Chapter 7
Perspectives

In this thesis, I have used 3 parameters to construct multiparametric phylogenetic trees. However, to obtain the Consensus Phylogenetic Tree, several hundreds or thousands of parameters have to be integrated. According to the Central Dogma of Molecular Biology, the genetic information in the Genes is coded in form of nucleotide sequence in DNA and then retrieved as RNA intermediates and translated into polypeptides that constitute the molecular phenotypes. Even though, several genomic sequences are available today, their application to construct multiparametric phylogenetic trees is limited, because precise segments containing the translatable information is not yet fully annotated for sufficient number of species as the existing tools for gene annotation are computationally expensive and based on heuristic algorithm; thus subject to inaccuracies. Therefore, an accurate annotation tool is also necessary to construct multiparametric phylogenetic trees. Furthermore, every morphological character of an organism constitutes a polygenic trait and is a result of expression of multiple genes. Darwin's phylogenetic tree as well as well established Linnaean taxonomical classification is based on morphological characteristics as its basis. Hence, analyzing multiple genes is a powerful way of obtaining a consensus phylogenetic tree.

In my studies, I have used only 3 parameters to obtain a phylogenetic tree and the methodology can be extended to any number of them by constructing trees from mean Euclidean distances of any multiples of 3 parameters placed in a 3D space. In fact, recently the method was extended to the construction of phylogenetic tree for 130 species of Bacteria and Archaea based on comparison by MSA of 15 different Aminoacyl tRNA synthetases (R. Bargaje, M. Milner and S.P. Modak., 2009, unpublished). This shows the scalability of my
method that can be used for multiple genes from genomes and it would be interesting to apply it to all coding regions from full genomes.

In a phylogenetic tree, the divergence time is an important factor, which I have not integrated. Therefore, I visualize that divergence time should be integrated with multiparametric phylogenetic, which would become more meaningful than the one without divergent time.

Examination of our multiparametric phylogenetic tree clearly offers the possibility that terminal nodes/species that are positioned on relatively distant branches may converge in the 3-D space, which would be suggestive of Convergent Evolution. To test this concept, however, the sample size should be sufficiently large, say, at least 500 genes or polypeptides. Furthermore there are other topological considerations that may contain information that needs to be extracted and exploited. For example, in a phylogenetic tree positioned in 3D space, the branches are positioned at different angles with reference to the parental or ancestral stem. Similarly, the distance between the siblings is not necessarily identical and it is possible that both the intra-nodal angles and inter-nodal distances may provide greater insight on the fine structure of such a phylogenetic tree. The evolutionary distance between any two nodes may be more than the distance between them in the 3D representation. This shows the possibility of convergent evolution. Therefore, the intra- and inter-clade angles and distances can be investigated for inferring convergent evolution.

I have constructed multiparametric trees using Euclidian distances which are in a spherical space. However there exists another distance called Mahalanobis distance which is in an ellipsoidal space. One can investigate the construction of phylogenetic trees using Mahalanobis distance.
Finally, it may be possible to develop an algorithm to recognize from the available tree topology the absence of taxa or the Missing links that could be subsequently predicted to complete the consensus evolutionary tree.

If one were to classify organisms on the basis of their motility, Emu or Ostrich would have been classified along with bipedal vertebrates. However recognition and comparison of their other morphological traits places them squarely along with birds which are mostly winged flying animals. Thus among birds, Ostrich exhibits adaptive radiation. And there are many thousands of examples of such cases in the biological kingdom. It would be a task to generate a model that could integrate taxa exhibiting adaptive radiation in a consensus phylogenetic tree.

I have used PYTHON as the object-oriented high level programming language to develop multiparametric phylogenetic trees. While this language is easy to implement it is slower that low level languages like C/C++. Therefore, for a large number of parameters, translation and parallelization of these programs into C/C++ may be useful. Indeed, Python is known to be a language that serves as e-glue and can interface with most other object-oriented languages, so that the present attempts to construct multiparametric consensus phylogenetic trees can also be enriched using other quantitative parameters including genetic, morphological, geographical, environmental, physiological, etc.

Finally, I have enclosed a CD that shows an animated NJ tree in 3D space based on the present data on MSA of ATP8, ND and COX polypeptides. The visualization was achieved through python programming and is shown on Real Player.