

A Review on Phylogenetic Analysis: A Journey through Modern Era

Sourav Singha Roy, Rakhi Dasgupta*, Angshuman Bagchi*

Department of Biochemistry and Biophysics, University of Kalyani, Kalyani, India
Email: [*rdgadg@gmail.com](mailto:rdgadg@gmail.com), [*angshuman_bagchi@yahoo.com](mailto:angshuman_bagchi@yahoo.com)

Received 15 August 2014; revised 5 September 2014; accepted 29 September 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Phylogenetic analysis may be considered to be a highly reliable and important bioinformatics tool. The importance of phylogenetic analysis lies in its simple manifestation and easy handling of data. The simple tree representation of the evolution makes the phylogenetic analysis easier to comprehend and represent as well. The varied applications of phylogenetics in different fields of biology make this analysis an absolute necessity. The different aspects of phylogenetic analysis have been described in a comprehensive manner. This review may be useful to those who would like to have a firsthand knowledge of phylogenetics.

Keywords

Phylogeny, Tree Structure, Evolutionary Time, Evolution of Function

1. Introduction

The basic definition of “evolution” can be given in versatile ways in different contexts. From the biologist point of view evolution can be defined as the development of a biological form from other preexisting forms or its origin to the current existing form through natural selections and modifications *i.e.* change across successive generation. The driving force behind evolution is natural selection in which “unfit” forms are eliminated through changes of environmental conditions or sexual selection so that only the fittest are selected (Darwinism). The underlying mechanism of evolution is genetic mutations that occur spontaneously. The mutations on the genetic material provide the biological diversity within a population; hence, the variability of individuals within the population to survive successfully in a given environment. Genetic diversity thus provides the source of raw material for the natural selection to act on.

The term “phylogenetics” derived from the Greek terms *phyle* and *phylon* means “tribe” and “race”; and the

*Corresponding authors.

term “genetikos” imply “relative to birth”, from “genesis” *i.e.* “birth”. Phylogenetics is the study of evolutionary relatedness among groups of organisms (e.g. species, populations). In other words, phylogenetic analysis of a family is to determine how the family might have been derived during evolution.

2. Representation of Phylogenetic Relationship: Phylogenetic Tree

Phylogenetic tree is a two dimensional representation of relatedness among various biological species. It is a line drawing that provides a visual means of representation for a group of sequences or species and indicates their time series of origin. The phylogenetic tree is represented in three forms: Phylogram, Dendrogram, Cladogram.

Merits and Demerits of Tree Building Methods

A phylogenetic tree may be built by mainly either distance based methods or character based methods.

The most commonly used distance based methods include UPGMA (unweighted paired group method with arithmetic mean) [1], NJ (neighbor joining) [2], ME (minimum evolution method) [3], and FM (Fitch-Margoliash method) [4].

Character based method derives trees that optimize the distribution of the actual data pattern for each character. The most commonly used character based method includes Maximum Parsimony (MP) method [5] and Maximum Likelihood (ML) method [6].

There are some important criteria such as computational speed, consistency of estimated topology, statistical consistency of phylogenetic trees, probability of obtaining the correct topology, reliability of estimated branch length, depending on which we can compare different established tree-building methods. The computational speed of each tree-building method depends on the algorithms that have been used in each case. According to this criterion (*i.e.*, computational speed), the NJ method is the superior one from other tree-building methods which are currently in use. This method can handle a large number of sequences with bootstrap tests with ease whereas MP, ME, and ML methods examine all possible topologies searching for the MP, ME and ML trees, respectively. We all know that the possible number of topologies increases sharply with number of input sequences and it becomes hard to use these methods when number of experimental sequences is high. We are hopping for simplified algorithms to be developed for these methods as well. In the case of ME, simplified advanced algorithms has been developed which is efficient in frame of timescale for obtaining the correct tree and also for MP methods the branch and bound method is often used when number of sequence is relatively high. During nineties algorithm suggested by Rzhetsky & Nei may be used for determining trees rapidly. If no bias is applied during the estimation of distance through substitution NJ, ME methods are found consistent for estimating trees but MP is often inconsistent. A tree-building method is considered as a “consistent estimator” if the method tends to give the correct topology as the number of experimental sequence tends to infinity. ML methods on the other hand have the additional advantage of being more flexible in choosing the evolutionary model. But this method is leangthy and time consuming.

