Chapter 9

PHYLOGENETIC ANALYSIS: A BRIEF OVERVIEW

Kurtuluş ÖZGİŞİ[1]

◆ Basis of Phylogenetics

Phylogenetics is the science of estimating and analyzing evolutionary relationships of life on Earth. Initially, phylogenetic relationships of living organism were constructed using morphological data. The phenotype is the genetic information's result [1]. This genetic information pass on to the next generation and offspring that similar to parent occurs. Although phenotype is the results of proteins, RNA, DNA and their interacts with the environment, the function of some genetic materials, such as introns, are still not entirely clear. Because of the technological advances, molecular data are widely used to estimate phylogenetic relationships among organism [2].

In molecular systematic studies, proteins and DNA are the two molecules which have been mostly used. Since both of these molecules are heritable, they are assumed that they contain phylogenetic information in a linear array [3]. DNA has a double helix which have two antiparallel polynucleotide strands which consists of deoxyriboses with a phosphodiester linking each 5' carbon with the 3' carbon of the next sugar [Fig. 1.]. On each sugar, one of the four bases, adenine (A), or guanine (G) (the purines); thymine (T) or cytosine (C) (the pyrimidines), is linked to the 1' carbon. Hydrogen bonds between bases results in the two DNA strands binding together [4].

A nucleotide sequence of an organism is represented by four letters. Each tree bases a code for a specific amino acid [5]. During duplication, the DNA polymerase can occasionally include a non-complementary nucleotide. Also environmental factors, such as UV light, can cause damage on the bases. In such cases, DNA strand can be chemically modified [6]. When repair mechanism of genetic material miss these damages, the point mutation occurs. At the third codon position, point mutation rarely make a change in an amino acid whereas damage at first codon position mostly results in an amino acid change. Mutations that do not result in amino acid changes are called silent or synonymous mutations [7].

1   Dr., Eskisehir Osmangazi University, Faculty of Science and Letters, Department of Biology, Eskisehir, Turkey. kurtulusozgisi@gmail.com
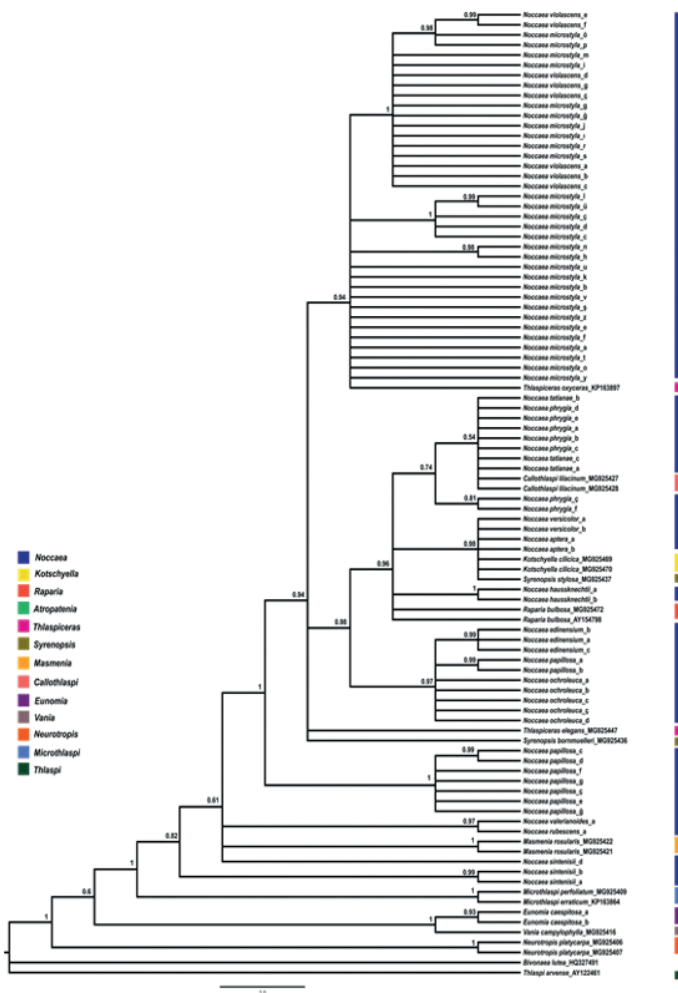
Fig. 7. A sample for phylogenetic tree based on Bayesian analysis. The posterior probabilities are indicated on the nods.

◆ **REFERENCES**

1. Sinden RR. 2012. DNA structure and function. Elsevier.
2. Montegut Felkner AE, Triemer RE. 1997. Phylogenetic Relationships of Selected Euglenoid Genera Based on Morphological and Molecular data 1. Journal of phycology, 33(3), 512-519.
3. Page RD, Holmes EC. 2009. Molecular evolution: a phylogenetic approach. John Wiley & Sons.
4. Borodovsky M, McIninch J. 1993. GENMARK: parallel gene recognition for both DNA strands. Computers & chemistry, 17(2), 123-133.
5. Johnson DB, Wang L. 2010. Imprints of the genetic code in the ribosome. Proceed-

ings of the National Academy of Sciences, 107(18), 8298-8303.

