456–461 Nucleic Acids Research, 1998, Vol. 26, No. 2

Three different group I introns in the nuclear large subunit ribosomal DNA of the amoeboflagellate *Naegleria*

Johan F. De Jonckheere* and Susan Brown[1]

Protozoology Laboratory, Scientific Institute of Public Health-Louis Pasteur, B-1050 Brussels, Belgium and [1]Culture Collection of Algae and Protozoa, Institute of Freshwater Ecology, Ambleside, Cumbria LA22 OLP, UK

Received October 3, 1997; Revised and Accepted November 25, 1997 DDBJ/EMBL/GenBank accession nos. AJ001314–AJ001316, AJ001399

ABSTRACT

We have amplified the large subunit ribosomal DNA (LSUrDNA) of the 12 described *Naegleria* spp. and of 34 other *Naegleria* lineages that might be distinct species. Two strains yielded a product that is longer than 3 kb, which is the length of the LSUrDNA of all described *Naegleria* spp. Sequencing data revealed that the insert in one of these strains is a group I intron without an open reading frame (ORF), while the other strain contains two different group I introns, of which the second intron has an ORF of 175 amino acids. In the latter ORF there is a conserved His-Cys box, as in the homing endonucleases present in group I introns in the small subunit ribosomal DNA (SSUrDNA) of *Naegleria* spp. Although the group I introns in the LSUrDNA differ in sequence, they are more related to each other than they are to the group I introns in the SSUrDNA of *Naegleria* spp. The three group I introns in the LSUrDNA in *Naegleria* are at different locations and are probably acquired by horizontal transfer, contrary to the SSUrDNA group I introns in this genus which are of ancestral origin and are transmitted vertically.

INTRODUCTION

The genus *Naegleria* consists of free-living amoebae that can differentiate into flagellates. One species, *Naegleria fowleri*, is highly pathogenic for man, causing primary amoebic meningo-encephalitis which almost invariably leads to death. Two other species, *Naegleria australiensis* and *Naegleria italica*, kill experimental animals but have never been identified as the cause of disease in humans (1). The ribosomal DNA of *Naegleria*, and of all other genera that belong to the family Vahlkampfiidae, is on a circular plasmid (2). In five of the 12 described *Naegleria* spp., and in 11 other *Naegleria* lineages, a 1.3 kb group I intron is present in the small subunit ribosomal DNA (SSUrDNA) (3, Brown and De Jonckheere, unpublished). This group I intron is carried at the same location in loop 20/21 of the SSUrDNA (4) of all these lineages. It contains an open reading frame (ORF) coding for 245 amino acids (5). De Jonckheere (3) demonstrated that two different sets of conserved sequences (P, Q, R, S) are present in these group I introns, which allowed him to construct three different possibilities for the secondary structure. Only recently, it was demonstrated experimentally that these two sets of conserved elements belong to different introns, forming a twin-ribozyme intron (6). The ORF is thought to be a putative homing endonuclease (7) but there is no evidence for horizontal transfer of this group I intron in the genus *Naegleria*. On the contrary, sequence data support the hypothesis that the intron is transmitted vertically and was lost from the majority of *Naegleria* spp. (5). Re-analysing the data, with the two group I introns of the twin-ribozyme introns treated separately, yielded the same conclusion (6). One *Naegleria* lineage was detected in which the group I intron had lost the ORF (8). Re-evaluation of the data in the light of the twin-ribozyme intron indicates that this lineage had actually lost the ORF-containing group I intron (6). We have now amplified the large subunit ribosomal DNA (LSUrDNA) of the 12 described *Naegleria* spp. and of 34 *Naegleria* lineages that might be distinct species. Two strains yielded a product that is longer than 3 kb, which is the length of the LSUrDNA of all described *Naegleria* spp., suggesting these two strains have a group I intron in the LSUrDNA. We report here the sequencing results of these inserts, which demonstrate that there are three different group I introns in the LSUrDNA of *Naegleria*, of which one contains an ORF with a His-Cys box.

