Volume 309, number 2, 123-126 FEBS 11490 September 1992

Structure of the small ribosomal subunit RNA of the pulmonate snail, *Limicolaria kambeul*, and phylogenetic analysis of the Metazoa

Birgitta Winnepennickx[a], Thierry Backeljau[b], Yves van de Peer[a] and Rupert De Wachter[a]

[a]*Departement Biochemie, Universitaire Instelling Antwerpen, Universiteitsplein 1, B-2610 Antwerpen, Belgium* and [b]*Koninklijk Belgisch Instituut voor Natuurwetenschappen, Afdeling Malacologie, Vautierstraat 29, B-1040 Brussel, Belgium*

Received 11 June 1992; revised version received 20 July 1992

The complete nucleotide sequence of the small ribosomal subunit RNA of the gastropod, *Limicolaria kambeul*, was determined and used to infer a secondary structure model. In order to clarify the phylogenetic position of the Mollusca among the Metazoa, an evolutionary tree was constructed by neighbor-joining, starting from an alignment of small ribosomal subunit RNA sequences. The Mollusca appear to be a monophyletic group, related to Arthropoda and Chordata in an unresolved trichotomy.

18 S rRNA sequence; Metazoan evolution; Mollusca; Gastropoda; *Limicolaria kambeul*; Phylogeny

1. INTRODUCTION

In spite of many morphological and embryological studies, the phylogenetic relationships among and within many metazoan taxa still remain poorly defined. This is due to the lack of informative morphological and anatomical characters common to different taxa, and the doubtful homology of similar traits.

For example, the phylogeny of the Mollusca still remains a point of discussion although studies concerning this issue already date from the previous century (e.g. [1]). Mollusca form a huge phylum, showing such large differences in body plan among different classes that the phylum can only be defined by a set of characteristics [2], such as the presence of a mantle (pallium), a pallial cavity, a foot, and a radula. Morphological anatomical and embryological features do not indicate clearly whether the molluscs were directly derived from an acoelomate (platyhelminthomorph) ancestor [2], or whether they are true coelomates [3]. The hypotheses concerning the origin of the branching pattern within the phylum [2–8] also remain controversial.

Molecular approaches, in addition to traditional ones, can contribute to clarify metazoan evolutionary branching patterns. Today the most appropriate molecules to infer phylogenies at a broad range of levels are the large ribosomal RNAs (18 S and 28 S rRNA for eukaryotes). By far the largest set of complete sequences is available for 18 S rRNA, viz. for 181 eukaryotic species among which 48 are metazoans [9]. Field et al. [10] determined an additional set of partial 18 S rRNA sequences representing 22 classes belonging to ten different metazoan phyla. Analyses [10–13] of the latter data using different tree construction methods, uniformly pointed to a coelomate origin of the Mollusca. However, the monophyly of the molluscs and their relationship to other protostome coelomates remained in doubt. Field et al. [10] and Patterson [13] could not resolve the branching pattern of Mollusca and a group consisting of Annelida, Pogonophora, Brachiopoda, and Sipuncula. Ghiselin [11] considered the Mollusca as a sister taxon of the latter group, whereas Lake [12] found the Mollusca to be paraphyletic.

In this paper we present the first complete 18 S rRNA sequence of a gastropod viz., the terrestrial snail, *Limicolaria kambeul* (Bruguière, 1789) (Mollusca, Gastropoda, Pulmonata). We also describe some improvements in the secondary structure model for eukaryotic small subunit rRNA and give the preliminary results of a phylogenetic study concerning the Mollusca. Voucher material of *L. kambeul* is deposited in the Royal Belgian Institute of Natural Science, Brussels (general inventory number, I.G. No. 27.834).

2. MATERIALS AND METHODS

2.1. *DNA isolation*

L. kambeul was collected in the woods of Ouahigouya (Burkina Faso) and frozen alive (−80°C). The ovotestis of a single specimen was homogenized in the presence of powdered dry ice in a pre-chilled mortar and transferred to 15 ml of pre-heated (60°C) 2% (w/v) CTAB (hexadecyltrimethylammoniumbromide) buffer [14] containing 100 μg/ml proteinase K. DNA was extracted as described by Doyle [14].

Correspondence address: R. De Wachter, Departement Biochemie, Universiteit Antwerpen (UIA), Universiteitsplein 1, B-2610 Antwerpen, Belgium. Fax: (32) (3) 820 22 48.

