Volume 11 Number 9 1983 Nucleic Acids Research

The nucleotide sequences of the 5S rRNAs of four mushrooms and their use in studying the phylogenetic position of basidiomycetes among the eukaryotes

Erik Huysmans, Erna Dams, Antoon Vandenberghe and Rupert De Wachter*

Departement Celbiologie, Universiteit Antwerpen, Universiteitsplein 1, B-2610 Wilrijk, Belgium

Received 22 February 1983; Revised and Accepted 7 April 1983

ABSTRACT

The nucleotide sequences of the 5 S ribosomal RNAs of the mushrooms *Russula cyanoxantha, Pleurotus ostreatus, Agaricus edulis,* and *Auricularia auricula-judae* were determined. The sequences fit in a universal five-helix secondary structure model for 5 S RNA. As in most other 5 S RNAs, some helical areas contain non-standard base pairs. A clustering method was used to reconstruct an evolutionary tree from 82 eukaryotic 5 S RNA sequences. It allows to make a choice between alternative systematic classifications for basidiomycetes and reveals that the fungal kingdom is highly polyphyletic.

INTRODUCTION

The collection of known eukaryotic 5 S rRNA primary structures presently comprises some 80 sequences, which can be used to study macroevolutionary relationships among eukaryotic taxa. We have enriched this collection by determining the 5 S RNA sequences of 4 mushrooms belonging to the orders Agaricales and Auriculariales. As is the case for all 5 S RNAs, whether originating from eukaryotic, bacterial, or organelle ribosomes, the sequences fit in the universal five-helix secondary structure model[1]. While our work was in progress, Walker and Doolittle[2] reported the 5 S RNA sequences for 8 basidiomycetes belonging to orders different from those studied by us. To examine the evolutionary origin of the basidiomycetes, we have aligned the 12 known sequences with a set of 70 other eukaryotic 5 S RNA sequences and derived an evolutionary tree by cluster analysis.

MATERIALS AND METHODS

The mushrooms *Agaricus edulis* and *Pleurotus ostreatus* were bought in a greengrocery. *Russula cyanoxantha* and *Auricularia auricula-judae* were collected in the field. Batches of 200 g of mushrooms were worked up as previously described for mollusca[3], to yield 200 to 500 OD_{260} units of ribosomes, from which 2 to 5 OD_{260} units of 5 S RNA could be obtained by gel electro-

phoresis.

The 5 S RNA preparations were labeled with [5'-^{32}P]pCp at the 3'-terminus[4], and purified and examined for length heterogeneity by electrophoresis on an 8% polyacrylamide gel. This showed the presence of a single component, later found to have a chain length of 118 nucleotides, in *A. edulis, P. ostreatus* and *A. auricula-judae*, whereas *R. cyanoxantha* contained a 118 and a 119 nucleotide component in about equal amounts. Individual bands were extracted from the gels and the sequences determined, with the exception of the 5'-terminal pNp, by Peattie's[4] chemical degradation-gel reading method. To identify the 5'-terminal end groups and to confirm the adjacent sequences the 5 S RNAs were ligated to $(Ap)_4A$, labeled at the 5'-end, and examined by a nuclease degradation-gel reading method as described for bacterial 5 S RNAs[5]. The examination of an alkaline hydrolysate of unlabeled 5 S RNA by HPLC[6] confirmed that the 5'-termini consist of pAp and do not carry any di- or triphosphate groups.

The phenogram represented in Fig. 2 was constructed from 82 eukaryotic 5 S RNA sequences, including those reported here, now in our collection. Most of these sequences can be found in a recent review[7] where they are presented in the same alignment as used for the phenogram construction. The alignment has 150 positions because some gaps have to be introduced to accomodate all presently known sequences. Additional sequences processed in the phenogram can be found in last year's review[8], or were taken from recent papers[2,9-17]. The alignment of these additional sequences is self-evident except for 3 basidiomycete species belonging to the order Ustilaginales (see Table 1), which contain a bulge in area D' of the secondary structure model[1] and sequence alignment[7]. In these cases the G residue at position 106 from the 5'-terminus was considered as an insertion and placed at alignment position 128 which is a gap in most other sequences. The phenogram was constructed by a clustering procedure known as the weighted pair grouping method using arithmetic average, or WPGMA[18]. Clustering starts from a matrix containing association coefficients for each of the $(n^2-n)/2$ pairwise comparisons that can be made in a set of n sequences. The association coefficient for a pair of sequences was calculated as follows. Each of the 150 alignment positions was compared for the two sequences. Positions where there is a gap in both sequences, or where at least one of them shows a heterogeneity, were not taken into account. For the positions where both sequences contain a nucleotide, the number of substitutions was corrected for multiple hits and back mutation[19]. Next, the number of positions was counted where one sequence contains a gap and the

