Belg. J. Zool. - Volume 129 (1999) - issue 1 - pages 5-16 - Brussels 1999

MOLECULAR EVOLUTION AND THE INCORPORATION OF SITE-TO-SITE RATE VARIATION IN DISTANCE TREE CONSTRUCTION METHODS

YVES VAN DE PEER (1)
Department of Biochemistry, University of Antwerp (UIA),
Universiteitsplein 1, B-2610 Antwerpen, Belgium,
e-mail : yvdp@uia.ua.ac.be

Abstract. The construction of evolutionary trees based on sequence data is not self-evident. Apart from the plethora of methods and software tools to choose from if one wants to infer phylogenetic tree topologies, one has also to be cautious about the sequence data themselves. In this paper, we discuss how systematic errors can be introduced by one of the phenomena that often characterize sequence data, i.e. differences in substitution rates among the different sites of the molecule. Regarding pairwise distance methods, these systematic errors can often be avoided if an appropriate substitution model is applied to the construction of phylogenetic trees. This is demonstrated for a phylogeny based on animal small subunit ribosomal RNA sequences.

Key words : molecular evolution – phylogenetic trees – substitution models – among-site rate variation.

INTRODUCTION

Radical advances in biotechnology and sequencing techniques have led to an explosive growth in the accumulation of biological data. Specifically, new applications of the Polymerase Chain Reaction method and the use of automatic sequencing quickly provide us with a huge amount of sequence data that can be used to study the evolutionary history of organisms, genes, and gene families. However, parallel to the development and improvement of rapid sequencing techniques, advances in computer technologies in particular have been responsible for breakthroughs in our exploration of molecular evolution. Since there is an ever-greater need for faster hardware and new computational tools in order to cope with the exponential growth of sequence data, these advances were and are extremely important for molecular phylogeny to become established. Additionally, modern networks and network facilities such as the World Wide Web (WWW) have made life much easier for molecular biologists and evolutionists. Through the WWW, a wealth of information becomes available with a few mouse clicks or keystrokes : sequence databases

(1) Present address : Department of Biology, University of Konstanz, D-78457 Konstanz, Germany.

can be browsed, servers with biological information can be consulted on-line, data can be exchanged rapidly, and software can easily be downloaded and installed.

Because of these advances in computer hardware and software, one should expect the construction of phylogenetic trees – which is highly computer dependent – to have become quite easy. This is only partly true. Although fast and user-friendly programs are now available, the number of different methods for tree topology inference has increased rapidly during the last few years. As a result, people are often bewildered by the vast range of computer algorithms that can be applied to sequence data. Moreover, literature on the construction of phylogenetic trees is extensive and the pros and cons of different methods are frequently debated. On the other hand, the powers and pitfalls of the different algorithms and methods are becoming more and more understood. Lately, much effort goes into the study of specific substitution models that try to explain the evolutionary change of the molecules used for tree construction. If the "true" evolutionary process could be described accurately by a certain mathematical model, trees inferred on the basis of that model would suffer less from systematic errors. One of the recently well-studied phenomena that often cause systematic errors in a tree topology, is site-to-site rate variation in molecules. In this paper, we try to show and explain how differences in substitution rates among sites in the small subunit ribosomal RNA (SSU rRNA) can influence tree topologies that are inferred on the basis of evolutionary distances. For a more general discussion about tree construction methods and models of evolution, we refer to some of the nice reviews that have been written about phylogenetic inference (NEI, 1987; FELSENSTEIN, 1988; SWOFFORD *et al.*, 1996).

PAIRWISE DISTANCE TREES AND SITE-TO-SITE RATE VARIATION

Distance methods for tree construction first fit a tree to a matrix of pairwise evolutionary distances. For every two sequences, the distance is a single value based on the fraction of positions in which both sequences differ, defined as dissimilarity. This dissimilarity is actually an underestimation of the true evolutionary distance, because some of the individual sequence differences are the result of multiple events. Since mutations are fixed in the genes, there is an increasing chance over time of multiple mutations having occurred at the same sequence position. As a result, later mutations can hide previous ones. Therefore, in distance-based methods, the actual number of substitutions is usually estimated by applying a specific evolutionary model that makes assumptions about the nature of evolutionary changes. When all the pairwise distances have been computed for a set of sequences, a tree topology can then be inferred by a variety of methods, the most well-known of which is probably the neighbor-joining method (SAITOU & NEI, 1987).