MATERIALS AND METHODS

Strains representing putative new *Naegleria* species (9, B.S. Robinson, unpublished) were cultured with *Escherichia coli*. When the agar plates were totally covered with amoebae the cells were harvested and concentrated by centrifugation. The DNA was extracted using a guanidinium thiocyanate-Sarkosyl method (10). For the *Naegleria* type species samples from DNA prepared in previous studies were used. The LSUrDNA was amplified using conserved primers at the 5′-end and the 3′-end of the LSUrDNA: forward primer 5′-ATATTAATAAGGGGAGGAAA and reverse primer 5′-AGGGTAAAACTAACCTGTCT. Amplification conditions were 1 min at 94°C, 1 min at 50°C and 2 min at 72°C for 30 cycles, with 10 min at 72°C at the end. Products were visualised on 0.7% agarose gels. To assess sequence homology between SSUrDNA and LSUrDNA group I introns, PCR products

*To whom correspondence should be addressed. Tel: +32 2 6425054; Fax +32 2 6425253; Email: jdjonckh@ben.vub.ac.be

of the LSUrDNA from two strains (NG236 and NG872) that yielded a PCR product longer than 3 kb were blotted onto Nylon filters (Hybond, Amersham Life Science Inc., Cleveland, Ohio, USA) from agarose gels. Also transferred were the amplified SSUrDNA of *Naegleria clarki*, strain NG597 and *N.fowleri*, and the total DNA of the same strains. These blots were hybridized at low stringency with either the PCR amplified intron, or ORF of the intron, of the SSUrDNA of *N.clarki* labeled with ^{32}P. The blots were autoradiographed for 3 h with intensifying screens. The PCR products used for sequencing were purified using the enzymes supplied with the Sequenase PCR product sequencing kit (Amersham Life Science Inc., Cleveland, Ohio, USA). The enzyme exonuclease I removes residual single-stranded primers and any extraneous single stranded DNA produced by the PCR. The enzyme shrimp alkaline phosphatase removes the remaining dNTPs from the PCR mixture. After purification the PCR products were sequenced using internal primers corresponding to sequences that were surrounding the introns and internal primers which were designed as the intron sequences became known. The locations of the group I introns in the LSUrDNA were determined by restriction frament length comparison of strains with and without the introns. Sequencing reaction products were separated on 6% acrylamide-urea sequencing gels and autoradiographed overnight at room temperature. Sequences were aligned by eye using the Eyeball Sequence Editor (ESEE) (10). The nucleotide sequence data reported in this paper are in the EMBL, GenBank and DDBJ nucleotide sequence databases under the accession nos. AJ001314–AJ001316 and AJ001399.

RESULTS

Two strains (NG236 and NG872) yielded a LSUrDNA amplification product longer than the 3 kb product amplified from all described *Naegleria* spp. The LSUrDNA length of 3.4 kb in strain NG872 was suggestive of a group I intron without an ORF, while in strain NG236 a length of 4.1 kb suggested a group I intron with an ORF. Both strains are high temperature tolerant and, based on SSUrDNA sequences, are situated on the *N.fowleri–Naegleria lovaniensis* branch of the phylogenetic tree (12).

In blots hybridized with the group I intron from *N.clarki* SSUrDNA as a probe the SSUrDNA PCR product of *N.clarki* and *Naegleria* spp. NG597 lit up as well as the total DNA of both strains (not shown). Neither the LSUrDNA PCR product of NG236 and NG872, nor the SSUrDNA PCR product of *N.fowleri* lit up. In similar blots hybridized with the PCR amplified ORF from the *N.clarki* SSUrDNA group I intron, only the SSUrDNA product and the total DNA of the latter species lit up. This gave a strong indication that if group I introns are present in the LSUrDNA of NG236 and NG872, their sequence must be quite different from that of the group I introns found in the SSUrDNA of *Naegleria* spp.