3. Dimension of Evolution: Evolutionary Time

The time period taken for evolution of a group of protein or DNA from a common ancestor is called as evolutionary time. Number of changes occurring in evolution can be identified by these phylogenetic analysis methods. It estimates No. of changes *i.e.* No. of mutations in protein sequence. It is done by multiple sequence alignment as the first step. Therefore it is based upon distance scores/sequence similarity score.

3.1. Molecular Clock Hypothesis [7]-[9]

In 1960’s Zuckerkandl & Pauling proposed the molecular clock hypothesis, which changes the concepts in modern evolutionary biology, proposes that genes and gene products evolve at rates that are roughly constant over time and across evolutionary lineages. It gives the idea about time scales of natural events even in the absence of fossil evidence. Molecular clock hypothesis is defined as the nucleic acids and proteins evolves at rates that are constant over times, also this evolution relates to mutations that an organism uses to progress to next generation without loss of function and not lethal.

Molecular clock simply aims at finding the number of mutations in a given protein given the time it has taken to evolve since rates of evolution are constant *i.e.* all the mutations occurs in same rate in all the branches and

the rate of mutations are same for all the positions along the sequence. The protein that functions well in keeping up with a molecular clock is alpha globins although at the structural level this clock does not tick without variation.

Divergence of Molecular Clock Hypothesis

The difference in rate of molecular evolution among lineages is only one of the potential problems faced by the evolutionary biologist interested in using molecular clocks to date divergence events. All molecular clocks must be calibrated using independent evidence, such as dates of speciation events inferred from the fossil record or dates estimated for particular biogeographic events. Attempts to estimate divergence times are obviously simpler when the taxa in question share a similar rate of molecular evolution. However, in the real world researchers may often be faced with rate variation among lineages. There are a number of potential methods available to solve this problem. Many methods like linearized tree method [10] [11] and the quartet method [12] estimate the divergence times by removing the nonclock-like subsets of the data. These methods have been used in diverse-cases such as avian biogeography [13], molecular evolution [14], and mammalian [15] [16] diversification. Quartet method identifies the pairs of taxa that have good fossil data with which we can calibrate absolute rates of molecular evolution between the pair. These pairs can in turn be assembled into quartets consisting of two pairs of taxa, each of which has a known fossil date of divergence. The problem of envisioning non-clock-like data was solved by two methods [16] [17] which includes nonparametric rate smoothing (NPRS) and penalized likelihood. These two are distinct from the previous methods because, rather than throwing out non-clock-like data, aforementioned methods estimate local rates, *i.e.*, for specific branches or clades. This is possible because these methods use constraint during the calculation of rate of molecular evolution, which can vary among lineages.

4. World Revealing by Phylogenetic Analysis

4.1. Evolution of Function [18]-[22]

Selection of advantageous mutations by natural procedure *i.e.*, positive selection, is an exciting field for evolutionary biologists to work on, because adaptive changes in genes are eventually responsible for evolutionary modernism. So natural selection has become a powerful approach for molecular biologists, biochemists, and virologists to understand the functions of new genes. Some studies using phylogenetic approaches have identified a number of genes under positive selection, especially genes involved in host-pathogen interactions. In a recent issue of PNAS described a remarkable study in which phylogenetic sequence comparison identified a small segment of the primate TRIM5 α protein to be under positive selection, and functional analysis using mutagenesis confirmed the importance of the segment in species-specific retroviral inhibition.

So we can say that information about protein sequences of ancestral organisms is important for identifying critical amino acid substitutions that have caused the functional change of proteins in evolution.

4.2. Ancestral Sequences Prediction [23]-[25]

The prediction of ancestral protein sequences from multiple sequence alignments is useful for many bioinformatics analyses. Predicting ancestral sequences is not a simple procedure and it depends on accuracies of alignments and proper phylogenetic analysis. Several algorithms exist based on Maximum Parsimony or Maximum Likelihood methods but many current implementations are unable to process residues with gaps, which may represent insertion/deletion (INDEL) events or sequence fragments. Predicting ancestral protein sequences from a multiple sequence alignment is a useful tool in bioinformatics. Many evolutionary sequence analyses require such predictions in order to map substitutions to a particular lineage. In other situations, the predicted ancestral sequence alone may provide a more representative functional sequence than a simple consensus sequence constructed from an alignment. Strict consensus methods are quick but can suffer from overrepresentation of larger clades of related sequences, which contribute more sequences to the consensus than more sparsely populated clades. Maximum Parsimony (MP) method overcomes this problem by minimising mutational steps, rather than maximising agreement with the terminal sequences. MP, however, cannot distinguish between several equally parsimonious predictions. More sophisticated likelihood-based methods exist that can give probabilities for different ancestral sequences and implementation such as CODEML and FASTML provides good balance between speed and accuracy.