Correct estimation of the evolutionary distance is crucial and several studies have shown that the use of an unrealistic substitution model can cause serious artifacts in tree topologies (OLSEN, 1987; VAN DE PEER *et al.*, 1993; LOCKHART *et al.*, 1994; DE RIJK *et al.*, 1995; VAN DE PEER *et al.*, 1996a, b). However, since we do not have an exact historical record of the events that took place in the evolution of sequences, it is not obvious how to correctly estimate the evolutionary distance. One of the first models used in the estimation of evolutionary distances is the one of JUKES & CANTOR (1969). This model starts from the

assumptions that all substitutions are independent, that all sequence positions are equally subject to change, that substitutions occur randomly among the four types of nucleotides – in other words, there is no bias in the direction of change –, and that no insertions or deletions have occurred. Based on these pre-assumptions, the authors derived an equation for estimating evolutionary distances from observed dissimilarity (see also Fig. 1):

$$d_{AB} = -\frac{3}{4} \ln \left(1 - \frac{4}{3} f_{AB}\right)$$

(Equation 1)

where f_{AB} is the dissimilarity (fraction of observed differences) between sequences A and B, and d_{AB} is the estimated evolutionary distance (fraction of expected differences) between sequences A and B.

Several other equations were subsequently proposed for the estimation of evolutionary distances. For example, KIMURA (1980) provided a method based on a model of evolution in which transitions and transversions may occur at different rates. Other equations are based on substitution models in which the four different nucleotides are not used in equal proportions (TAJIMA & NEI, 1984), or where a bias in the direction of change is accounted for (TAMURA & NEI, 1993 ; ZHARKIKH, 1994).

An important drawback of these models is that they ignore differences in substitution rate among the sites of a molecule. However, it has been known for some time that substitution rates differ among sites is nearly all genes. For example, OLSEN (1987) demonstrated that application of the JUKES & CANTOR model to sequences composed of sites with unequal rates could lead to artifacts in tree topology. He proposed a different evolutionary model that assumed a log-normal distribution of substitution rates over the sequence positions. JIN & NEI (1990) followed a similar approach but assumed that substitution rates were gamma distributed. On the basis of this distribution they derived several equations to compute evolutionary distances from observed sequence dissimilarities, using a parameter α that describes the extent of the rate variation. The main problem with applying gamma distances is the estimation of this parameter α, the accuracy of which depends on the estimation method, predefined tree topology, and on the number of sequences used in the estimation (SULLIVAN *et al.*, 1996 ; YANG, 1996).

Quantitative estimation of the substitution rates or variabilities of nucleotide sites is not obvious. Maximum parsimony estimates can be heavily biased while maximum likelihood estimates may experience computational difficulties (YANG, 1996). Recently, we developed a new method called "substitution rate calibration" for measuring the relative substitution rate of individual sites in a nucleotide sequence alignment on the basis of a distance approach (VAN DE PEER *et al.*, 1993 ; 1996a). The main advantage of this approach is that nucleotide variability estimates are independent from a given tree topology, contrary to estimates inferred from maximum parsimony or maximum likelihood methods (SULLIVAN *et al.*, 1996), and that they can be based on a large number of sequences (VAN DE PEER *et al.*, 1996c ; VAN DE PEER & DE WACHTER, 1997a). The latter point is particularly important since the accuracy of the substitution rate estimate increases with the num-

ber of sequences taken into account. When the substitution rates of the individual nucleotides of the molecule are computed as described previously (VAN DE PEER *et al.*, 1993; VAN DE PEER *et al.*, 1996a), an equation can be derived that describes a more realistic substitution model, and that discriminates more selectively between sequence dissimilarity and evolutionary distance (see also Fig. 1):

$$d_{AB} = p\left[\left(1 - \frac{4}{3} f_{AB}\right)^{-\frac{3}{4}p} - 1\right]$$

(Equation 2)

This equation is similar to the general formula proposed by RZHETSKY & NEI (1994) (with parameter $p = \frac{3}{4}\alpha$) to compute gamma distances.

