J Mol Evol (2001) 53:155–159
DOI: 10.1007/s002390010203

JOURNAL OF MOLECULAR EVOLUTION

Letter to the Editor

Phylogenetic Analysis of Viroid and Viroid-Like Satellite RNAs from Plants: A Reassessment

Santiago F. Elena,[1] Joaquín Dopazo,[2] Marcos de la Peña,[3] Ricardo Flores,[3] Theodor O. Diener,[4] Andrés Moya[1]

[1] Institut *Cavanilles* de Biodiversitat i Biologia Evolutiva and Departament de Genètica, Universitat de València, Apartat 2085, 46071 València, Spain
[2] Unidad de Bioinformática, Centro Nacional de Investigaciones Oncológicas Carlos III, Carretera Majadahonda-Pozuelo Km 2, 28220 Majadahonda, Madrid, Spain
[3] Instituto de Biología Molecular y Celular de Plantas (UPV-CSIC), Universidad Politécnica de Valencia, Avenida de los Naranjos, 46022 Valencia, Spain
[4] Center for Agricultural Biotechnology, University of Maryland Biotechnology Institute, College Park, MD 20742, USA

Received: 17 January 2001 / Accepted: 16 March 2001

Abstract. The proposed monophyletic origin of a group of subviral plant pathogens (viroids and viroid-like satellite RNAs), as well as the phylogenetic relationships and the resulting taxonomy of these entities, has been recently questioned. The criticism comes from the (apparent) lack of sequence similarity among these RNAs necessary to reliably infer a phylogeny. Here we show that, despite their low overall sequence similarity, a sequence alignment manually adjusted to take into account all the local similarities and the insertions/deletions and duplications/rearrangements described in the literature for viroids and viroid-like satellite RNA, along with the use of an appropriate estimator of genetic distances, constitutes a data set suitable for a phylogenetic reconstruction. When the likelihood-mapping method was applied to this data set, the tree-likeness obtained was higher than that corresponding to a sequence alignment that does not take into consideration the local similarities. In addition, bootstrap analysis also supports the major groups previously proposed and the reconstruction is consistent with the biological properties of this RNAs.

Correspondence to: Dr. Andrés Moya; *email:* andres.moya@uv.es

Key words: Likelihood-mapping — Molecular phylogeny — Tree-likeness — Viroid — Viroid-like satellite RNA

Viroids, subviral pathogens of higher plants, are small (246–400 nucleotide residues, nt), unencapsidated, single-stranded, circular RNAs characterized by highly base-paired secondary structures (Diener 1991; Flores et al. 2000). Viroids do not code for any protein, yet they replicate autonomously (without the assistance of helper viruses) in susceptible cells. Viroid-like satellite RNAs resemble viroids structurally and like them replicate through a rolling-circle mechanism, but they are found within the capsids of certain helper viruses required for their replication (Bruening et al. 1991; Keese and Symons 1987). In a previous work (Elena et al. 1991), we proposed a classification of viroids and viroid-like satellite RNAs, based on a phylogenetic study of the available sequences, that additionally gave support to the previously posited monophyletic origin of all these molecules (Diener 1989). The phylogenetic tree grouped these entities according to the presence of conserved sequence motifs, with the existence or absence of hammerhead ribozymes (which in some of them mediate self-cleavage

of the replicative intermediates into unit-length forms), and with their biological properties. A test for the tree-likeness of the aligned sequences (Eigen et al. 1988) supported the feasibility of performing a phylogenetic analysis.