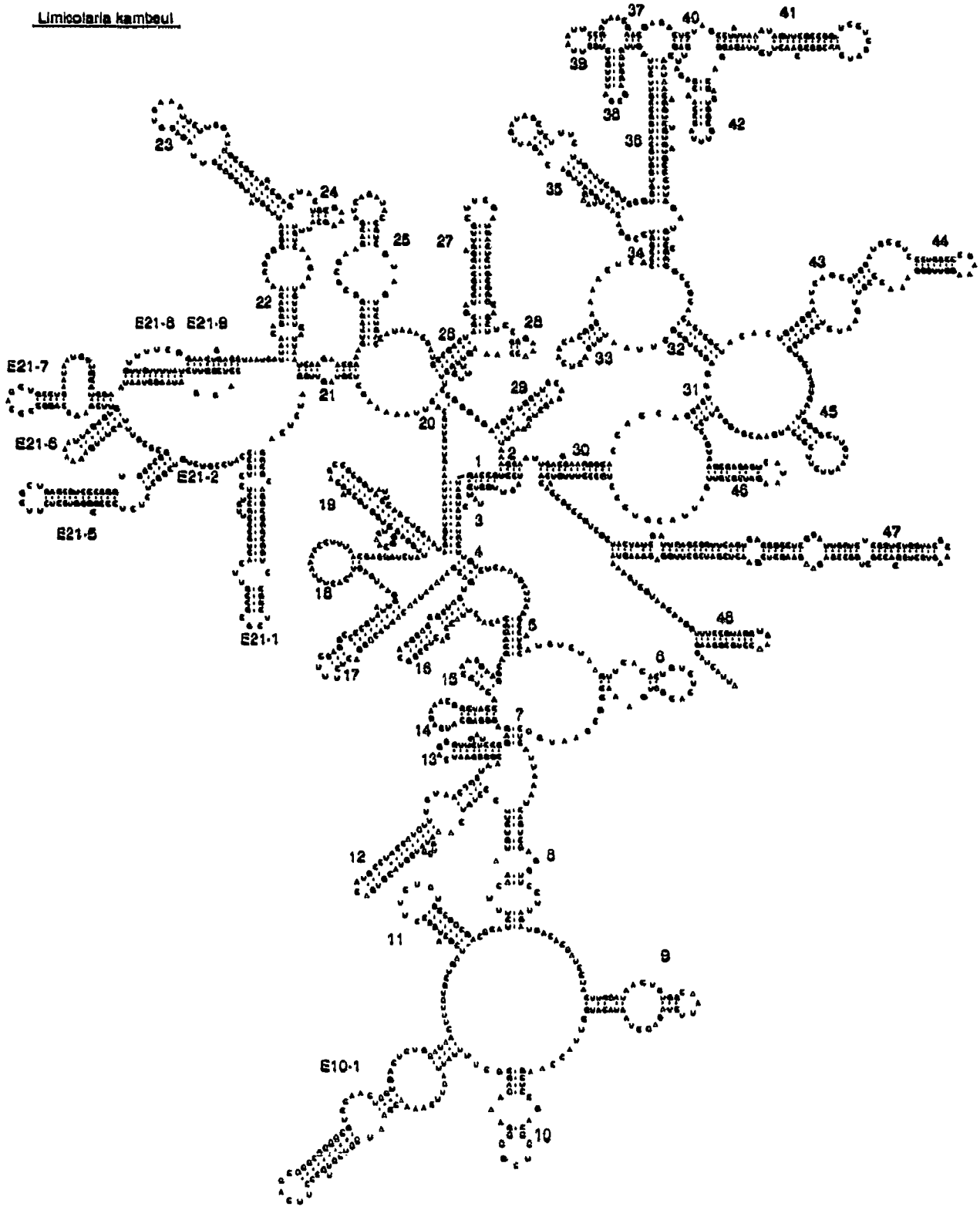

Fig. 1. Secondary structure model for 18 S rRNA of *Limicolaria kambeul*. Helix-numbering is as described by De Rijk et al. [9].