APPLICATION OF SUBSTITUTION RATE CALIBRATION TO TREE CONSTRUCTION

When substitution rate calibration is applied to tree construction, sequence dissimilarity is converted into evolutionary distance, but for a set of nucleotides mutating with variable rates, the conversion allows for a slower increase in dissimilarity as a function of distance than for a randomly mutating set (see Fig. c). This is achieved using equation 2

Legend to the figure (see opposite page)

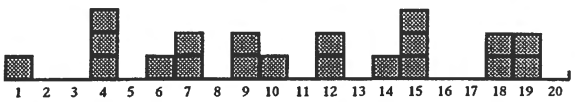
Fig. 1. – A. Hypothetical distribution of substitutions in a sequence of 20 nucleotides. It is assumed that the rate of substitution per site is the same for all sites in the sequence. In other words, substitutions (represented by gray squares) occur randomly. In this particular example, 11 substitutions are observed although 20 have really occurred. Several sites have undergone multiple substitutions; *e.g.* site 4 has mutated 3 times. When the dissimilarity (11/20) is converted into evolutionary distance by using the equation of Jukes & Cantor (1969; equation 1, see text) a value of about 1 is obtained. This means that, on average, every site has been substituted once (20 substitutions in a sequence of 20 nucleotides) which is indeed correct.

B. Hypothetical distribution of substitutions in a sequence of 20 nucleotides but here it is assumed that the rate of substitution in site 20 is 100 times that of site 1 (thus assuming a ratio of 1/100). As a result, the majority of substitutions will take place near the end of the sequence. Consequently, the number of observed substitutions will be smaller than in sequences where substitutions occur randomly (see a). In this particular example, the number of observed substitutions is only 8. If the distance is computed according to JUKES & CANTOR (equation 1, see text), the evolutionary distance is about 0.57. Therefore, the evolutionary distance is seriously underestimated, since the "true" evolutionary distance should be close to 1.

C. Graphic representation of the functions describing the relationship between evolutionary distance (expected fraction of substitutions) and dissimilarity (observed fraction of substitutions) when substitutions are assumed to occur randomly (upper curve; inverse of equation 1, see text), and when substitution rates are assumed to differ among sites (lower curve, inverse of equation 2, see text). See text for details.

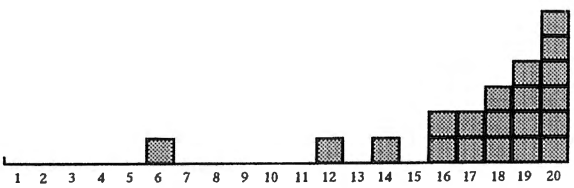
with parameter p adapted to the shape of a rate spectrum constructed by grouping alignment positions of similar variability (see VAN DE PEER *et al.*, 1993 ; 1996c). Using the latter approach yielded some significant improvements in tree topology for the evolution of eukaryotic SSU rRNA sequences of different groups of protists (VAN DE PEER *et al.*, 1996a, 1996b; VAN DE PEER & DE WACHTER, 1997a). In particular, the rate calibration method largely avoids tree distortions due to the presence of sequences with increased evolutio-

A) 20 sites with identical substitution rates



Substitutions : occurred	**20**	Distance (JC) =	**≈1**
observed	**11**	Dissimilarity =	**0.55**

B) 20 sites with different substitution rates (ratio 1 : 100)



Substitutions : occurred	**20**	Distance (JC) =	**0.57**
		Distance (cal) =	**≈1**
observed	**8**	Dissimilarity =	**0.4**