Recently, Jenkins et al. (2000) questioned this analysis by arguing that the sequence similarities needed for inferring a trustable phylogeny were not present. We agree that the overall similarity is low when viroids from the two different families are compared, or when the comparison involves viroids and viroid-like satellite RNAs, but maintain that for any phylogenetic analysis to be valid, existing local similarities must be taken into account. Also, viroids and viroid-like satellites RNAs accumulate predominantly as circular molecules (or they go through a circular stage during replication). Consequently, the decision of assigning the +1 position to a given nucleotide is in principle arbitrary and conditions the result of the alignment must be considered. Furthermore, because there are considerable differences in length between these sequences, and comparisons of individual sequences with others have disclosed the presence of insertions/deletions and duplications/rearrangements, manual adjustments are required for aligning the stretches of homologous positions prior to phylogenetic analysis. It is important to note that Jenkins et al. (2000) used simple dissimilarity to measure the distances. This estimator, however, is well known to give erroneous estimations of the distance when the global divergence is high (Nei 1987), which is the case. When sequences with a high degree of divergence are compared, it becomes necessary to use estimators that correct for multiple substitutions. Here we used the nucleotide substitution model proposed by Hasegawa et al. (1985) corrected for gamma-distributed rates among sites (HKY-Γ). By doing so, we obtained bootstrap support for clusters of sequences that in the analysis by Jenkins et al. (2000) were predicted to be simply unrelated to each other. This result could not be obtained in absence of a true phylogenetic signal. In addition to this, the Monte-Carlo method employed by Jenkins et al. (2000) is not the most appropriate tool to assess the significance of the similarities when they are restricted to certain regions of the molecule, because the method masks the signal provided by these local similarities. In such a situation, a different approach based on a procedure avoiding sequence jumbling is needed. Here, we propose the use of the likelihood-mapping algorithm advanced by Strimmer and von Haeseler (1997), a technique specially developed for testing the existence of a phylogenetic signal in a set of distantly related aligned sequences.

The accepted viroid classification (Flores et al. 2000) is based on the existence of conserved sequence and structural motifs, and allocates these RNAs into two families. Members of family *Pospiviroidae* (25 of the 28 known viroids), whose type species is PSTVd, have a central conserved region (CCR) in their most stable rod-like secondary structure and do not exhibit self-cleavage mediated by hammerhead ribozymes. The three members of family *Avsunviroidae,* whose type species is ASBVd, lack a CCR and their strands of both polarities self-cleave through hammerhead ribozymes. Within each family, viroid species are grouped into genera according to different criteria. In family *Pospiviroidae,* five genera are distinguished, depending on the sequences of the CCR (formed by two sets of conserved nt varying in size from 13 to 30). A second criterion is the presence or absence of two other conserved motifs: the terminal conserved region (TCR) of 13–16 nt, found in all members of genera *Pospi-* and *Apscaviroid* as well as in the two largest members (CbVd2 and CbVd3) of genus *Coleviroid,* and the terminal conserved hairpin (TCH) of 13 nt found in HSVd and in all members of genus *Cocadviroid.* The CCR, TCR, and TCH are motifs conserved not only in sequence but also in location within the rod-like conformation and even in secondary structure (this is the case of the TCH, as reflected by its name, and of the CCR upper strand, which forms part of the so-called hairpin I within an alternative metaestable conformation) (for a review see Flores et al. 1997). In addition, there is an A-rich region and a U-rich region conserved in similar positions in the P domain of the rod-like structure of members of genera *Pospi-, Hostu-, Apsca-,* and *Coleviroid.* In family *Avsunviroidae,* two genera have been established according to several criteria that include base composition, type of hammerhead, and overall secondary structure (Flores et al. 2000).

In the present work, we first produced an alignment of viroids using CLUSTAL-X (Thompson et al. 1997), manually editing the result to preserve all the local similarities indicated above. Numbering of sequences was based on the local similarities found between the upper strand of the CCRs of most members of family *Pospiviroidae* and the upper left-hand portions of the hammerhead structures of members of the family *Avsunviroidae* (Diener 1989). The phylogenetic tree derived by the neighbor-joining method from such an alignment of the complete sequences reproduces the same distribution in families and genera obtained when only using the conserved sequence motifs (Flores et al. 2000).