Table 1. Positions of group I introns in nuclear LSUrDNA ([a]with ORF)

Helix	Flanking sequence	Organism	Ref.
D6′	GAACCAT ⇓ CTAGTA	*C.ellipsoidea*	29
		H.ericae	28
E26′	GACT ⇓ CTCTTAAGG	*P.carinii*	22
		B.brongniartii (LSU1)	19
		G.graminis (LSU1)	20
		Naegleria spp. NG872	this paper
	GACTCT ⇓ CTTAAGG	*C.albicans*	25
		C.dubliensis	26
		C.stellatoidea	26
	GACTCTCT ⇓ TAAGG	*P.polycephalum* (LSU3[a])	23
		Tetrahymena spp.	24
	GACTCTCTTAA ⇓ GG	*R.elatiana*	27
E28	TCGTCAT ⇓ TTAATT	*P.polycephalum* (LSU1)	15
		P.flavicomum (LSU1)	16
		D.iridis (LSU1)	17
		Naegleria sp. NG236 (LSU1)	this paper
G2	GACCCT ⇓ GTTGAGC	*B.brongniartii* (LSU2)	19
G2′	GGGAT ⇓ AACTGGCT	*P.polycephalum* (LSU2)	
		P.flavicomum (LSU2)	
		D.iridis (LSU2)	
		A.adeninovorans	18
		B.brongniartii (LSU3)	19
		G.graminis (LSU2)	20
G19′	CCCACT ⇓ AATAGGG	*B.brongniartii* (LSU4)	19
		G.graminis (LSU3)	20
		Naegleria spp. NG236 (LSU2[a])	this paper