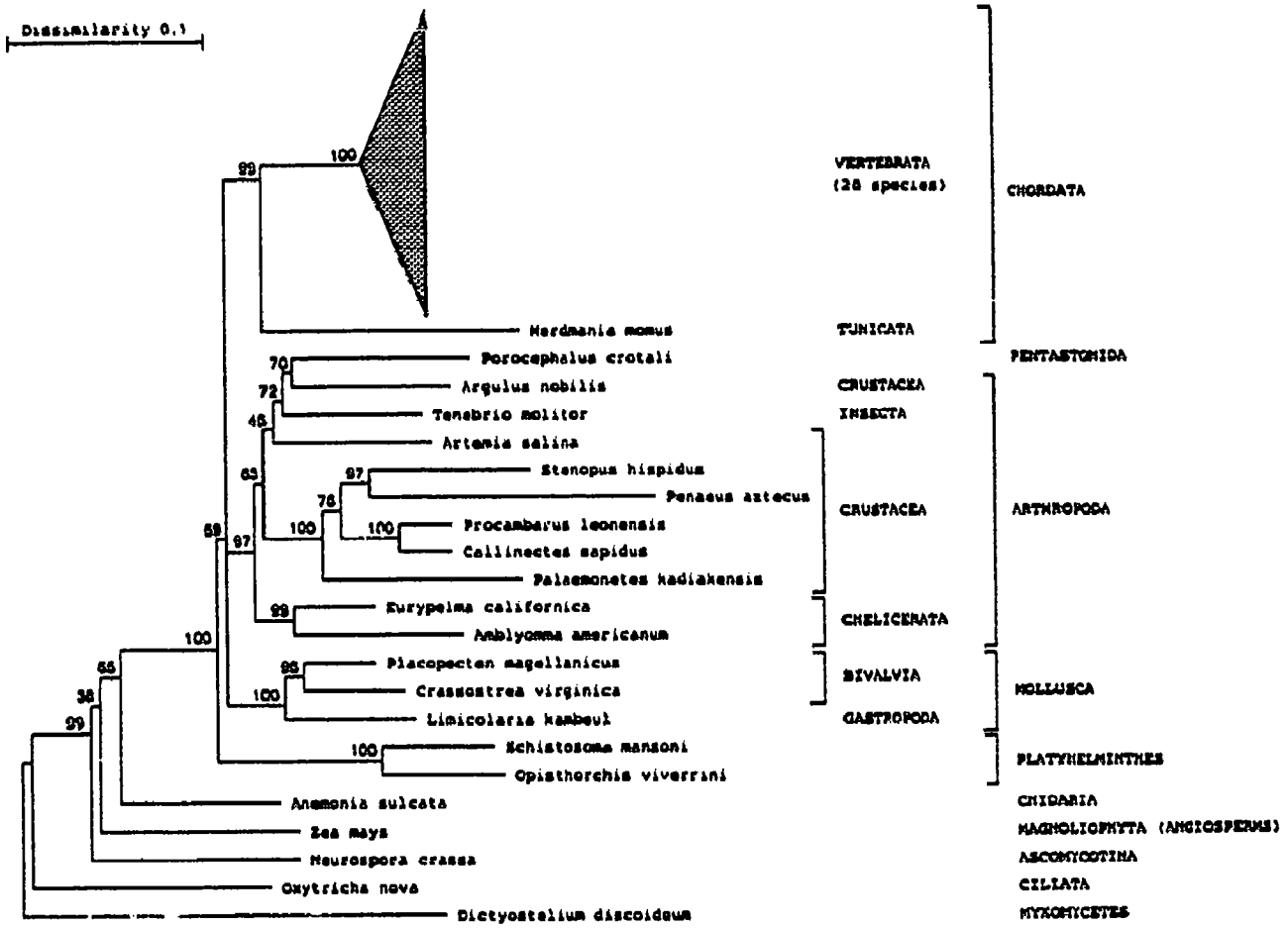

Fig. 2. Phylogenetic tree based on the 18 S rRNA sequences of 50 eukaryotes using neighbor-joining. *Dictyostelium discoideum* was used as the outgroup. Apart from the tree shown, 100 trees were constructed with re-sampling of nucleotide positions (bootstrap analysis [24]). Figures at the root of a cluster indicate the percentage of trees containing a cluster of this composition. The vertebrate cluster is represented by a triangle, the apex of which corresponds to the first divergence node and the apex-base distance is equal to the mean branch length.