Viroid-like satellite RNAs, which have been classified according to the type of helper virus they depend on, have in all cases hammerhead structures in one or both polarity strands (Bruening et al. 1991). The conserved sequence motifs present in these ribozymes serve as a link between viroid-like satellite RNAs (and viroids), as well as to adopt a consistent numbering for all of them. Moreover, viroid-like satellite RNAs from *Nepoviruses* have in their minus polarity strands hairpin ribozymes, which also contain conserved sequence motifs (De-Young et al. 1995), and other conserved motifs also exist in viroid-like satellite RNAs from *Sobemoviruses,* most

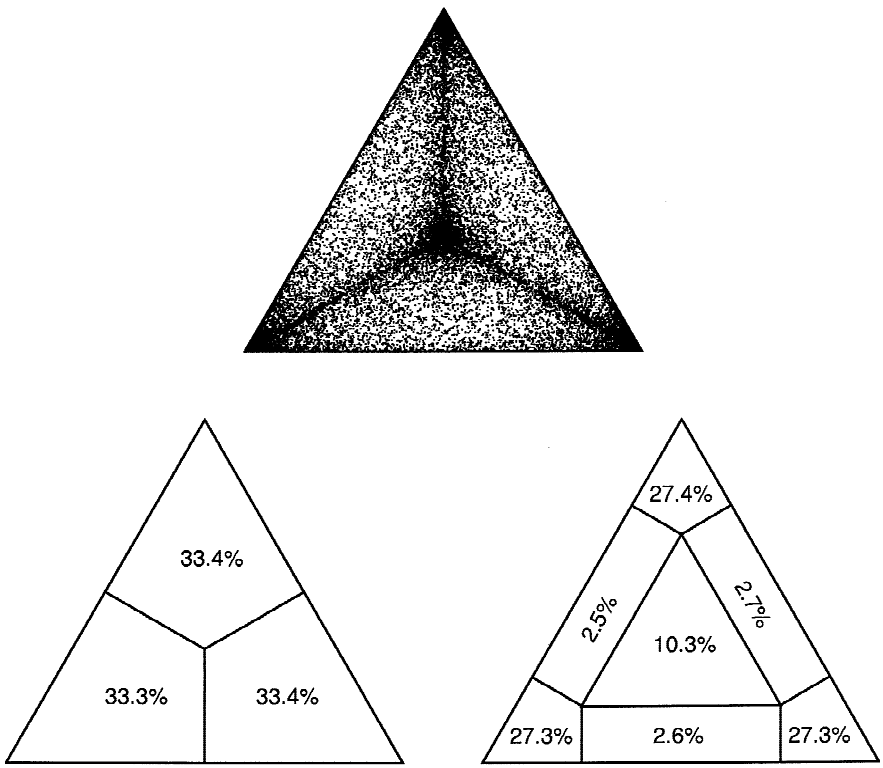

Fig. 1. Likelihood-mapping analysis for viroid and viroid-like satellite RNAs. The upper diagram shows the distribution pattern of all possible quartets. The lower left diagram shows the percentage of quartets in the neighborhood of each corner (completely resolved quartets) and the lower right diagram the percentage of quartets in each of the seven attraction basins. Only those quartets found in the three corners represent well-resolved phylogenetic relationships. The nucleotide substitution model employed in the analysis was that proposed by Hasegawa et al. (1985) with gamma-distributed rates (16 categories).

of them short and interspersed but one long and covering the left terminal domains of vLTSV and vRYMV (Collins et al. 1998). When the alignment of viroid-like satellite RNAs is manually adjusted to include these similarities, the resulting phylogenetic tree obtained by the neighbor-joining method reproduces a distribution into three groups concordant with their biological properties (data not shown). We have excluded in the present analysis the viroid-like domain of hepatitis delta virus RNA which is now known to self-cleave through a specific class of ribozyme (Been 1994), different from the hammerhead and hairpin structures present in viroid and viroid-like satellite RNAs from plants. However, a common origin has been recently proposed for the hammerhead, hairpin, and hepatitis delta ribozymes (Harris and Elder 2000).

In an ultimate step, we made a composed alignment of the two previous partial alignments, first by manually anchoring the similarities described in viroids and viroid-like satellite RNAs, and then with CLUSTAL-X, realigning those nonhomologous regions to maximize the overall similarity. A copy of this alignment can be obtained from ftp://serbio.uv.es/pub/incomming/elena/viroids_sats.msf by anonymous FTP.

We next addressed the question of whether this data set contains a significant phylogenetic signal. Fig. 1 shows the result of the likelihood-mapping analysis for our alignment. Adding the tree-likeness values obtained for each corner that represents a fully determined quartet, a value of 82% was obtained. Because a tree-likeness value well above 50% has been suggested as advisable to be trusted (Strimmer and von Haeseler 1996), we conclude that despite the high heterogeneity observed among viroid and viroid-like satellite RNAs, still enough room exists for making phylogenetic inferences.