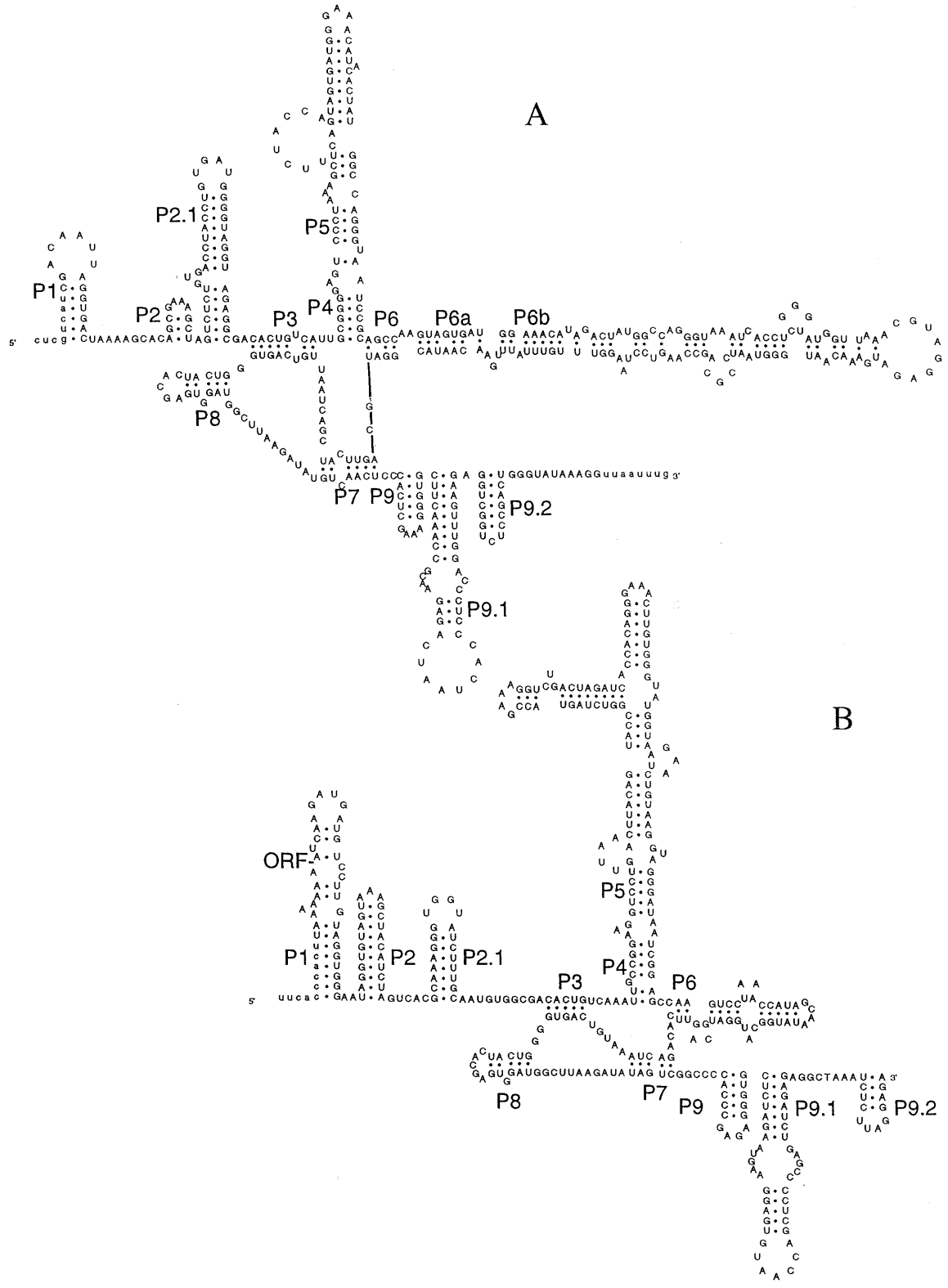

Figure 1. (Above and overleaf) Secondary structure models for the group I introns in the LSUrDNA of *Naegleria* spp. (**A**) First intron of strain NG236, (**B**) second intron of strain 236 (**C**), intron of strain NG872. Lower case letters are the LSUrDNA at the 5′- and 3′-end and upper case letters are the intron.