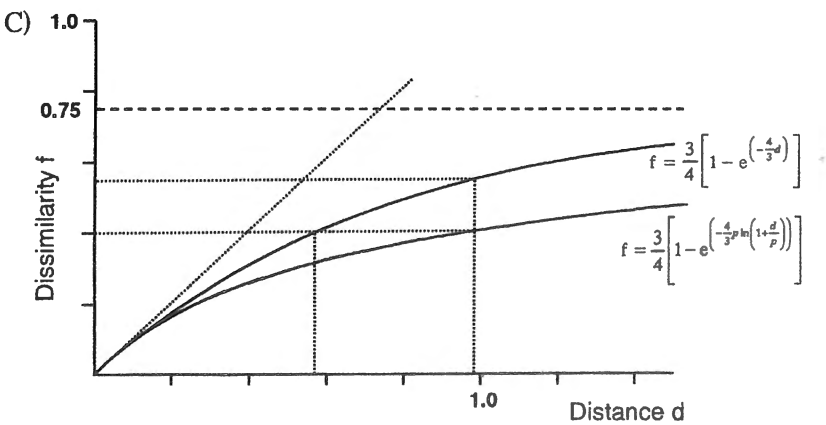


Fig. 1

nary rates. These long-branch distortions are usually caused by the underestimation of large distances with respect to smaller ones if distances are computed assuming equal variability of all nucleotides in a sequence (see Fig. 1c).