other a nucleotide, in other words the number of insertions and deletions distinguishing the two sequences. The corrected number of substitutions was then added to the number of insertions/deletions. The sum, divided by the number of positions compared, is the association coefficient. It increases with the dissimilarity between sequences. The average association coefficient between sequence groups, as calculated by the WPGMA clustering procedure and expressed as %, is the quantity defining the branch lengths and indicated on the scale in the phenogram (Fig. 2).

RESULTS AND DISCUSSION

Primary and secondary structure of the examined 5 S RNAs.

Fig. 1 shows the nucleotide sequence of *Russula cyanoxantha* 5 S RNA, folded in the universal secondary structure model[1]. There are 14 positions where at least one of the three other species examined displays a different nucleotide. These positions are indicated by arrows. Inside the arrows the nucleotide composition for *Pleurotus ostreatus*, *Auricularia auricula-judae* and *Agaricus edulis* can be read, in that order, starting from the arrowhead. Only *Russula cyanoxantha* 5 S RNA shows length heterogeneity at the 3'-terminus, with about half of the molecules terminating in U_{118} and the other half having an extra U_{119}.

By analogy with the situation in most other 5 S RNAs, we have depicted the secondary structure with $A_1 \cdot U_{118}$ base-paired although the adjacent nucleotides cannot form a standard base pair. Classical secondary structure estimation rules[20] admit only the existence of G·C, A·U and G·U pairs. According to a modified set of rules proposed by Ninio[21], non-standard pairs such as U·U may exist within helices, but can only occupy the penultimate position of a helix segment if the ultimate pair is G·C. In the basidiomycete 5 S RNAs, the shape of helix A would approach the standard model more closely if we admit that U_2 base-pairs with U_{117}, and there are many similar examples in other 5 S RNAs. Another non-standard base pair is the $A_{80} \cdot C_{96}$ found in helix E. This occurs, adjacent to a $C_{81} \cdot G_{95}$ pair, in three of the four examined species. In *Agaricus edulis*, these pairs are replaced by an $A_{80} \cdot U_{96}$ and a $U_{81} \cdot A_{95}$ pair. It looks as if two stacked A·U's can be replaced without harm by a stack consisting of a stronger G·C and a weaker A·C.

The G_{83} forming a bulge on helix E could exchange places with G_{82}, which would then form a bulge one position higher up. This possibility exists in many eukaryotic 5 S RNAs. It is conceivable that the bulge can actually jump from one position to the other and that this process confers flexibility to

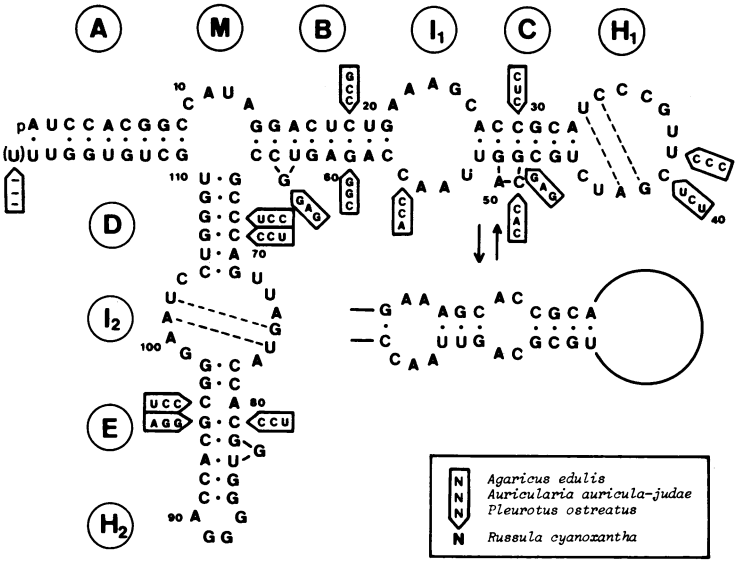

Fig. 1. Sequence and secondary structure model of the 5 S RNAs from four mushrooms.

