negative correlation was found for most other nucleotides comparisons which indicates that mutation pressure might influence the codon usage pattern in ATP8 gene (Uddin and Chakraborty).

Chen *et al* (2014) found neutrality analysis did not find a significant correlation between GC12 and GC3 but ENc- GC3 plot suggested that mutational bias played a major role in shaping the codon usage. Parity Rule 2 plot (PR2) analysis of their study showed that GC and AT were not used proportionally and proposed that codons containing A or U at third position are used preferentially in nemertean species, in spite of whether corresponding tRNAs are encoded in the mitochondrial DNA. They also found that the nucleotide at the second codon position slightly affects synonymous codon choices based on the context-dependent analysis. Their results suggested that mutational and selection forces were most likely acting to codon usage bias in nemertean mitochondrial genomes (Chen Haixia *et al.* 2014).

Ding *et al.* (2014) found that there is no link between the variation of the overall tRNA abundance and the specific folding units in the nsp1α, of porcine reproductive and respiratory syndrome virus and the low translation speed of ribosome caused by the tRNA abundance exists in the *nsp1*α. They found high correlation between some synonymous codon usage and the specific folding units in the nsp1α and the phenomenon of CDCB exists in the specific folding units of the nsp1α. Their findings provide an insight into the roles of the synonymous codon usage and CDCB in the formation of PRRSV nsp1α structure (Ding *et al*. 2014).

Butt *et al.* (2014) reported the patterns of synonymous codon usage in 141 CHIKV genomes by calculating several codon usage indices and applying multivariate statistical methods. The preferred synonymous codons were G/C and A-ended as suggested from relative synonymous codon usage (RSCU) analysis. Comparison of RSCU between CHIKV and its hosts illustrates that codon usage patterns of CHIKV are a mixture of coincidences and antagonism. The overall codon usage patterns of CHIKV have been strongly influenced by *Pan troglodytes* and *Aedes albopictus* during evolution as revealed from similarity index analysis. Effective number of codon suggested that overall codon usage bias was low in CHIKV genomes.Their result suggested that although mutation pressure dominates codon usage in CHIKV, patterns of codon usage in CHIKV are also under the influence of natural selection from its hosts and geography (Butt *et al.* 2014).

Chen *et al.* (2013) analyzed synonymous codon usage patterns of the mitochondrial genomes in 43 parasitic platyhelminth species. Among the mitochondrial genomes of 19 trematode species, the GC content of third codon positions varied from 0.151 to 0.592, with a mean of 0.295 ± 0.116. In cestodes, the mean GC content of third codon positions was 0.254 ± 0.044. A comparison of the nucleotide composition at 4-fold synonymous sites revealed that, on average, there was a greater abundance of codons ending on U (51.9%) or A (22.7%) than on C (6.3%) or G (19.14%). Twenty-two codons, including UUU, UUA and UUG, were frequently used. They found, in addition to compositional constraints, the degree of hydrophobicity and the aromatic amino acids also influenced codon usage in the mitochondrial genomes of these 43 parasitic platyhelminth species (Chen L *et al*. 2013).

Campos *et al.* (2013) reported codon usage bias (CUB) in *Drosophila* is higher for X-linked genes than for autosomal genes. Their result shows that selection for preferred versus un-preferred synonymous variants is stronger on the X chromosome than the

autosomes, which accounts for the higher CUB of genes on the X chromosome. The stronger selection on CUB on the X chromosome leads to a lower rate of synonymous site divergence compared with the autosomes; this will cause a stronger upward bias for X chromosome than A in estimates of the proportion of nonsynonymous mutations fixed by positive selection, for methods based on the McDonald–Kreitman test (Campos *et al.* 2013).

Nair *et al.* (2013) reported mutational pressure dictates synonymous codon usage in freshwater unicellular α – cyanobacterial descendant *Paulinella chromatophora* and β - cyanobacterium *Synechococcus elongates* PCC6301. They investigated various factors associated with synonymous codon usage in the genomes of *P. chromatophora* and *S. elongatus* PCC6301 and found mutational pressure was the major force behind codon usage variation in both genomes but correspondence analysis revealed that the intensity of mutational pressure was higher in *S. elongatus* than in *P. chromatophora*. (Nair *et al*. 2013).