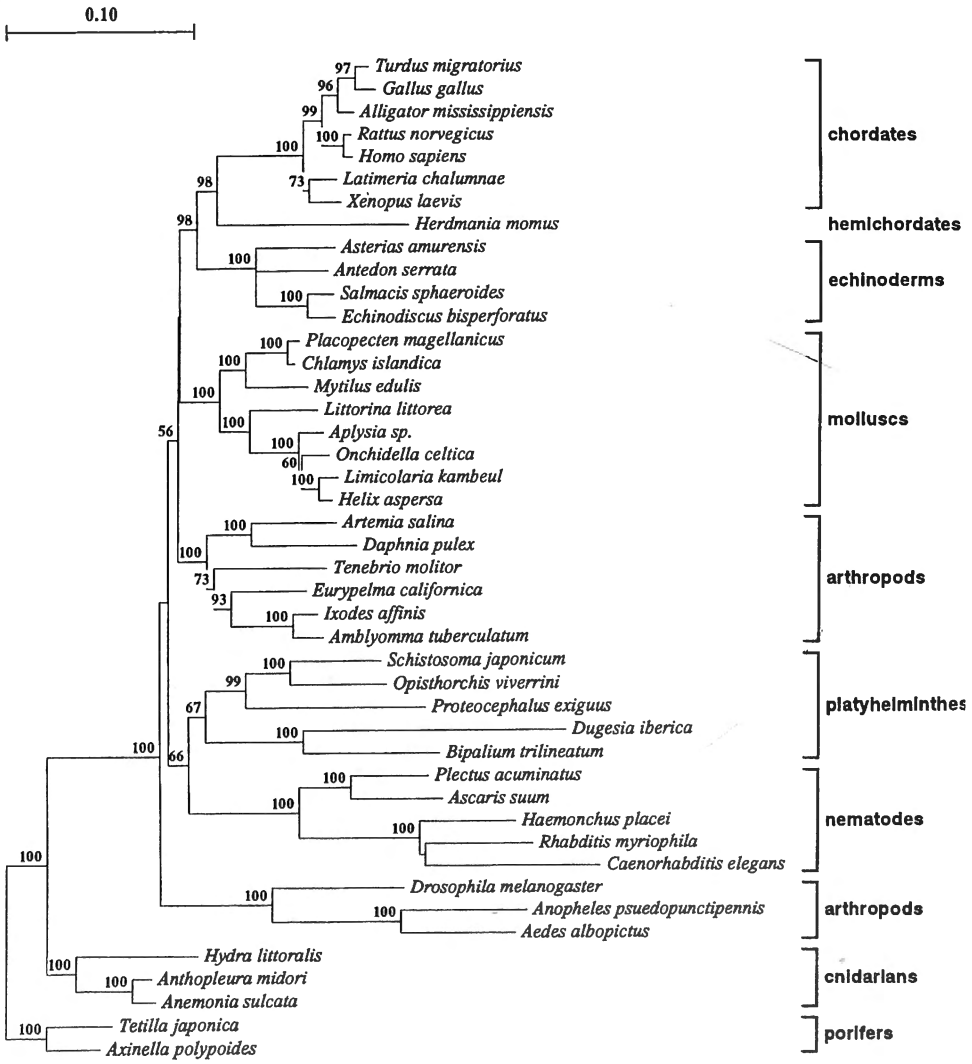


Fig. 2. – Evolutionary tree of 44 SSU rRNA sequences of animals retrieved from the Antwerp SSU rRNA database (VAN DE PEER *et al.*, 1998). Evolutionary distances were computed according to JUKES & CANTOR (1969) while the tree topology was inferred by neighbor-joining (SAITOU & NEI, 1987). The tree was rooted with the porifers. Bootstrap values (FELSENSTEIN, 1985) above 50 % (out of 500 replications) are indicated. The scale on top measures evolutionary distance in substitutions per nucleotide. Taxon designations are placed to the right of the corresponding clusters. All analyses were performed with the software package TREECON for Windows (VAN DE PEER & DE WACHTER, 1997b).

In the following example, rate calibration was applied to a phylogeny of Metazoa on the basis of SSU rRNA. Fig. 2. shows an evolutionary tree of 44 SSU rRNA sequences derived from different animals. The tree was constructed by neighbor-joining (SAITOU & NEI, 1987) on the basis of JUKES & CANTOR (1969) distances. Fig. 3 shows an evolutionary tree of the same set of sequences but with differences in substitution rates among the various sites of the SSU rRNA taken into account. Evolutionary distances were computed as described previously (VAN DE PEER *et al.*, 1993; 1996a) and a "p" value of 0.44 was obtained (see equation 2). Although the same groups of animals can be found in both trees, there are some remarkable differences, the most important of which is the position of the