The sequence of *Russula cyanoxantha* 5 S RNA is represented in the universal five-helix model[1]. Positions that are not identical in the four sequences examined are indicated by an arrow. Inside the arrow, the base composition for the three other species is shown in the order indicated in the figure insert. The lettering of helices (A to E), multibranched loop (M), internal loops (I_1, I_2) and hairpin loops (H_1, H_2) is as in previous papers on 5 S RNA secondary structure[1] and sequence alignment[7]. The universal alternative structure in area I_1 - C is drawn for the *R. cyanoxantha* sequence only. Adjacent base pairs connected by broken lines in loop H_1 are found in the large majority of 5 S RNAs, those in loop I_2 in nearly all eukaryotes.

helix E. We have previously described[1] a similar case of "bulge migration" in helix C of chloroplast 5 S RNAs. All other 5 S RNAs show, in area I_1 - C, the possible equilibrium between a bulge and an internal loop form indicated in Fig. 1. The potential equilibria in helices E and C may bestow on the 5 S RNA structure a mobility required in the functioning ribosome.

Molecular evolution

The publication of 5 S RNA evolutionary trees[22-34], ranging in size from 5 to 71 organisms, has kept pace with the growth of the 5 S RNA sequence collection. Some of the trees[22-26] were built using the minimal length criterion, the others[27-34] were constructed by a clustering procedure. The former method rapidly becomes prohibitive in computer time with an increasing number of species. This can only be avoided when some kind of restriction is

included in the program so that it does not examine all possible topologies[26]. One of the 5 S RNA trees is limited to vertebrate sequences[24] and one[26] is only concerned with bacteria, but all the others comprise eukaryotic and eubacterial, and in some cases also archaebacterial and organelle sequences. This raises the question as to how sequences should be aligned, since each of these sequence groups shows deletions and/or insertions with respect to the others. The trees published by Hori et al.[29-31] are based on an alignment that assumes a profound difference between eukaryotic and eubacterial 5 S RNAs, both in sequence homology and in secondary structure. We have shown[1] that an alignment with less gaps and more homology can be obtained under the assumption that 5 S RNA secondary structure is essentially universal. A more complete version of this alignment and the matching secondary structure model has been published recently[7].

Due to its conservative structure, 5 S RNA is a low resolution phylogenetic marker, yielding useful evolutionary information only on a large time-scale. When the WPGMA clustering procedure, described in the methods section, is applied to some 150 5 S RNA sequences now in our collection the resulting phenogram (unpublished) shows a eukaryotic, a eubacterial and an archaebacterial branch. The phenogram represented in Fig. 2 resulted when the subset of 82 eukaryotic 5 S RNA sequences was clustered. In contrast to metazoa and green plants, which form homogeneous clusters, the fungi and protista are each represented by more than one branch, indicating the polyphyletic character of the latter groups. In the dendrograms published by Hori et al.[29-31] the number of protista and fungi is much smaller and the latter all belong to the Ascomycotina. This gives the impression of a more homogeneous descent for protista and fungi, although the tree topology for the species examined is not basically different from that shown in Fig. 2. The most recent dendrogram of Küntzel et al.[34] contains about half the number of eukaryotic sequences included in Fig. 2. In this case too, the protista and fungi each ramify from a single branch. This is due, however, to the method of tree construction[32]. Sequences are first classified on the basis of "established phylogenetic affinities" as metazoa, protozoa, mycophyta or plants. Subtrees are constructed for each of these taxa, and only subsequently linked into a general tree. No such constraints exist in the algorithm that we used to construct the tree represented in Fig. 2.