4.3. Amino Acid Sites under Positive Selection: Prediction of Adaptive Evolution [26]-[29]

Modern researcher of molecular evolutionary genomics shows their interest in the detection of positive selection on protein-coding DNA sequences. Nucleotide substitutions in the coding genes of amino acids of proteins can be either synonymous where amino acid changes or non-synonymous *i.e.*, silent substitutions where amino acid remains same. Usually, most non-synonymous changes would be expected to be eliminated by purifying selection, but under certain conditions Darwinian selection may allow their retention. Estimation of synonymous and non-synonymous substitution rates is important for revealing the dynamics of molecular evolution. In parsimony methods, substitutions are determined using parsimony reconstruction of ancestral sequences, and an excess of non-synonymous substitutions is tested independently for each site. The two methods differ in a way, first estimated the average ratio of non-synonymous rate (dN) to the synonymous rate (dS) *i.e.*, dN/dS along the sequence and then compared the non-synonymous/synonymous rate ratio at each site against this average. Likelihood method is a two-step procedure in which firstly “likelihood ratio test” is done for positive selection in the whole gene. If this test indicates statistical evidence for the presence of a proportion of sites evolving under positive selection, identification of putative positively selected sites can then proceed. The likelihood methods are used in the PAML package.

4.4. Simulation of Molecular Evolution [30]-[41]

What is the origin of life? A highly questionable field. In this context, computer simulation is played an important role. The idea was, there was once a prehistoric stage wherein RNA carried both the genetic function and the catalytic function, named “the RNA World”. However, still there was question to answer, how did the RNA World arise? A relatively direct and simple consideration is that, the RNA World originated *de novo* from non-living world, which involves several stages: stage 1, prebiotic synthesis of nucleotides; stage 2, prebiotic formation of poly-nucleotides from the nucleotides; stage 3, emergence of special RNA molecules catalyzing its own replication primordial “RNA replicases”; stage 4, evolution of the primordial replicases towards more efficient ones; stage 5, emergence and evolution of other catalytic RNA molecules favoring replication or existence in the background of natural selection. However, experimental evidence in this field still stays at level one of these stages *i.e.*, mineral-catalyzed synthesis of polynucleotides and non-enzymatic template-directed ligation of oligoribonucleotides or polymerization; RNA-catalyzed template-directed ligation or polymerization and recreating RNA replicases via *in vitro* directed molecular evolution; artificial construction of an autoevolving replicase system. Up to now, researchers seem to have outlined all the basic reaction mechanisms of these stages, but they were not sure if these stages could happen as a continuous and integrated process. This is a point where computer simulation provides the assistance. Monte Carlo simulation is a kind of computer simulation that mimicking random events in reality by determining the relative probabilities based on definitive rules. For instance, the scenario concerning the genesis of the widely accepted RNA World remains blurry, though we have gathered some circumstantial evidence and fragmented knowledge on several supposed stages, including formation of polynucleotides from a prebiotic nucleotide pool, emergence of RNA replicases (RNA molecules catalyzing their own replication), and evolution of RNA replicases. It is highly valuable to simulate the stages as a continuous process to evaluate the plausibility of the supposition and study the rules involved.

4.5. Modern Trends in Phylogenetics [42]-[57]

With third-generation sequencing technology rapidly approaching, it will become more feasible to obtain large multilocus data sets to infer evolutionary relationships (Genome 10 k Community of Scientists 2009). These enormous quantities of data have spawned the development of several new programs for phylogenetic inference for these highly heterogeneous data sets. From multiple sequence alignment (MSA) to species tree construction, these new methods are changing the way we gather and manipulate data and analyze and interpret results. Following the construction of an MSA for the traditional 2-step MSA phylogeny estimation procedure, the researcher is left with the decision of how to handle the gaps inserted into the data set by the MSA algorithm to account for INDEL events. For most traditional maximum parsimony (MP) analyses, gaps have been either coded as missing data (most cases) or coded as a fifth character state. Both of these methods are potentially problematic in that the former completely discards relevant evolutionary information, whereas the latter assumes that gaps represent independent evolutionary events; a highly unlikely scenario. These issues also extended into