2.2. *Gene cloning*

In order to identify a restriction enzyme suitable for isolating a DNA fragment containing the entire 18 S rRNA gene, the DNA was digested with seven restriction enzymes, *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, *Sal*I, *Sst*I and *Xba*I. DNA fragments were subsequently separated on a 0.8% (w/v) agarose gel and transferred via Southern blotting [15] to a Hybond-N membrane (Amersham, UK). This was hybridized with a plasmid containing the 18 S rRNA gene from *Artemia salina* [16], radioactively labeled by nick translation [17] using the Amersham Kit. Competent *E. coli* DH5α cells were transformed with a 7 kb *Pst*I restriction fragment containing the 18 S rRNA gene ligated into pUC19 [18]. The plasmid was isolated from recombinant *E. coli* DH5α cells using the alkaline lysis method [19].

2.3. *Sequencing*

Sequencing of both strands of the 18 S rRNA gene was performed by the dideoxynucleotide method [20], using 17 primers complementary to conserved regions. Sixteen of these have been published elsewhere [21] and a seventeenth has the sequence, TCTCAGGCTCCYTCTCCGG, complementary to positions 421–403 of *L. kambeul* 18S rRNA. For the sequencing reactions we used Sequenase 2.0 (USB; Cleveland, OH, USA) according to the manufacturer's instructions.

2.4. *Alignment and phylogenetic tree construction*

The 18S rRNA gene sequence of *L. kambeul* was aligned with those of other Metazoa present in the alignment of De Rijk et al. [9]. Dissimilarity matrices were calculated by pairwise sequence comparison as described by Van de Peer et al. [22]. Evolutionary trees were constructed using the neighbor-joining method [23]. Bootstrap re-sampling analysis was performed according to the principles of Felsenstein [24].

3. RESULTS AND DISCUSSION

3.1. *Primary and secondary structure of the 18 S rRNA*

We sequenced 1,940 nucleotides, comprising the entire 18 S rRNA gene. The presumed termini of the 18 S rRNA were identified by similarity with those of other eukaryotic sequences. The 1,839-nucleotide long *L. kambeul* 18 S rRNA sequence has been deposited in the EMBL nucleotide sequence library under accession number X66374. Fig. 1 shows a secondary structure model in accordance with our latest insights into the folding of the eukaryotic small ribosomal subunit RNA, deduced on a comparative basis from a sequence alignment of gradually increasing size [9]. Knowledge of a detailed secondary structure model of 18 S rRNA is

important because the boundaries of the secondary structure elements serve as markers during sequence alignment.

3.2. *Phylogenetic analysis*

We used 48 metazoan 18 S rRNA sequences [9]. This set consists of 19 complete sequences and 29 partial ones for which a minimum of 1,289 nucleotides were analysed. The partial sequences determined by Field et al. [10], for which 854–1,041 nucleotides were analysed, were not included. The sequences of two arthropods, *Aedes albopictus* and *Drosophila melanogaster*, and one nematod, *Caenorhabditis elegans*, were excluded from the analysis because their high evolutionary rates tend to induce systematic errors [25]. We included representatives of three non-metazoan eukaryotic phyla, viz. a plant, a fungus and a ciliate. The slime mold, *Dictyostelium discoideum*, was chosen as the outgroup. All nucleotide positions were taken into account for computation of the dissimilarities used to construct the tree shown in Fig. 2. The probability of occurrence of each cluster in the bootstrap resampling analysis [24] is indicated at its base. The branching pattern within the vertebrate cluster, represented as a triangle in Fig. 2, is described elsewhere [26,27].

The Mollusca appear as a monophyletic group, with *L. kambeul* as a sister taxon to the Bivalvia. This monophyly, which is strongly supported by bootstrap values (100/100), is in contrast with the results of Lake [12], who analysed the partial 18 S rRNA sequences of Field et al. [10] using evolutionary parsimony. Yet, this monophyletic character is only a very preliminary conclusion since many invertebrate phyla are not represented and only three complete mollusc sequences are available.

The monophyly of the Arthropoda, as proposed on the basis of partial sequences by Field et al. [10], Patterson [13] and Turbeville [28], but questioned by Lake [12], is strongly supported in our tree (bootstrap value 97/100). The pentastomid, *Porocephalus crotali*, though, is included in the arthropod cluster, as previously observed by Abele et al. [29].