Fig. 2 shows the neighbor-joining tree obtained for all the sequences incorporated on this analysis with a bootstrap based on 1000 replicates. The substitution model employed (HKY-Γ) produced the largest reduction in log-likelihood among many tested (data not shown) (Yang 1997). The tree reproduces the groups obtained previously (Elena et al. 1991). Bootstrap frequencies separate members of family *Pospiviroidae* into three groups ($P \geq 0.98$ in all cases), one including genera *Pospi-*, *Hostu-* and *Cocadviroid*, consistent with the extensive similarities existing between their CCRs, and the other two groups including genera *Apsca-* and *Coleviroid*, respectively. Further division into nested classes gets more difficult, but is still possible: the first group can be split into two clusters containing members of genera *Pospi-* and *Cocadviroid*, although with low *P*-values for IrVd, HLVd, and HSVd. Similarly, members of genus *Apscaviroid* can be divided into two groups, the first highly significant incorporating ASSVd, ADFVd, and CVd3, and the second encompassing all other viroids classically assigned to this genus ($P = 0.94$).

Regarding the autocatalytic RNAs, members of family *Avsunviroidae* appear in the tree in an intermediate position between viroids of the other family and viroid-like satellites RNA, being ASBVd closer to viroids and the two components of genus *Pelamoviroid* closer to satellites, but forming a fully monophyletic group. On