Another point requiring comment is the unexpected position of some species within the metazoan branch. Examples are the crustacean *Artemia salina*, which clusters far from the four insects, and the sponge *Haliclona oculata*, which

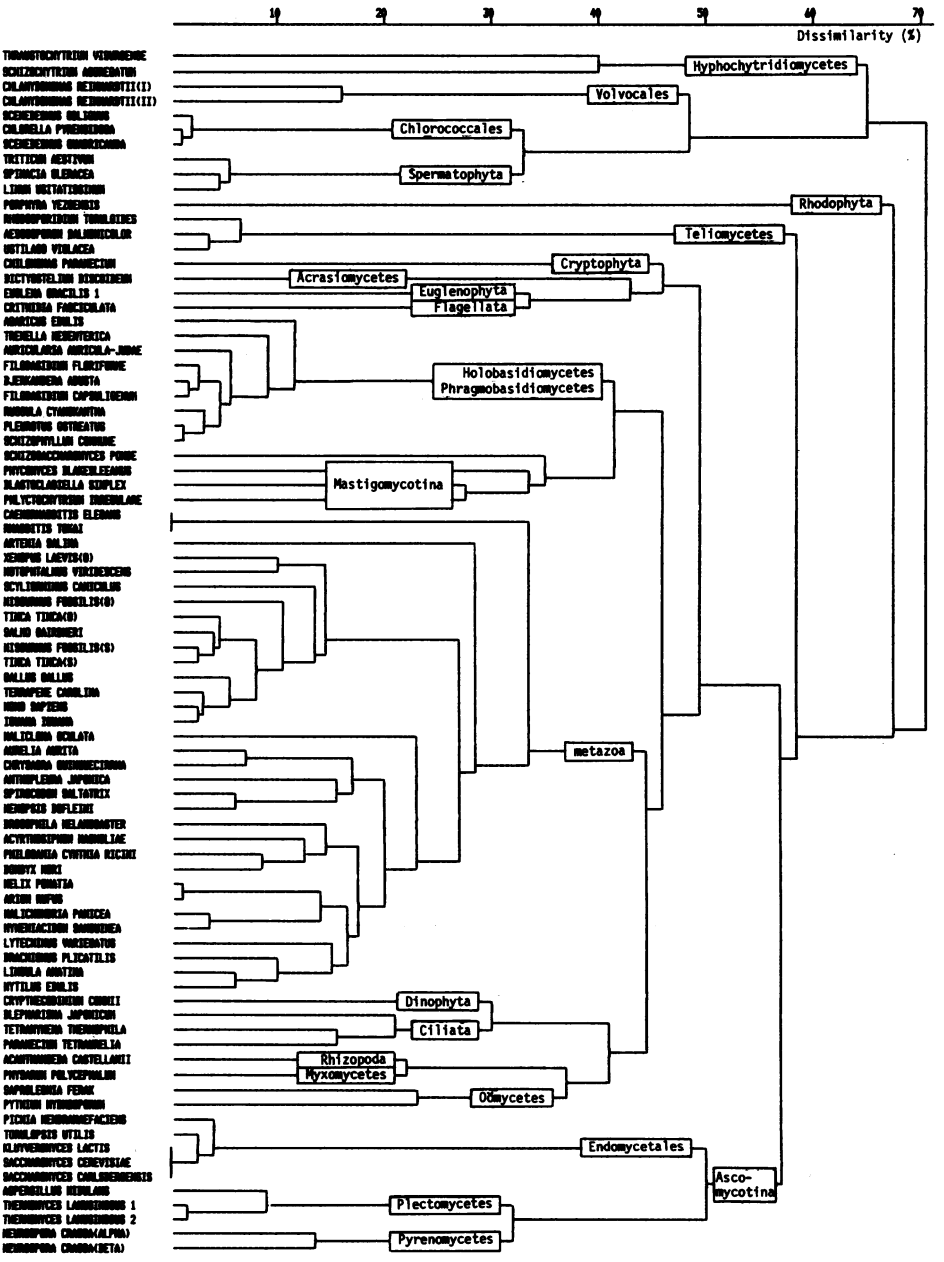

Fig. 2. Phenogram constructed from 82 eukaryotic 5 S RNA sequences. In naming the different lineages we did not take into account the systematic position of form-taxa.