Xu *et al.* (2013) discovered a relationship between codon usage and a general property of circadian rhythms called conditionality. They showed that in the cyanobacterium *Synechococcus elongate*, non-optimal codon usage was selected as a post-transcriptional mechanism to switch between circadian and non-circadian regulation of gene expression as an adaptive response to environmental conditions. Their results indicate that natural selection against circadian systems in cyanobacteria that are intrinsically robust at cool temperature and modulation of circadian amplitude is therefore, crucial to its adaptive significance. Furthermore, these results illustrate the direct effects of codon usage on a complex phenotype and organismal fitness. Their work also challenges the long-standing view of directional selection towards optimal codons 4–7, and provides a key example of natural selection against optimal codons to achieve adaptive responses to environmental changes (Xu *et al.* 2013).

Chung and Lee (2012) developed novel computational procedures for estimating the relative value of optimizing individual codon usage (ICU) and codon context (CC) for enhancing protein expression. In their *in silico* validation of the resultant optimized DNA sequences for *Escherichia coli*, *Lactococcus lactis*, *Pichia pastoris* and *Saccharomyces cerevisiae* suggests that CC is a more relevant design criterion than the normally considered ICU. The CC optimization framework can match and increase the capabilities of current gene design tools, with potential applications to heterologous protein production and even vaccine development in synthetic biotechnology (Chung and Lee 2012).

Behura and Severson (2012) analyzed 15 species of dipteran and seven species of hymenopteran insects. They found GAA is the most frequent codon in dipteran species while GAG is preferred in the hymenopteran species. Codons ending with G or C were frequently used in dipteran species and those ending with A or T is frequently used in hymenopteran species. Synonymous codon usage pattern varied within genomes in a pattern that seemed to be distinct for each species. They found negative correlation between the effective number of codon and codon adaption index indicating that codon usage bias of genes of these species had very distinct relationships with nucleotide composition of coding sequences (Behura and Severson 2012).

Rao *et al.* (2011) carried out chicken genome analysis by the use of the relative synonymous codon usage (RSCU) method and identified 11 putative optimal codons, all of them ending with U, which is significantly departing from the pattern, observed in other eukaryotes. In the chicken genome, optimal codons are probably the ones equivalent to highly expressed transfer RNA (tRNAs) or tRNA gene copy numbers in the cell. They also found codon bias, measured as the frequency of optimal codons (Fop), is negatively correlated with the GC content, recombination rate, but positively correlated with gene expression, protein length, gene length and intron length. The positive correlation of codon bias and protein, gene and intron length are the same as that of unicellular organisms but are quite different from other multi-cellular organisms. Their study suggests that both mutation bias and selection contribute to codon bias. However, mutation bias is the driving force of the codon usage in the *Gallus gallus* genome. They also provide evidence that the negative correlation between codon bias and recombination rates in *Gallus* and is determined mostly by recombination-dependent mutational patterns (Rao *et al.* 2011).

Sharp *et al.* (2010) showed that the strength of selected codon usage bias is highly correlated with the bacterial growth rate, suggesting that selection has favoured translational efficiency. They reported that highly expressed genes in *Escherichia coli* appear to be under continuing strong selection, whereas selection is very weak in genes expressed at low levels (Sharp *et al*. 2010).

Jia and Higgs (2008) reported that frequencies of bases at 4-fold degenerate sites are strongly influenced by context-dependent mutation, which causes correlations between pairs of neighboring bases. They compared the codon usage in conserved and variable sites in the same genes to detect selection for translational accuracy. Selection for translational efficiency might lead to preference for codons that match the limited repertoire of

anticodons on the mitochondrial tRNAs. In their mitochondrial genes under study, there is very little evidence for translational accuracy selection. They found several cases where unusual bases occur at the wobble position of the tRNA, and in these cases, some evidence for selection on codon usage was found and suggests the way that these unusual cases are associated with codon reassignments in the mitochondrial genetic code (Jia Wenli and Higgs 2008).