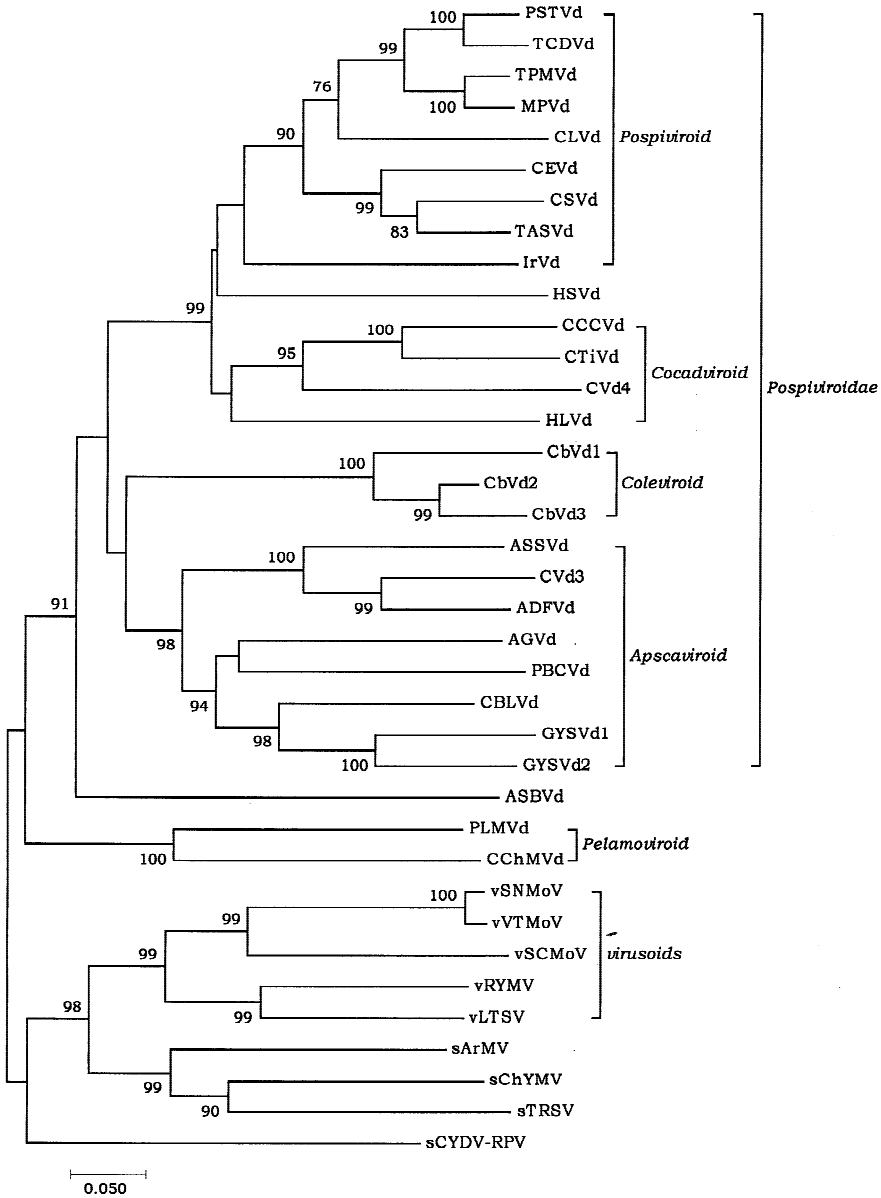

Fig. 2. Neighbor-joining phylogenetic tree obtained from the alignment of viroid and viroid-like RNA satellites. The distance matrix was obtained according to Hasegawa et al. (1985) but with gamma-distributed rates among sites (16 categories). Bootstrap values were based on 1000 random replicates (only values >70% are reported). Viroids: PSTVd (potato spindle tuber); TCDVd (tomato chlorotic dwarf); MPVd (Mexican papita); TPMVd (tomato planta macho); CSVd (chrysanthemum stunt); CEVd (citrus exocortis); TASVd (tomato apical stunt); IrVd (iresine 1); CLVd (columnea latent); HSVd (hop stunt); CCCVd (coconut cadang-cadang); CTiVd (coconut tinangaja); HLVd (hop latent); CVd4 (citrus IV); ASSVd (apple scar skin); CVd3 (citrus III); ADFVd (apple dimple fruit); GYSVd1 (grapevine yellow speckle 1); GYSVd2 (grapevine yellow speckle 2); CBLVd (citrus bent leaf); PBCVd (pear blister canker); AGVd (Australian grapevine); CbVd1 (*Coleus blumei* 1); CbVd2 (*Coleus blumei* 2); CbVd3 (*Coleus blumei* 3); ASBVd (avocado sunblotch); PLMVd (peach latent mosaic); CChMVd (chrysanthemum chlorotic mottle). Viroid-like satellite RNAs: vLTSV (lucerne transient streak virus); sRYMV (rice yellow mottle virus); vSCMoV (subterranean clover mottle virus); vSNMoV (*Solanum nodiflorum* mottle virus); vVTMoV (velvet tobacco mottle virus); sTRSV (tobacco ringspot virus); sArMV (*Arabis* mosaic virus); sChYMV (chicory yellow mottle virus); sCYDV-RPV (cereal yellow dwarf virus-RPV).

the other hand, viroid-like satellite RNAs cluster according to the type of their helper virus: those from *Sobemoviruses* (virusoids) and *Nepoviruses* constituting a monophyletic group ($P = 0.98$), and the only representative with a *Luteovirus* as a helper (sCYDV-RPV) forming an outgroup for the viroid-like satellite RNA cluster.

The likelihood-mapping analysis, which is a helpful tool to investigate how well supported the internal branches of a tree topology are, was used to assess the likelihood of specific branches of the tree. The high tree-likeness value (96.1%) of the branch separating viroids from viroid-like satellite RNAs gives support to the notion of a real relationship between these two groups of subviral RNAs. On the other hand, the tree-likeness of the branch separating the autocalytic RNAs (members of family *Avsunviroidae* and viroid-like satellite RNAs)

from viroids of the other family was also highly significant (93.9%).

When the same analysis was applied to an alignment without the extensive modifications made above, the result was clearly different. For instance, the tree-likeness of the data was reduced to 62.9% and the neighbor-joining tree generated with this alignment did not reproduce any of the groups indicated above (except in the case of members of genus *Pospiviroid*), with viroids and viroid-like satellite RNAs appearing intermingled. Therefore, in spite of the blurring effects of recombination, the tree-likeness values derived from our alignment indicate that the true relationships among these RNAs are preserved.