Table 1. Basidiomycetes included in the evolutionary tree

Species[a]	Order	Class	
		according to Talbot[37]	according to Patouillard[35]
Ustilago violacea *Aessosporon salmonicolor*[b] *Rhodosporidium toruloides*[b]	Ustilaginales	Teliomycetes	Hetero-basidiomycetes
Auricularia auricula-judae★	Auriculariales	Phragmo-basidiomycetes	
Tremella mesenterica	Tremellales		
Bjerkandera adusta *Schizophyllum commune*	Aphyllophorales	Holo-basidiomycetes	Homo-basidiomycetes
Agaricus edulis★ *Russula cyanoxantha*★ *Pleurotus ostreatus*★	Agaricales		
Filobasidium capsuligenum *Filobasidium floriforme*	basidiomycetous yeasts		

(a) The 5 S RNA sequences of the species identified by an asterisk were determined by us, the others by Walker and Doolittle[2].

(b) The assignment of *Aessosporon* to the order Ustilaginales is according to van der Walt[39], that of *Rhodosporidium* according to Banno[40].

is rather distant from the other porifera *Halichondria panicea* and *Hymeniacidon sanguinea*. These species are not included in previously published trees, hence a comparison is not possible. As mentioned above, 5 S RNA is not a very useful molecule to study relationships among relatively recently diverged organisms such as the metazoa. Since mutation is a random process and the information content of 5 S RNA is limited, a few such discrepancies have to be expected. It is not excluded, however, that the incorporation of additional invertebrate sequences improves the local topology of the tree in the future. The fact that the echinoderm *Lytechinus variegatus* clusters with the protostomian metazoa and not with the vertebrates has been previously observed and

discussed[32].

The 12 basidiomycetous fungi included in our study and their systematic position are listed in Table 1. The division in Heterobasidiomycetes and Homobasidiomycetes[35] which is still frequently made, and the phylogenetic schemes based on it[36], are incompatible with the topology shown in Fig. 2. A division in Teliomycetes, Phragmobasidiomycetes and Holobasidiomycetes[37] seems more reasonable. According to our phenogram, and also on the basis of ultrastructural data[38], the genus *Filobasidium* belongs to the Holobasidiomycetes or the Phragmobasidiomycetes. The latter two classes seem to have diverged too recently to be resolved on the basis of 5 S RNA sequencing data.

A surprising feature of the phenogram is the large evolutionary distance between Teliomycetes on the one hand and Holo- and Phragmobasidiomycetes on the other. The lineage leading to the Teliomycetes originates at approximately the same time as that leading to the Ascomycotina. These divergences can be estimated, by comparison with the metazoan radiation, to have taken place approximately 1200 million years ago. In contrast, the Holo- and Phragmobasidiomycetes seem to be related to the Mastigomycotina, which diverged from the metazoan lineage more recently. If we presume a monophyletic origin for the Basidiomycotina, we would expect the Holo- and Phragmobasidiomycetes to cluster with the Teliomycetes. The similarity between the former two classes and the Mastigomycotina may be due to convergence. Another possibility is that the divergence time of the Teliomycetes is overestimated in the phenogram. The examination of sequences from taxa with characters intermediate between those of ascomycetous and basidiomycetous fungi, such as Taphrinales and Uredinales, may help to clarify this point in the future.

ACKNOWLEDGEMENTS

We thank Raymond De Baere for assistance in the preparation of ribosomes and 5 S RNA. Willi Banwarth, Martin Cunningham, Elisabeth Hoey and Philip Lehrbach participated in the sequencing of *Pleurotus ostreatus* 5 S RNA on the occasion of an EMBO workshop on nucleic acid sequencing techniques at Leuven. Erna Dams holds an I.W.O.N.L. scholarship. Our research was supported in part by F.K.F.O. grant nr. 2.0017.80.