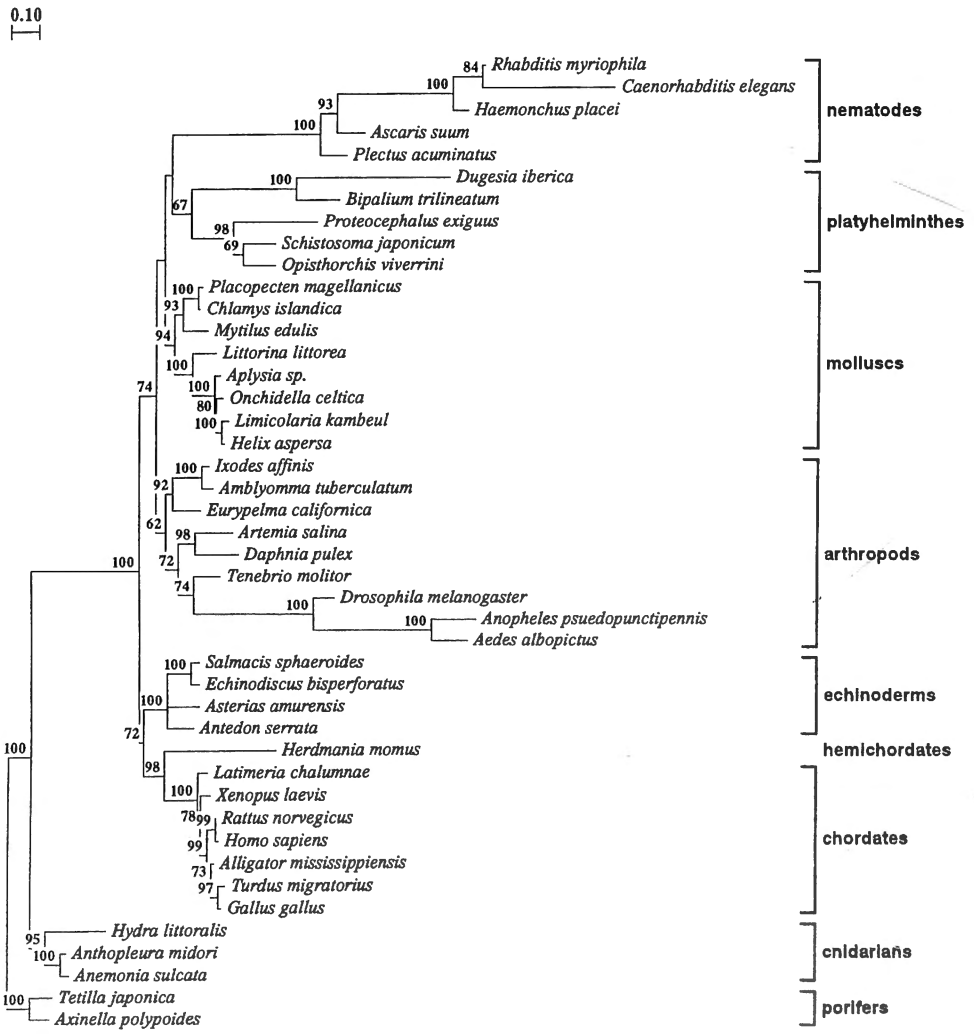


Fig. 3. – Evolutionary tree of the same set of animals as in Fig. 2., but based on "substitution rate calibration". Interpretation is as in Fig. 2. See text for details.

insects *Drosophila*, *Anopheles*, and *Aedes*. As can be seen in the calibrated tree (Fig. 3), these three SSU rRNA sequences are characterized by an increased evolutionary rate. As a result they form long branches in the tree with respect to most other sequences. Nevertheless, when rate calibration is applied, they are clustered as expected: with the other arthropods and more specifically with the other insect *Tenebrio*. Contrarily, in the tree based on JUKES & CANTOR distances, *Drosophila*, *Anopheles*, and *Aedes* seem to form an independent evolutionary lineage and diverge near the base of the tree, which is often characteristic for fast-evolving lineages.

As can be seen in Figs 2 and 3, the phylogenetic relationships and divergence order between the different animal phyla are hard to resolve on the basis of SSU rRNA data (see also *e.g.* ADDOUTE & PHILIPPE, 1993; MACKEY *et al.*, 1996; PAWLOWSKI *et al.* 1996; WINNEPENNINCKX *et al.*, 1996; ABOUHEIF *et al.*, 1998). This is probably due to a massive radiation of new evolutionary lineages within a small time interval during the Cambrium (ERWIN, 1991; ADOUTTE & PHILIPPE; 1993; PHILIPPE *et al.*, 1994; but see WRAY *et al.*, 1996 for a different opinion). As a result, most internodes between the animal taxa are very short and therefore difficult to reconstruct. Possibly, the addition of more sequences representative for the different animal taxa can further stabilize the animal tree, as previously suggested (TURBEVILLE *et al.*, 1992; LECOINTRE *et al.*, 1993). Alternatively, the problem of short internodes in animal phylogeny can be tackled by combining different genes into one long alignment. In this so-called multigenic approach, even complete (mitochondrial) genomes can be compared (CUMMINGS *et al.*, 1995, OTTO *et al.* 1996). Additionally, information such as the gene order and inferred number of gene rearrangements can then be taken into account in the study of evolutionary relationships (BOORE & BROWN, 1994; BOORE *et al.*, 1995). For a more general discussion about animal phylogenies taking into account site-to-site rate calibration, we refer to WINNEPENNINCKX *et al.* (1998)

DISCUSSION

Pairwise distance methods are often regarded as being inferior to character-based methods such as maximum parsimony because they strongly reduce the phylogenetic information of the sequences. However, as shown in this study and elsewhere (VAN DE PEER *et al.*, 1993; 1996b, VAN DE PEER & DE WACHTER, 1997a), distance methods can be of great value as long as the distances are estimated accurately. Moreover, it is indeed just one of the advantages and strengths of methods that are based on an explicit model of evolution (such as distance and maximum likelihood methods), that appropriate substitution models can be developed to correct for multiple mutations. Furthermore, tree inferring methods such as neighbor-joining have the added bonus of being very fast, which allows the construction of large trees, including several hundreds of sequences (see *e.g.* VAN DE PEER & DE WACHTER, 1997a).

The use of a more sophisticated (and realistic) substitution model such as the one presented here can thus make a big difference in the inferred tree topology (see *e.g.* Figs 2 and 3). However, it is not always necessary to use such complicated models for estimating evolutionary distances. As can be seen in Fig. 1c, the effect of using different models of evolution can be quite extensive for large evolutionary distances, but for small distances

(<0.25), the effect is often only marginal. Moreover, if closely related sequences are being analyzed by distance methods, it is even better to use a simpler model such as the one of JUKES & CANTOR, because of the lower variance compared to more sophisticated methods (SWOFFORD *et al.*, 1996).