probabilistic phylogenetic inference in that parameters were estimated without taking indel events into account. An alternative to constructing an MSA prior to phylogenetic inference is to use DO (direct optimization) procedures. DO is different from other approaches in that the alignment and phylogenetic tree are estimated simultaneously. Optimization can be performed either under parsimony or under a probabilistic framework. The program POY, for example, estimates both the phylogenetic tree and the best alignment based on the MP criterion. Previous versions of POY were also able to implement DO in a likelihood framework. Newer programs such as Stat-Align, BALi-Phy, and BEAST incorporate models of sequence evolution to estimate the posterior distribution of a set of trees and alignments based on Bayesian inference (BI). The Bali-Phy software shows exceptional promise in that its models allow for nested or overlapping indel events, whereas other methods utilize the more common TKF1 and TKF2 indel models. However, joint estimation of alignment and phylogeny in a probabilistic framework is currently computationally intensive and feasible only with smaller data sets. These methods also fit a single model to the data, which may not be justified with multi-locus data sets. As multi-locus data sets become the norm across laboratories, some of the most commonly employed techniques for both MSA and tree reconstruction will no longer be adequate for generating phylogenetic hypotheses. Instead, alternate and more sophisticated search algorithms are required in order to fully exploit the information contained in these large quantities of data. As highly heterogeneous data sets become available, testing the accuracy of both modern alignment algorithms and DO methods through simulation will become even more important. For traditional phylogenetic inference, MP analysis will no doubt continue to play a role. In this regard, TNT (Tree Analysis Using New Technology) is showing promise for dealing with difficult phylogenetic problems. Furthermore, model-based concatenation methods using mixture models in Bayes Phylogenies seem promising for multi-locus data sets. However, there have been few simulations to quantify the accuracy of the model compared with other methods including direct species tree inference.

Acknowledgements

Authors are thankful to the Department of Science and Technology (DST Project No. SR/SO/BB-71/2007), Govt. of India, and UGC (RFSMS) for their financial support. The authors also acknowledge the DBT for Bioinformatics Infrastructure Facility (BIF) of University of Kalyani, Mr. Suman Nandy of Department of Biochemistry and Biophysics and Mr. Rajabata Bhuyan of BIF, University of Kalyani for their immense help.

References

- [1] Murtagh, F. (1984) Complexities of Hierarchic Clustering Algorithms: The State of the Art. *Computational Statistics Quarterly*, **1**, 101-113.
- [2] Saitou, N. and Nei, M. (1987) The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, **4**, 406-425.
- [3] Rzhetsky, A. and Nei, M. (1993) Theoretical Foundation of the Minimum-Evolution Method of Phylogenetic Inference. *Molecular Biology and Evolution*, **10**, 1073-1095.
- [4] Fitch, W.M. and Margoliash, E. (1967) Construction of Phylogenetic Trees. *Science*, **155**, 279-284. <http://dx.doi.org/10.1126/science.155.3760.279>
- [5] Sober, E. (1983) Parsimony in Systematics: Philosophical Issues. *Annual Review of Ecology and Systematics*, **14**, 335-357. <http://dx.doi.org/10.1146/annurev.es.14.110183.002003>
- [6] Felsenstein, J. (1981) Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach. *Journal of Molecular Evolution*, **17**, 368-376. <http://dx.doi.org/10.1007/BF01734359>
- [7] Zuekerkandl, E. and Pauling, L. (1962) Molecular Disease, Evolution, and Genie Heterogeneity. In: Kasha, M. and Pullman, B., Eds., *Horizons in Biochemistry*, Academic Press, New York, 189.
- [8] Zuekerkandl, E. and Pauling, L. (1964) Molecules as Documents of Evolutionary History. *Problems of Evolutionary and Technical Biochemistry*, Science Press, Academy of Sciences of the USSR, 54-62. (In Russian)
- [9] Zuekerkandl, E. and Pauling, L. (1965) Evolutionary Divergence and Convergence in Proteins. In: Bryson, V. and Vogel, H.J., Eds., *Evolving Genes and Proteins*, Academic Press, New York, 97-165. <http://dx.doi.org/10.1016/B978-1-4832-2734-4.50017-6>
- [10] Takezaki, N., Rzhetsky, A. and Nei, M. (1995) Phylogenetic Test of the Molecular Clock and Linearized Trees. *Molecular Biology and Evolution*, **12**, 823-833.
- [11] Cooper, A. and Penny, D. (1997) Mass Survival of Birds across the Cretaceous-Tertiary Boundary: Molecular Evi-