In conclusion, previous attempts to reconstruct the phylogenetic relationships between viroids, particularly those of family *Pospiviroidae* (Elena et al. 1991), appear well founded and are supported by the present analysis with a more complete set of sequences and an alternative methodology to test the tree likeness. Despite the overall low similarity that these sequences show, the existence of domains conserved in primary and secondary structure, as well as in location within the rod-like conformation proposed for these RNAs, still defines significant groups and must be taken into consideration for a proper phylogenetic reconstruction. A parallel argument can be extended to the case of viroid-like satellite RNAs, which also contain conserved domains. One of these domains, the hammerhead structure, is also present in members of family *Avsunviroidae* and serves as a connecting bridge between viroids and viroid-like satellite RNAs. Although the possibility exists that the hammerhead ribozyme could have emerged more that once in evolution, the ample variety of structurally different ribozymes that can be obtained by *in vitro* selection from a pool of randomized sequences (Tang and Breaker 2000), suggests a single appearance for this specific ribozyme.

Acknowledgments. This work was supported by grants PM97-0060-C02-02 from the Spanish DGES and 1FD1997-2328 from the U.E. (U. València), and grant PB98-0500 from DGES (UPV-CSIC).

References

Been M (1994) *Cis*- and *trans*-acting ribozymes from a human pathogen, hepatitis delta virus. Trends Biochem Sci 19:251–256

Bruening G, Feldstein PA, Buzayan JM, van Tol H, Passmore BK, deBear J, Gough GR, Gilham PT, Eckstein F (1991) Satellite tobacco ringspot virus RNA: self-cleavage and ligation reactions in replication. In: Maramorosch K (ed) Viroids and satellites: molecular parasites at the frontier of life. CRC Press, Boca Raton, pp 141–158

Collins RF, Gellatly DL, Sehgal OP, Abounhaidar MG (1998) Self-cleaving circular RNA associated with rice yellow mottle virus is the smallest viroid-like RNA. Virology 241:269–275

DeYoung MB, Siwkowski AM, Lian Y, Hampel A (1995) Catalytic properties of hairpin ribozymes derived from chicory yellow mottle virus and arabis mosaic virus satellite RNAs. Biochemistry 34: 15785–15791

Diener TO (1989) Circular RNAs: relics of precellular evolution? Proc Natl Acad Sci USA 86:9370–9374

Diener TO (1991) Subviral pathogens of plants: viroids and viroidlike satellite RNAs. FASEB J 5:2808–2813

Eigen M, Winkler-Oswatitsch R, Dress A (1988) Statistical geometry in sequence space: a method of quantitative comparative sequence analysis. Proc Natl Acad Sci USA 85:5913–5917

Elena SF, Dopazo J, Flores R, Diener TO, Moya A (1991) Phylogeny of viroids viroidlike satellite RNAs, and the viroidlike domain of hepatitis δ virus RNA. Proc Natl Acad Sci USA 88:5631–5634

Flores R, Di Serio F, Hernández C (1997) Viroids: the noncoding genomes. Semin Virol 8:65–73

Flores R, Randles JW, Bar-Joseph M, Diener TO (2000) Viroids. In: van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR, Wickner RB (eds) Virus taxonomy, 7th Rep Intern Com Taxon Viruses. Academic Press, San Diego, pp 1009–1024

Harris RJ, Elder D (2000) Ribozyme relationships: the hammerhead, hepatitis delta and hairpin ribozymes have a common origin. J Mol Evol 51:182–184

Hasegawa M, Kishino H, Yano T (1985) Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22: 160–174

Jenkins GM, Woelk CH, Rambaut A, Homes EC (2000) Testing the extent of sequence similarity among viroids, satellite RNAs, and the hepatitis delta virus. J Mol Evol 50:98–102

Keese P, Symons RH (1987) The structure of viroids and virusoids. In: Semancik JS (ed) Viroids and viroid-like pathogens. CRC Press, Boca Raton, pp 1–47

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, p 64

Strimmer, K, von Haeseler A (1996) Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol Biol Evol 13:964–969

Strimmer K, von Haeseler A (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci USA 94:6815–6819

Tang J, Breaker R (2000) Structural diversity of self-cleaving ribozymes. Proc Natl Acad Sci USA 97:5784–5789

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS 13:555–556