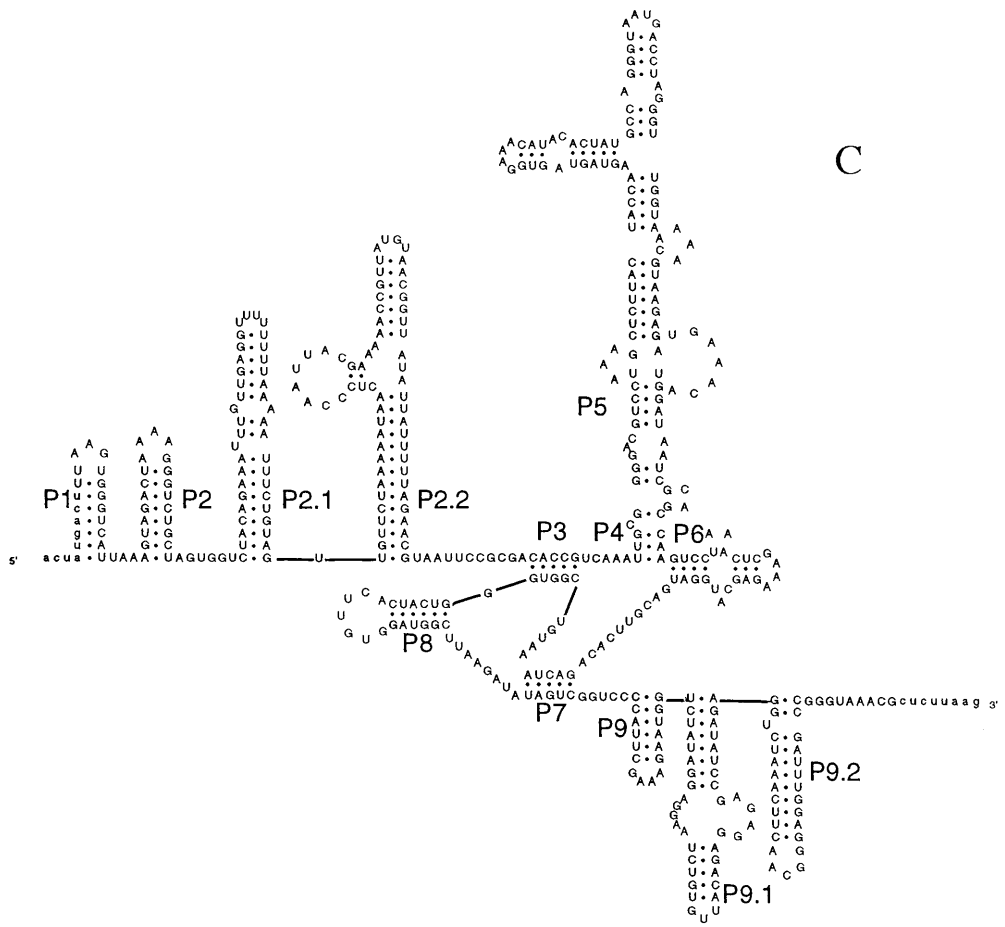

Figure 1. *continued*

The sequencing results confirm that the inserts in the LSUrDNA are group I introns and that they are quite different in sequence from the group I introns in the SSUrDNA of *Naegleria* spp. The LSUrDNA intron in strain NG872 is located in stem–loop E26 (Table 1) of the secondary structure of the LSUrDNA as proposed by De Rijck *et al.* (13). The sequencing results of the LSUrDNA of strain NG236 showed that it does not contain one group I intron with an ORF, but two different group I introns instead. The first intron is located in stem–loop E28 (Table 1), which is 28 nt to the 3′-end of the site where the group I intron of strain NG872 is found. The second intron is located in G19′ (Table 1). The group I introns in the LSUrDNA differ in length and sequence but are more related to each other than they are to the group I introns in the SSUrDNA of *Naegleria* spp. The intron in strain NG872 is 474 nt long and differs from the first intron in strain NG236, which is 389 nt long, especially by the presence of an insert producing an additional stem–loop P2.2 in the secondary structure (Fig. 1). The second intron in the LSU of NG236 is 919 nt long and has an ORF for 175 amino acids. The ORF in the second intron in the LSUrDNA of NG236 is not located in P6, as in the SSUrDNA group I introns of *Naegleria* spp., but in P1 instead (Fig. 1). Although this ORF is much shorter than the ORF in the group I introns of the SSUrDNA of *Naegleria* spp., it does have the conserved His-Cys box (Table 2) which is present in the latter. This His-Cys box in the ORF in the LSUrDNA group intron is less conserved than it is in the SSUrDNA. It differs as much from the latter as from the His-Cys box found in the myxomycetes *Physarum polycephalum* and *Didymium iridis*. In addition the His-Cys box in the LSUrDNA intron of strain NG236 is missing the third conserved Cysteine. Contrary to NG236, NG872 also has a group I intron in the SSUrDNA, and this group I intron has an ORF for 245 amino acids. Alignments show that the ORF of the group I intron that is found in the SSUrDNA of NG872 is related to the ORF in the SSU group I introns of described *Naegleria* spp. However, there is a major difference in the His-Cys box, in which there is an insertion of serine with a compensating deletion of asparagine three amino acids away (Table 2).

DISCUSSION

In the comparative database for group I introns (14) a total of 48 introns are reported in the nuclear SSUrDNA and 22 in the nuclear LSUrDNA. This difference in number of group I introns in the two molecules could be due to under-reporting in the LSUrDNA as the latter database is much smaller than the one for the SSUrDNA. However, in the genus *Naegleria*, group I introns (with or without ORF) are present in the nuclear SSUrDNA of several *Naegleria* lineages, while the presence of a group I intron in the LSUrDNA of *Naegleria* seems to be quite uncommon (Table 3). Of 12 *Naegleria* spp. and 34 lineages that could be regarded as distinct species, only two carry group I introns in the