- dence. *Science*, **275**, 1109-1113. <http://dx.doi.org/10.1126/science.275.5303.1109>
- [12] Voelker, G. (1999) Dispersal, Vicariance, and Clocks: Historical Biogeography and Speciation in a Cosmopolitan Passerine Genus (Anthus: Motacillidae). *Evolution*, **53**, 1536-1552. <http://dx.doi.org/10.2307/2640899>
- [13] Edwards, S.V., Chesnut, K., Satta, Y. and Wakeland, E.K. (1997) Ancestral Polymorphism of Mhc Class II Genes in Mice: Implications for Balancing Selection and the Mammalian Molecular Clock. *Genetics*, **146**, 655-668.
- [14] Hedges, S.B., Parker, P.H., Sibley, C.G. and Kumar, S. (1996) Continental Breakup and the Ordinal Diversification of Birds and Mammals. *Nature*, **381**, 226-229. <http://dx.doi.org/10.1038/381226a0>
- [15] Wang, D., Kumar, S. and Hedges, S. (1999) Divergence Time Estimates for the Early History of Animal Phyla and the Origin of Plants, Animals and Fungi. *Proceedings of the Royal Society B*, **266**, 163-171. <http://dx.doi.org/10.1098/rspb.1999.0617>
- [16] Bishop, J.G., Dean, A.M. and Mitchell-Olds, T. (2000) Rapid Evolution in Plant Chitinases: Molecular Targets of Selection in Plant-Pathogen Coevolution. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 5322-5327. <http://dx.doi.org/10.1073/pnas.97.10.5322>
- [17] Haydon, D.T., Bastos, A.D., Knowles, N.J. and Samuel, A.R. (2001) Evidence for Positive Selection in Foot-and-Mouth Disease Virus Capsid Genes from Field Isolates. *Genetics*, **157**, 7-15.
- [18] Shackelton, L.A., Parrish, C.R., Truyen, U. and Holmes, E.C. (2005) High Rate of Viral Evolution Associated with the Emergence of Carnivore Parvovirus. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 379-384. <http://dx.doi.org/10.1073/pnas.0406765102>
- [19] Polley, S.D. and Conway, D.J. (2001) Strong Diversifying Selection on Domains of the Plasmodium Falciparum Apical Membrane Antigen 1 Gene. *Genetics*, **158**, 1505-1512.
- [20] Götesson, A., Marshall, J.S., Jones, D.A. and Hardham, A.R. (2002) Characterization and Evolutionary Analysis of a Large Polygalacturonase Gene Family in the Oomycete Plant Pathogen *Phytophthora cinnamomi*. *Molecular Plant-Microbe Interactions*, **15**, 907-921. <http://dx.doi.org/10.1094/MPMI.2002.15.9.907>
- [21] Sawyer, S.L., Wu, L.I., Emerman, M. and Malik, H.S. (2005) Positive Selection of Primate TRIM5a Identifies a Critical Species-Specific Retroviral Restriction Domain. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 2832-2837. <http://dx.doi.org/10.1073/pnas.0409853102>
- [22] Stremlau, M., Owens, C.M., Perron, M.J., Kiessling, M., Autissier, P. and Sodroski, J. (2004) The Cytoplasmic Body Component TRIM5a Restricts HIV-1 Infection in Old World Monkeys. *Nature*, **427**, 848-853. <http://dx.doi.org/10.1038/nature02343>
- [23] Zhang, J. and Nei, M. (1997) Accuracies of Ancestral Amino Acid Sequences Inferred by the Parsimony, Likelihood, and Distance Methods. *Journal of Molecular Evolution*, **44**, S139-S146. <http://dx.doi.org/10.1007/PL00000067>
- [24] Fitch, W.M. (1971) Toward Defining Course of Evolution: Minimum Change for a Specific Tree Topology. *Systematic Zoology*, **20**, 406-416. <http://dx.doi.org/10.2307/2412116>
- [25] Yang, Z., Kumar, S. and Nei, M. (1995) A New Method of Inference of Ancestral Nucleotide and Amino Acid Sequences. *Genetics*, **141**, 1641-1650.
- [26] Kimura, M. (1980) A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide Sequences. *Journal of Molecular Evolution*, **16**, 111-120. <http://dx.doi.org/10.1007/BF01731581>
- [27] Suzuki, Y. and Gojobori, T. (1999) A Method for Detecting Positive Selection at Single Amino Acid Sites. *Molecular Biology and Evolution*, **16**, 1315-1328. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a026042>
- [28] Nielsen, R. and Yang, Z. (1998) Likelihood Models for Detecting Positively Selected Amino Acid Sites and Applications to the HIV-1 Envelope Gene. *Genetics*, **148**, 929-936.
- [29] Yang, Z. (2000) Maximum Likelihood Estimation on Large Phylogenies and Analysis of Adaptive Evolution in Human Influenza Virus A. *Journal of Molecular Evolution*, **51**, 423-432.
- [30] Ferris, J.P., Hill, A.R., Liu, R. and Orgel, L.E. (1996) Synthesis of Long Prebiotic Oligomers on Mineral Surfaces. *Nature*, **381**, 59-61. <http://dx.doi.org/10.1038/381059a0>
- [31] Ferris, J.P. (2002) Montmorillonite Catalysis of 30-50 Mer Oligonucleotides: Laboratory Demonstration of Potential Steps in the Origin of the RNA World. *Origins of Life and Evolution of the Biosphere*, **32**, 311-332. <http://dx.doi.org/10.1023/A:1020543312109>
- [32] Rohatgi, R., Bartel, D.P. and Szostak, J.W. (1996) Kinetic and Mechanistic Analysis of Nonenzymatic, Template-Directed Oligoribonucleotide Ligation. *Journal of the American Chemical Society*, **118**, 3332-3339. <http://dx.doi.org/10.1021/ja953712b>
- [33] Orgel, L.E. (1992) Molecular Replication. *Nature*, **358**, 203-209. <http://dx.doi.org/10.1038/358203a0>
- [34] Jaeger, L., Wright, M.C. and Joyce, G.F. (1999) A Complex Ligase Ribozyme Evolved *In Vitro* from a Group I Ribozyme.