*To whom correspondence should be addressed

REFERENCES

1. De Wachter, R., Chen, M.-W., Vandenberghe, A. (1982) Biochimie 64, 311-329.

2. Walker, W.F., Doolittle, W.F. (1982) Nature 299, 723-724.
3. Fang, B.-L., De Baere, R., Vandenberghe, A., De Wachter, R. (1982) Nucleic Acids Res. 10, 4679-4685.
4. Peattie, D.A. (1979) Proc. Natl. Acad. Sci. USA 76, 1760-1764.
5. Dams, E., Vandenberghe, A., De Wachter, R. (1983) Nucleic Acids Res. 11, 1245-1252.
6. Vandenberghe, A., De Wachter, R. (1982) J. Liq. Chromatogr. 5, 2079-2084.
7. Erdmann, V.A., Huysmans, E., Vandenberghe, A., De Wachter, R. (1983) Nucleic Acids Res. 11, r105-r133.
8. Erdmann, V.A. (1982) Nucleic Acids Res. 10, r93-r115.
9. Denis, H., Wegnez, M. (1977) Develop. Biol. 59, 228-236.
10. Wegnez, M., Denis, H., Mazabraud, A., Clerot, J.C. (1978) Develop. Biol. 62, 99-111.
11. Wildeman, A.G., Nazar, R.N. (1982) J. Biol. Chem. 257, 11395-11404.
12. Kumazaki, T., Hori, H., Osawa, S. (1982) FEBS Lett. 146, 307-310.
13. Kumazaki, T., Hori, H., Osawa, S. (1982) FEBS Lett. 149, 281-284.
14. Goldsbrough, P.B., Ellis, T.H.N., Lomonossoff, G.P. (1982) Nucleic Acids Res. 10, 4501-4514.
15. Kumazaki, T., Hori, H., Osawa, S., Ishii, N., Suzuki, K. (1982) Nucleic Acids Res. 10, 7001-7004.
16. Hori, H., Ohama, T., Kumazaki, T., Osawa, S. (1982) Nucleic Acids Res. 10, 7405-7408.
17. McKay, R.M., Doolittle, W.F. (1982) Nucleic Acids Res. 10, 8307-8310.
18. Sokal, R.R., Michener, C.D. (1958) Univ. Kansas Sci. Bull. 38, 1409-1438.
19. Kimura, M., Ohta, T. (1972) J. Mol. Evol. 2, 87-90.
20. Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M., Gralla, J. (1973) Nature New Biol. 246, 40-41.
21. Ninio, J. (1979) Biochimie 61, 1133-1150.
22. Sankoff, D., Morel, C., Cedergren, R.J. (1973) Nature New Biol. 245, 232-234.
23. Schwartz, R.M., Dayhoff, M.O. (1978) in Atlas of protein sequence, Dayhoff, M.O., Ed., Vol. 5, suppl. 3, pp. 327-337, National Biomedical Research Foundation, Silver Spring, Maryland.
24. Denis, H., Wegnez, M. (1978) J. Mol. Evol. 12, 11-15.
25. Hinnebusch, A.G., Klotz, L.C., Blanken, R.L., Loeblich, A.R. (1981) J. Mol. Evol. 17, 334-347.
26. Sankoff, D., Cedergren, R.J., McKay, W. (1982) Nucleic Acids Res. 10, 421-431.
27. Hori, H. (1975) J. Mol. Evol. 7, 75-86.
28. Hori, H. (1976) Mol. Gen. Genet. 145, 119-123.
29. Hori, H., Osawa, S. (1979) Proc. Natl. Acad. Sci. USA 76, 381-385.
30. Hori, H., Osawa, S. (1980) in Genetics and evolution of RNA polymerase, tRNA and ribosomes, Osawa, S., Ozeki, H., Uchida, H. and Yura, T., eds., pp. 539-551, Elsevier/North Holland, Amsterdam.
31. Hori, H., Itoh, T., Osawa, S. (1982) Zbl. Bakt. Hyg., I. Abt. Orig. C 3, 18-30.
32. Küntzel, H., Heidrich, M., Piechulla, B. (1981) Nucleic Acids Res. 9, 1451-1461.
33. Küntzel, H. (1982) Zbl. Bakt. Hyg., I Abt. Orig. C 3, 31-39.
34. Küntzel, H., Piechulla, B., Hahn, U. (1983) Nucleic Acids Res. 11, 893-900.
35. Patouillard, N. in Essai taxonomique sur les familles et les genres des Hyménomycètes, Declume, Lons-le-Saunier, France, 1900.
36. Savile, D.B.O. (1955) Can. J. Bot. 33, 60-104.
37. Talbot, P.H.B. (1968) Taxon 17, 620-628.
38. Kreger-van Rij, N.J.W. in The fungi, an advanced treatise, Ainsworth,

G.C., Sparrow, F.K. and Sussman, A.S. eds., Vol. IV A, pp. 11-32, Academic Press, New York, 1973.
39. van der Walt, J.P. (1970) J. Microbiol. Serol. 36, 49-55.
40. Banno, I. (1981) J. Gen. Appl. Microbiol. 13, 167-196.