LSUrDNA. One of the strains carries two different group I introns. The introns differ in length and are located in different places in the LSUrDNA. This is in contrast with group I introns in the SSUrDNA of *Naegleria* which are all located in the same site, whether they contain an ORF or not (5,8). The first and second LSUrDNA group I introns of strain NG236 are in location E28 and in stem–loop G19, respectively, of the LSUrDNA. The former is in the same location as the first group I intron of the myxomycetes *P.polycephalum* (15), *P.flavicomum* (16) and *D.iridis* (17) (Table 1). These three myxomycetes all have a second group I intron at the location of a group I intron found in the yeast *Arxula adeninivorans* (18), in the fungi *Beauveria brongniartii* (19) and *Gaeumannomyces graminis* (20). The latter two fungi have also a group I intron at the same location as the second group I intron of NG236. It is interesting to note that the LSUrDNA group I intron ORF at site E28 in *P.polycephalum* (15) and in *D.iridis* (17) and in the SSUrDNA group I introns of *Naegleria* (6), have the conserved His-Cys box as the second LSUrDNA group I intron of NG236. However, this His-Cys box is more conserved within the different *Naegleria* SSUrDNA ORFs than it is in the LSUrDNA ORF of NG236. The His-Cys box in NG236 is as different from the former as those in the two myxomycetes and in the red alga *Porhyra spiralis*. In the latter rhodophyte an ORF is present on the complementary strand (21) of a group I intron in the SSUrDNA. It was only reported later by Vader *et al*., (16) that this ORF also contains the His-Cys box. The LSUrDNA group I intron of strain NG872 is at the same location in E26 as the first group I intron in *Pneumocystis carinii* (22) and this is also where the two fungi *B.brongniartii* (19) and *G.graminis* (20) have an additional group I intron. Group I introns are found at different locations in E26′ in other organisms as well: the third group I intron of *P.polycephalum* (23) is located 4 nt further to the 3′-end, as are all group I introns of *Tetrahymena* spp. (24); the group I introns of *Candida albicans* (25), *Candida dubliniensis* and *Candida stellatoidea* (26) are located 2 nt further to the 3′-end; the group I intron of *Rotaliella elatiana* (27) is located 7 nt further to the 3′-end (Table 1). The only other location of group I introns in nuclear LSUrDNA is D6′, as found in *Hymenoscyphus ericae* (28) and in *Chlorella ellipsoidea* (29).

Table 2. Amino acid sequence of the His-Cys box in the ORF in group I introns

N.andersoni	SSU	TISHLC-GNGGCARPGH-LRIE-KKTVNDERTHCH
N.jamiesoni	SSU	......-..........-....-..S.........
N.gruberi	SSU	......-.....S....-....-..S.........
N.italica	SSU	......-..........-....-..S........L
N.clarki	SSU	......-..........-....-..S.........
Naegleria spp.NG872	SSU	......-.....C....-....S...-........
Naegleria spp.NG236	LSU	V.R.T.-.CKD.CN.E.-.KLG-T.SD.EYDKGI.
P.polycephalum	LSU	.A....-H.TR.HN.L.-.CW.SLDD-.KG.NW.P
D.iridis	SSU	HS....K.D.S.MELK.T..VP-AQ.NLADHEL.P
P.spiralis	SSU	EA..T.-H.AK.VNKA.-.TL.SGDL-.KS.IY.R

Table 3. Summary of the presence and length (in nt) of group I introns in the nuclear rDNA of the genus *Naegleria*

Species	SSU rDNA (refs. 5 and 7)	LSU rDNA (this paper)
N.fowleri	–	–
N.lovaniensis	–	–
Naegleria spp. NG872	1318 (+ ORF)	474 (– ORF)
Naegleria spp. NG236	–	389[b] (– ORF)
		919[b](+ ORF)
N.jadini	–	–
N.australiensis	–	–
N.italica	1319 (+ ORF)	–
N.andersoni	1309 (+ ORF)	–
N.jamiesoni	1307 (+ ORF)	–
N.gruberi	1316[a] (+ ORF)	–
(species complex)	–	
N.clarki	1305 (+ ORF)	–
Naegleria spp. NG434, NG650, NG597	375 (– ORF)	–
N.galeacystis	–	–
N.minor	–	–
N.pussardi	–	–

[a]Only one out of four clusters in *N.gruberi* complex contains a group I intron.
[b]Different group I introns at different locations in the same strain.

The group I intron in *Naegleria* SSUrDNA was demonstrated to be of ancestral origin and to have been lost from the majority of *Naegleria* lineages (5). In contrast, based on sequence differences, the different locations and the low incidence in the different species, we deduce that the LSUrDNA group I introns in *Naegleria* spp. must have been acquired by horizontal transfer on separate occasions. The sequences of the SSUrDNA and LSUrDNA group I introns in *Naegleria* spp. do not seem to be closely related to each other and, therefore, are probably not due to a transposition event from the SSUrDNA to the LSUrDNA. This is probably in contrast to *P.carinii* in which the LSUrDNA group I intron is closely related to the one located in its SSUrDNA (22).