- zyme Domain. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 14712-14717. <http://dx.doi.org/10.1073/pnas.96.26.14712>
- [35] Kozlov, I.A. and Orgel, L.E. (2000) Nonenzymatic Template-Directed Synthesis of RNA from Monomers. *Molecular Biology*, **34**, 781-789. <http://dx.doi.org/10.1023/A:1026663422976>
- [36] McGinness, K.E. and Joyce, G.F. (2002) RNA-Catalyzed RNA Ligation on an External RNA Template. *Chemistry & Biology*, **9**, 297-307. [http://dx.doi.org/10.1016/S1074-5521\(02\)00110-2](http://dx.doi.org/10.1016/S1074-5521(02)00110-2)
- [37] McGinness, K.E. and Joyce, G.F. (2003) In Search of an RNA Replicase Ribozyme. *Chemistry & Biology*, **10**, 5-14. [http://dx.doi.org/10.1016/S1074-5521\(03\)00003-6](http://dx.doi.org/10.1016/S1074-5521(03)00003-6)
- [38] Lawrence, M.S. and Bartel, D.P. (2003) Processivity of Ribozyme-Catalyzed RNA Polymerization. *Biochemistry*, **42**, 8748-8755. <http://dx.doi.org/10.1021/bi0342281>
- [39] Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. and Bartel, D.P. (2001) RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension. *Science*, **292**, 1319-1325. <http://dx.doi.org/10.1126/science.1060786>
- [40] Bartel, D.P. (1999) Re-Creating an RNA Replicase. In: Gesteland, R.F., Cech, T.R. and Atkins, J.F., Eds., *The RNA World*, Cold Spring Harbor Laboratory Press, New York, 143-162.
- [41] Szostak, J.W., Bartel, D.P. and Luisi, P.L. (2001) Synthesizing Life. *Nature*, **409**, 387-390. <http://dx.doi.org/10.1038/35053176>
- [42] Genome 10k Community of Scientists (2009) Genome 10k: A Proposal to Obtain Whole-Genome Sequence for 10000 Vertebrate Species. *Journal of Heredity*, **100**, 659-674. <http://dx.doi.org/10.1093/jhered/esp086>
- [43] Swofford, D.L., Olsen, G.J., Waddell, P.J. and Hillis, D.M. (1996) Phylogenetic Inference. In: Hillis, D.M., Moritz, C. and Mable, B.K., Eds., *Molecular Systematics*, 2nd Edition, Sinauer Associates, Sunderland (MA), 407-514.
- [44] Simmons, M.P. and Ochoterena, H. (2000) Gaps as Characters in Sequence-Based Phylogenetic Analysis. *Systematic Biology*, **49**, 369-381. <http://dx.doi.org/10.1093/sysbio/49.2.369>
- [45] Simmons, M.P., Ochoterena, H. and Carr, T.G. (2001) Incorporation, Relative Homoplasy, and Effect of Gap Characters in Sequence-Based Phylogenetic Analyses. *Systematic Biology*, **50**, 454-462. <http://dx.doi.org/10.1080/106351501300318049>