6. Freier SM, Altmann KH. 1997. The ups and downs of nucleic acid duplex stability: structure-stability studies on chemically-modified DNA: RNA duplexes. Nucleic acids research, 25(22), 4429-4443.

7. Chamary JV, Parmley JL., Hurst LD. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. Nature Reviews Genetics, 7(2), 98-108.

8. Freese EB. 1961. Transitions and transversions induced by depurinating agents. Proceedings of the National Academy of Sciences of the United States of America, 47(4), 540.

9. Allers T, Lichten M. 2001. Differential timing and control of noncrossover and crossover recombination during meiosis. Cell, 106(1), 47-57.

10. Foster PG, Hickey DA. 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. Journal of molecular evolution, 48(3), 284-290.

11. Baum DA, Donoghue MJ. 1995. Choosing among alternative" phylogenetic" species concepts. Systematic botany, 560-573.

12. Hillis DM. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Systematic Biology, 47(1), 3-8.

13. Snel B, Bork P, Huynen MA. 1999. Genome phylogeny based on gene content. Nature genetics, 21(1), 108-110.

14. Revell LJ, Harmon LJ, Collar DC. 2008. Phylogenetic signal, evolutionary process, and rate. Systematic biology, 57(4), 591-601.

15. Suarez DL, Perdue ML. 1998. Multiple alignment comparison of the non-structural genes of influenza A viruses. Virus research, 54(1), 59-69.

16. Rambaut A. 1996. Se-Al: sequence alignment editor. version 2.0. University of Oxford.

17. Hall T, Biosciences I, Carlsbad C. 2011. BioEdit: an important software for molecular biology. GERF Bull Biosci, 2(1), 60-61.

18. Li KB. 2003. ClustalW-MPI: ClustalW analysis using distributed and parallel computing. Bioinformatics, 19(12), 1585-1586.

19. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. 1998. Multiple sequence alignment with Clustal X. Trends in biochemical sciences, 23(10), 403-405.

20. Sievers F, Higgins DG. 2014. Clustal omega. Current protocols in bioinformatics, 48(1), 3-13.

21. Katoh K, Asimenos G, Toh H. 2009. Multiple alignment of DNA sequences with MAFFT. In Bioinformatics for DNA sequence analysis (pp. 39-64). Humana Press.

22. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32(5), 1792-1797.

23. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular biology and evolution, 33(7), 1870-1874.

24. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4(4), 406-425.

25. Strimmer K, Von Haeseler A. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Molecular Biology and Evolution, 13(7), 964-969.

26. Gehrke J, Ganti V, Ramakrishnan R, Loh WY. 1999. BOAT-optimistic decision tree construction. In Proceedings of the 1999 ACM SIGMOD international conference on

Management of data (pp. 169-180).

27. Sober E. 2002. Instrumentalism, parsimony, and the Akaike framework. Philosophy of Science, 69(S3), S112-S123.

28. Anderberg A, Tehler A. 1990. Consensus trees, a necessity in taxonomic practice. Cladistics, 6(4), 399-402.

29. Bergsten J. 2005. A review of long-branch attraction. Cladistics, 21(2), 163-193.

30. Guindon S, Delsuc F, Dufayard JF, Gascuel O. 2009. Estimating maximum likelihood phylogenies with PhyML. In Bioinformatics for DNA sequence analysis (pp. 113-137). Humana Press.

31. Lockhart PJ, Steel MA, Hendy MD, Penny D. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Molecular Biology and Evolution, 11(4), 605-612.

32. Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics (Oxford, England), 14(9), 817-818.

33. Nylander JAA. 2004. MrModeltest ver. 2. Evolutionary Biology Centre, Uppsala University, Sweden.

34. Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In Mammalian protein metabolism (H. Munro, ed.) (pp. 21-132). Academic Press, New York.

35. Felsenstein J. 1981. Continuous-genotype models and assortative mating. Theoretical Population Biology, 19(3), 341-357.

36. Kimura DK. 1980. Likelihood methods for the von Bertalanffy growth curve. Fishery bulletin, 77(4), 765-776.

37. Hasegawa M, Kishino H, Yano TA. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution, 22(2), 160-174.

38. Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. Journal of Molecular Evolution, 39(3), 306-314.

39. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), 1312-1313.

40. Huelsenbeck JP, Larget B, Swofford D. 2000. A compound Poisson process for relaxing the molecular clock. Genetics, 154(4), 1879-1892.

41. Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics, 17(8), 754-755.

42. Huelsenbeck JP, Larget B, Miller RE, Ronquist F. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. Systematic Biology, 51(5), 673-688.

43. Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution, 39(4), 783-791.