ACKNOWLEDGEMENTS

Yves Van de Peer is Research Fellow of the Fund for Scientific Research – Flanders and acknowledges support by the Special Research Fund of the University of Antwerp (UIA).

REFERENCES

ABOUHEIF, E., R. ZARDOYA & A. MEYER (1998) – Limitations of metazoan 18S rRNA sequence data: implications for reconstructing a phylogeny of the animal kingdom and inferring the reality of the Cambrian Explosion. *J. Mol. Evol.*, **47**: 394-405.

ADOUTTE, A. & H. PHILIPPE (1993) – The major lines of Metazoan evolution: summary of traditional evidence and lessons from ribosomal RNA sequence analysis. In: *Comparative molecular neurobiology*. PICHON, Y. (ed.). Birkhäuser Verlag, Basel, p. 1-30.

BOORE, J.L. & W.M. BROWN (1994) – Mitochondrial genomes and the phylogeny of mollusks. *Nautilus*, **108** (Suppl. 2): 61-78.

BOORE, J.L., T.M. COLLINS, D. STANTON, L. DAEHLER & W.M. BROWN (1995) – Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature*, **376**: 163-165.

CUMMINGS, M., S. OTTO & J. WAKELEY (1995) – Sampling properties of DNA sequence data in phylogenetic analysis. *Mol. Biol. Evol.*, **12**: 814-822.

DE RIJK, P., Y. VAN DE PEER, I. VAN DEN BROECK & R. DE WACHTER (1995) – Evolution according to large ribosomal subunit RNA. *J. Mol. Evol.*, **41**: 366-375.

ERWIN, D.H. (1991) – Metazoan phylogeny and the cambrian radiation. *Trends Ecol. Evol.*, **6**: 131-134.

FELSENSTEIN, J. (1988) – Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.*, **22**: 521-565.

FELSENSTEIN, J. (1985) – Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**: 783-791.

JIN, L. & M. NEI (1990) – Limitations of the evolutionary parsimony method of phylogenetic analysis. *Mol. Biol. Evol.*, **7**: 82-102.

JUKES, T.H. & C.R. CANTOR (1969) – Evolution of protein molecules. In: *Mammalian protein metabolism*. MUNRO, H.H. (ed.). Academic Press, New York, p. 21-132

KIMURA, M. (1980) – A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, **16**: 111-120.

LECOINTRE, G., H. PHILIPPE, H.L.V. LÊ & H. LE GUYADER (1993) – Species sampling has a major impact on phylogenetic inference. *Mol. Phyl. Evol.*, **2**: 205-224.

LOCKHART, P.J., M.A. STEEL, M.D. HENDY & D. PENNY (1994) – Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.*, **11**: 605-612.

MACKEY, L.Y., B. WINNEPENNINCKX, R. DE WACHTER, T. BACKELJAU, P. EMSCHERMANN & J.R. GAREY (1996) – 18S rRNA suggests that entoprocta are protostomes, unrelated to Ectoprocta. *J. Mol. Evol.*, **42**: 552-559.

NEI, M. (1987) – Phylogenetic trees. In: *Molecular Evolutionary Genetics*. Columbia Press, New York: 287-326.

OLSEN, G.J. (1987) – Earliest phylogenetic branchings: comparing rRNA-based evolutionary trees inferred with various techniques. *Cold Spring Harbor Symposia on Quantitative Biology*, LII: 825-837.

OTTO, S.P., M.P. CUMMINGS & J. WAKELEY (1996) – Inferring phylogenies from sequence data: the effect of sampling. In: *New uses for new phylogenies*. HARVEY, P.H., A.J.L. BROWN, J. MAYNARD SMITH & S. NEE (eds.). Oxford University Press, Oxford, p. 103-115.

PAWLOWSKI, J., J.-I. MONTOYA-BURGOS, J.F. FAHRNI, J. WIIEST & L. ZANINETTI L. (1996) – Origin of the Mesozoa inferred from 18S rRNA gene sequences. *Mol. Biol. Evol.*, **13**: 1128-1132.