- [46] Kawakita, A., Sota, T., Ascher, J.S., Ito, M., Tanaka, H. and Kato, M. (2003) Evolution and Phylogenetic Utility of Alignment Gaps within Intron Sequences of Three Nuclear Genes in Bumble Bees (*Bombus*). *Molecular Biology and Evolution*, **20**, 87-92. <http://dx.doi.org/10.1093/molbev/msg007>
- [47] Wheeler, W. (1996) Optimization Alignment: The End of Multiple Sequence Alignment in Phylogenetics. *Cladistics*, **12**, 1-9. <http://dx.doi.org/10.1111/j.1096-0031.1996.tb00189.x>
- [48] Varón, A., Vinh, L.S. and Wheeler, W.C. (2009) POY Version 4: Phylogenetic Analysis Using Dynamic Homologies. *Cladistics*, **26**, 72-85. <http://dx.doi.org/10.1111/j.1096-0031.2009.00282.x>
- [49] Novák, A., Miklós, I., Lyngso, R. and Hein, J. (2008) StatAlign: An Extendable Software Package for Joint Bayesian Estimation of Alignments and Evolutionary Trees. *Bioinformatics*, **24**, 2403-2404. <http://dx.doi.org/10.1093/bioinformatics/btn457>
- [50] Suchard, M.A. and Redelings, B.D. (2006) Bali-Phy: Simultaneous Bayesian Inference of Alignment and Phylogeny. *Bioinformatics*, **22**, 2047-2048. <http://dx.doi.org/10.1093/bioinformatics/btl175>
- [51] Lunter, G., Miklós, I., Drummond, A., Jensen, J.L. and Hein, J. (2005) Bayesian Coestimation of Phylogeny and Sequence Alignment. *BMC Bioinformatics*, **6**, 83. <http://dx.doi.org/10.1186/1471-2105-6-83>
- [52] Redelings, B.D. and Suchard, M.A. (2005) Joint Bayesian Estimation of Alignment and Phylogeny. *Systematic Biology*, **54**, 401-418. <http://dx.doi.org/10.1080/10635150590947041>
- [53] Redelings, B.D. and Suchard, M.A. (2007) Incorporating Indel Information into Phylogeny Estimation for Rapidly Emerging Pathogens. *BMC Evolutionary Biology*, **7**, 40. <http://dx.doi.org/10.1186/1471-2148-7-40>
- [54] Thorne, J.L., Kishino, H. and Felsenstein, J. (1991) An Evolutionary Model for Maximum Likelihood Alignment of DNA Sequences. *Journal of Molecular Evolution*, **33**, 114-124. <http://dx.doi.org/10.1007/BF02193625>
- [55] Thorne, J.L., Kishino, H. and Felsenstein, J. (1992) Inching towards Reality: An Improved Likelihood Model of Sequence Evolution. *Journal of Molecular Evolution*, **34**, 3-16. <http://dx.doi.org/10.1007/BF00163848>
- [56] Goloboff, P.A., Farris, J. and Nixon, K. (2008) TNT, a Free Program for Phylogenetic Analysis. *Cladistics*, **24**, 774-786. <http://dx.doi.org/10.1111/j.1096-0031.2008.00217.x>
- [57] Goloboff, P.A., Farris, J. and Nixon, K. (2003) TNT. Tree Analysis Using New Technology. Program and Documentation. <http://www.zmuc.dk/public/phylogeny/tnt>

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or [Online Submission Portal](#).