Mukhopadhyay *et al.* (2007) studied synonymous codon usage in different protein secondary structural classes of human genes using 401 *Homo sapiens* proteins extracted from Protein Data Bank (PDB). It was found that synonymous codon families show non-randomness in codon usage in four different secondary structural classes. On the other hand, when the genes were classified according to their GC3 levels there was an increase in non-randomness in high GC3 group of genes. They also reported that in high GC3 group of genes, the non-random behaviour of synonymous codons increases in sheet structure of all secondary structural class of protein (Mukhopadhyay *et al.* 2007).

Gibson *et al.* (2004) carried out a study on the composition of 69 mammalian mitochondrial genomes. Variation in base composition across genes and species is known to adversely affect the performance of phylogenetic inference methods. They, therefore, developed a customized three-state general time-reversible DNA substitution model, implemented in the PHASE phylogenetic inference package, which lumps C and T into a composite pyrimidine state. They compared the phylogenetic tree obtained using the new three-state model with that obtained using a standard four-state model. Their results in a three-state model are more congruent with recent studies using large sets of nuclear genes and help decide some of the apparent conflicts between studies using nuclear and mitochondrial proteins (Gibson *et al.* 2005).

Gustafsson *et al.* (2004) studied the codon bias and heterologous protein expression. The expression of functional proteins in heterologous hosts is a corner stone of modern biotechnology. Improvements in the speed and cost of gene synthesis have facilitated the complete redesign of entire gene sequences to maximize the likelihood of high protein expression. They discussed the redesign strategies, including modification of translation initiation regions, alteration of mRNA structural elements and use of different codon biases (Gustafsson *et al.* 2004).

Rocha (2004) suggested a much more universal trend in the evolution of anticodon and codon choice and provided new evidence that a selective force for the optimization of the translation machinery is the maximization of growth (Rocha 2004).

Urrutia and Hurst (2001) presented a new way to measure codon bias that corrects for background nucleotide content using 2396 human genes. They observed nearly all genes (99%) exhibit a higher amount of codon bias than expected by chance, and the patterns connected with selectively driven codon bias are weakly recovered. Broadly expressed genes have a higher level of bias than do tissue-specific genes; the bias is higher for genes with lower rates of synonymous substitutions, and certain codons are repeatedly preferred (Urrutia and Hurst 2001).

Karlin and Mrazek (1996) presented data that codon choices for human genes are largely a consequence of two factors: (1) amino acid constraints, (2) maintaining DNA structures dependent on base-step conformational tendencies consistent with the organism's genome signature that is determined by genome-wide processes of DNA modification, replication and repair. The related codon signature, defined as the dinucleotide relative abundances at the distinct codon positions {1,2}, {2,3}, and {3,4} (4 = 1 of the next codon), accommodated both the global genome signature and amino acid constraints. In human gene, codon positions {2,3] and {3,4} containing the silent site have similar codon signatures reflecting DNA symmetry. Strong CG and TA dinucleotide underrepresentation is observed at all codon positions as well as in non-coding regions. Estimates of synonymous codon usage based on codon signatures are in excellent agreement with the actual codon usage in human and general vertebrate genes (Karlin and Mrázek 1996).

Hay *et al.* (1995) used nucleotide sequence comparisons to investigate ordinal and familial relationships within the class Amphibia. Approximately 850 base pairs of the mitochondrial 16S ribosomal RNA (rRNA) gene from representatives of 28 of the 40 families of extant amphibians were sequenced. Phylogenetic analyses of these data together with published data of the 12S rRNA gene for the same families and both genes for three more taxa (approximately 1,300 base pairs total for 35 taxa) support the monophyly of each of the three amphibian orders: Anura (confidence value with the interior-branch test: P(c) = 99%), Caudata (P(c) = 100%), and Gymnophiona (P(c) = 99%) (Hay *et al.* 1995).

Zardoya and Meyer (1997) collected about 3,500 base pairs from seven species of the large 28S nuclear ribosomal gene. All phylogenetic analyses (maximum parsimony, neighbor-joining, and maximum likelihood) point toward the hypothesis that lungfishes and coelacanths form a monophyletic group and are equally closely related to land vertebrates. This evolutionary hypothesis complicates the identification of morphological or physiological preadaptations that might have permitted the common ancestor of tetrapods

to colonize land. This is because the reconstruction of its ancestral conditions would be hindered by the difficulty to separate uniquely derived characters from shared derived characters in the coelacanth/lungfish and tetrapod lineages. This molecular phylogeny aids in the reconstruction of morphological evolutionary steps by providing a framework; however, only paleontological evidence can determine the sequence of morphological acquisitions that allowed lobefinned fishes to colonize land (Zardoya and Meyer 1997).