PHILIPPE, H., A. CHENUIL & A. ADOUTTE (1994) – Can the Cambrian explosion be inferred through molecular phylogeny? *Development* (Suppl.): 15-25.

RZHETSKY, A. & M. NEI (1994) – Unbiased estimates of the number of nucleotide substitutions when substitution rate varies among different sites. *J. Mol. Evol.*, **38**: 295-299.

SAITOU, N. & M. NEI (1987) – The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**: 406-425.

SULLIVAN, J., K.E. HOLSINGER & C. SIMON (1996) – The effect of topology on estimation of among-site rate variation. *J. Mol. Evol.*, **42**: 308-312.

SWOFFORD, D.L., G.J. OLSEN, P.J. WADDELL & D.M. HILLIS (1996) – Phylogenetic Inference. In: *Molecular Systematics*. HILLIS, D.M., C. MORITZ & B.K. MABLE (eds.). Sinauer Associates, Sunderland, USA.: 407-514.

TAJIMA, F. & M. NEI (1984) - Estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.*, **1**: 269-285.

TAMURA, K. & M. NEI (1993) – Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.*, **10**: 512-526.

TURBEVILLE, J.M., K.G. FIELD & R.A. RAFF (1992) – Phylogenetic position of the Nemertini, inferred from 18S rRNA sequences: molecular data as a test of morphological character homology. *Mol. Biol. Evol.*, **9**: 235-249.

VAN DE PEER, Y. & R. DE WACHTER (1997a) – Evolutionary relationships among the eukaryotic crown taxa taking into account site to site rate variation in 18S rRNA. *J. Mol. Evol.*, **45**: 619-630.

VAN DE PEER, Y. & R. DE WACHTER (1997b) – Construction of evolutionary distance trees with TREECON for Windows: accounting for variation in nucleotide substitution rate among sites. *Comput. Applic. Biosci.*, **13**: 227-230.

VAN DE PEER, Y., J.-M. NEEFS, P. DE RIJK & R. DE WACHTER (1993) – Reconstructing evolution from eukaryotic small ribosomal subunit RNA sequences: calibration of the molecular clock. *J. Mol. Evol.*, **37**: 221-232.

VAN DE PEER, Y., G. VAN DER AUWERA & R. DE WACHTER (1996a) - The evolution of stramenopiles and alveolates as derived by "substitution rate calibration" of small ribosomal subunit RNA. *J. Mol. Evol.* **42**: 201-210.

VAN DE PEER, Y., S.A. RENSING, U.-G. MAIER & R. DE WACHTER (1996b) – Substitution rate calibration of small subunit ribosomal RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. *Proc. Natl. Acad. Sci. USA*, **93**: 7732-7736.

VAN DE PEER, Y., S. CHAPELLE & R. DE WACHTER (1996c) – A quantitative map of nucleotide substitution rates in bacterial ribosomal subunit RNA. *Nucleic Acids Res*, **24**: 3381-3391.

VAN DE PEER, Y., A. CAERS, P. DE RIJK, & R. DE WACHTER (1998) – Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Res.*, **26**: 179-182.

WINNEPENNINCKX, B., T. BACKELJAU & R. DE WACHTER (1996) – Investigation of molluscan phylogeny on the basis of 18S rRNA sequences. *Mol. Biol. Evol.*, **13**: 1306-1317.

WINNEPENNINCKX, B., Y. VAN DE PEER & T. BACKELJAU (1998) – Metazoan relationships on the basis of 18S rRNA sequences: a few years later ... *Amer. Zool.*, **38**, (in press)

WRAY, G.A., J.S. LEVINTON & L.H. SHAPIRO (1996) – Molecular evidence for deep precambrian divergences among metazoan phyla. *Science*, **274**: 568-573.

YANG, Z. (1996) – Among-site rate variation and its impact on phylogenetic analyses. *TREE*, **11**: 367-372.

ZHARKIKH, A. (1994) – Estimation of evolutionary distances between nucleotide sequences. *J. Mol. Evol.*, **39**: 315-329.