The branching pattern between Chordata, Arthropoda and Mollusca appears to be an unresolved trichotomy. The choice of different outgroups yielded trees (not shown) with various branching patterns, viz. (i) the Mollusca as a sister group of the Platyhelminthes, (ii) the Mollusca branching off between the Platyhelminthes lineage and the coelomate cluster (represented by Chordata, Pentastomida and Arthropoda), (iii) a trichotomy as shown in Fig. 2. These preliminary results show that 18 S rRNA sequences from representatives of many more invertebrate phyla and classes of Mollusca must be determined in order to unravel this pattern, as well as the branching pattern of classes within the Mollusca.

Acknowledgements: We are indebted to Lic. T. Warmoes for collecting the specimens and to Dr. A.R. Mead (University of Arizona) for confirming their identification. This work was supported by FGWO Grant 3.0080.89 and FKFO Grant 2.0004.91.

REFERENCES

[1] von Ihering, H. (1876) Jahrb. Deutsch. Malakozool. Ges. 3, 97–148.
[2] von Salvini-Plawen, L. (1990) Iberus 9, 1–33.
[3] Götting, K.-J. (1980) Z. Zool. Syst. Evolut.-forsch. 18, 24–27.
[4] Milburn, P.W. (1960) Veliger 3, 43–48.
[5] von Salvini-Plawen, L. (1969) Malacologia 9, 191–216.
[6] von Salvini-Plawen, L. (1990) Lavori S.I.M. 23, 5–30.
[7] Scheltema, A.H. (1978) Malacologia 17, 99–109.
[8] Stasek, C.R. (1972) in: Chemical Zoology, vol. VII (Florkin, M. and Scheer, B.T., eds.) pp. 1–44.
[9] De Rijk, P., Neefs, J.-M., Van de Peer, Y. and De Wachter, R. (1992) Nucleic Acids res. 20, 2075–2089.
[10] Field, K.G., Olsen, G.J., Lane, D.J., Giovannoni, S.J., Ghiselin, M.T., Raff, E.C., Pace, N.R. and Raff, R.A. (1988) Science 239, 748–753.
[11] Ghiselin, M.T. (1988) in: Oxford Surveys in Evolutionary Biology, vol. 5 (Harvey, P. and Partridge, L., eds.) pp. 66–95, Oxford University Press.
[12] Lake, J.A. (1990) Proc. Natl. Acad. Sci. USA 87, 763–766.
[13] Patterson, C. (1988) in: The Hierarchy of Life (Fernholm, B., Bremer, K. and Järnvall, H., eds.) pp. 471–487, Elsevier, Amsterdam.
[14] Doyle, J. (1991) in: Molecular Techniques in Taxonomy, vol. H57 (Hewitt, G.M., Johnston, A.W.B. and Young, J.P.W., eds.) pp. 283–285, NATO ASI Series.
[15] Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.
[16] Nelles, L., Fang, B.-L., Volckaert, G., Vandenberghe, A. and De Wachter, R. (1984) Nucleic Acids Res. 12, 8749–8768.
[17] Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237–251.
[18] Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103–119.
[19] Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513–1522.
[20] Sanger, F., Nicklen, S. and Coulson, R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[21] Hendriks, L., Goris, A., Neefs, J.-M., Van de Peer, Y., Hennebert, G. and De Wachter, R. (1989) System. Appl. Microbiol. 12, 223–229.
[22] Van de Peer, Y., Neefs, J.-M. and De Wachter, R. (1990) J. Mol. Evol. 30, 463–476.
[23] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.
[24] Felsenstein, J. (1985) Evolution 39, 783–791.
[25] Swofford, D.L. and Olsen, G.J. (1990) in: Molecular Systematics (Hillis, D.M. and Moritz, C., eds.) pp. 411–499, Sinauer Associates, Sunderland, MA, USA.
[26] Hedges, S.B., Moberg, K.D. and Maxsons, L.R. (1990) Mol. Biol. Evol. 7, 607–633.
[27] Van de Peer, Y., Neefs, J.-M., De Rijk, P. and De Wachter, R. (1992) Biochem. Syst. Ecol. (in press).
[28] Turbeville, J.M., Pfeifer, D.M., Field, K.G. and Raff, R.A. (1991) Mol. Biol. Evol. 8, 669–686.
[29] Abele, L.G., Kim, W. and Felgenhauer, B.E. (1989) Mol. Biol. Evol. 6, 685–691.