Zardoya *et al.* (1998) analyzed the phylogenetic relationships of the African lungfish (*Protopterus dolloi*) and the coelacanth (*Latimeria chalumnae*) with respect to tetrapods were analyzed using complete mitochondrial genome DNA sequences. A lungfish + coelacanth clade was favored by maximum parsimony (although this result is dependent on which transition : transversion weights are applied), and a lungfish + tetrapod clade was supported by neighbor-joining and maximum-likelihood analyses. These two hypotheses received the strongest statistical and bootstrap support to the exclusion of the third alternative, the coelacanth + tetrapod sister group relationship. All mitochondrial protein coding genes combined favor a lungfish + tetrapod grouping. When the complete mitochondrial sequence data were combined with nuclear 28S rRNA gene data, a lungfish 1 coelacanth clade was supported by maximum parsimony and maximum likelihood, but a lungfish + tetrapod clade was favored by neighbor-joining. The seemingly conflicting results based on different data sets and phylogenetic methods were typically, not statistically, strongly supported based on Kishino-Hasegawa and Templeton tests, although they were often supported by strong bootstrap values. Differences in the rate of evolution of the different mitochondrial genes (slowly evolving genes such as the cytochrome oxidase and tRNA genes favoured a lungfish + coelacanth clade, whereas genes of relatively faster substitution rate, such as several NADH dehydrogenase genes, supported a lungfish + tetrapod grouping), as well as the rapid radiation of the lineages back in the Devonian, rather than base compositional biases among taxa seem to be directly responsible for the remaining uncertainty in accepting one of the two alternate hypotheses (Zardoya and Meyer 1998).

Rasmussen and Arnason (1998) sequenced the complete mitochondrial genome of the spiny dogfish, *Squalus acanthias,* and included it in a phylogenetic analysis together with a number of bony fishes and amniotes. The phylogenetic reconstructions placed the dogfish among the bony fishes. Thus, contrary to the common view, the analyses have shown that the position of the sharks is not basal among the gnathostomes. The presently recognized phylogenetic position of the dogfish was identified irrespective of the out

group used, echinoderms or agnathan fishes. The lungfish was the most basal gnathostomes fish, while the teleosteans had an apical position in the piscine tree. A basal position of the dogfish among the gnathostomes was statistically rejected, but the phylogenetic relationship among the coelacanth, spiny dogfish, and teleosts was not conclusively resolved. The results questioned the status of several morphological characters commonly used in piscine phylogenetic reconstruction. The study also confirmed recent findings demonstrating that the origin of the amniotes is deeper than the diversification of extant bony fishes (Rasmussen and Arnason 1999).

Zhang *et.al* (2009) determined the complete nucleotide sequence of the mitochondrial (mt) genome of the large-headed frog, *Limnonectes bannaensis* (Amphibia, Anura) by using polymerase chain reaction (PCR). The entire mtDNA sequence is 16,867 bp in length with a novel case of tRNAs in vertebrates. This mt genome is characterized by three distinctive features: (1) a tandem duplication of tRNA (Met) gene is observed, (2) the tRNA (Ala), tRNA (Asn), tRNA (Cys) and tRNA (Glu) genes coded on the L-strand are absent from the *L. bannaensis* mtDNA, the tRNA (Cys) and tRNA (Glu) genes change into tRNA pseudogenes by reason of degenerative anticodon, and a noncoding sequence of 206 nt long (NC1) has replaced the original position of other two tRNAs, (3) besides NC1, another three noncoding spacers (NC2-4) longer than 50 bp are found in the broken WANCY region and the region NC3-ND5-NC4-ND6-PsiE-Cytb-CR of the new sequence. These features could be explained by a model of gene duplication and deletion. The new sequence data was used to assess the phylogenetic relationships among 25 species of Anura using neighbor-joining, Bayesian, maximum likelihood methods and the phylogenetic tree showed the rice frog *Fejervarya limnocharis* is closest to *L. bannaensis* in the reported study (Zhang Ji-Feng *et al*. 2009).