In the genus *Naegleria*, we observe the unique combination of vertical transmission of the group I introns in the SSUrDNA and of horizontal transmission in the LSUrDNA. According to insertion site, the first NG236 LSUrDNA intron is related to one of the group I introns in the LSUrDNA of the slime molds *P.polycepalum* and *D.iridis*, while the NG872 LSUrDNA intron is related to the group I intron in the LSUrDNA of *P.carinii* and the first intron in both *B.brongniartii* and *G.graminis*. However, there seems to be no sequence homology, although they are located at the same position in the LSUrDNA. Therefore, all introns that occur at a particular site of the LSUrDNA are not necessarily closely related.

The LSU3 intron in *P.polycephalum* and the second LSUrDNA intron in *Naegleria* strain NG236 are the only introns reported in the LSUrDNA with an ORF. The intron with an ORF in *P.polycephalum* has been found in only one particular strain, as is the case in the genus *Naegleria*. ORFs in nuclear SSUrDNA are found in *Naegleria* (5) and *Didymium* (30). In both cases the ORF belongs to twin-ribozyme introns (6,30), but this seems not to be the case in the ORF of the LSU rDNA in *Naegleria* and the myxomycete *P.polycephalum*.

ACKNOWLEDGEMENTS

We would like to thank Bret Robinson (Australian Centre for Water Quality Research) for providing the strains NG236 and NG872.

REFERENCES

1 De Jonckheere, J.F. (1994) *Eur. J. Protistol*., **30**, 333–341.
2 Clark, C.G. and Cross, A.M. (1988) *J. Protozool*., **35**, 326–329.
3 De Jonckheere, J.F. (1993) *J. Euk. Microbiol*., **40**, 179–187.
4 Van de Peer, Y., Van den Broeck, I., De Rijck, P. and De Wachter, R. (1994) *Nucleic Acids Res*., **22**, 3488–3494.
5 De Jonckheere, J.F. (1994) *J. Euk. Microbiol*., **41**, 457–463.
6 Einvick, C., Decatur, W.A., Embley, T.M., Vogt, V.M. and Johansen, S. (1997) *RNA*, **3**, 710–720.
7 Johansen, S., Embley, T.M. and Willassen, N.P. (1993) *Nucleic Acids Res*., **21**, 4405.
8 De Jonckheere, J.F. and Brown, S. (1994) *Nucleic Acids Res*., **22**, 3925–3927.
9 Robinson, B.S., Christy, P., Hayes, S.J. and Dobson, P.J. (1992) *J. Protozool*., **39**, 702–712.
10 Pitcher, D.G., Sanders, N.A. and Owen, R.J. (1989) *Lett. Appl. Microbiol*., **8**, 151–156.
11 Cabot, E.L. and Beckenbach, A.T. (1989) *Comput. Appl. Biosci*., **5**, 233–234.
12 De Jonckheere, J.F. and Brown, S. (1997) *Acta Protozool*., **36**, 273–278.
13 De Rijck, P., Van de Peer, Y., Chapelle, S. and De Wachter, R. (1994) *Nucleic Acids Res*., **22**, 3495–3501.
14 Damberger, S. H., and Gutell, R. R. (1994) *Nucleic Acids Res*., **22**, 3508–3510.
15 Ruoff, B., Johansen, S. and Vogt, V. M. (1992) *Nucleic Acids Res*., **20**, 5899–5906.
16 Vader, A., Naess, J., Haugli, K., Haugli, F and Johansen, S. (1994) *Nucleic Acids Res*., **22**, 4553–4559.
17 Johansen, S., Johansen, T. and Haugli, F. (1992) *Curr. Genet*. **22**, 297–304.
18 Rosel, H. and Kunze, G. (1996) *Yeast*, **12**, 1201–1208.
19 Neuvéglise, C., Brygoo, Y. and Riba, G. (1997) *Mol. Ecol*., **6**, 373–381.
20 Tan, M.K. (1997) *J. Mol. Evol*., **44**, 637–645.
21 Oliviera, M. C. and Ragan, M. A. (1994) *Mol. Biol. Evol*., **11**, 195–207.
22 Liu, Y., Rocourt, M., Pan, S., Liu, C. and Leibowitz, M. J. (1992) *Nucleic Acids Res*., **20**, 3763–3772.
23 Nomiyama, H., Kuhara, S., Kukita, T., Otsuka, T. and Sakaki, Y. (1981) *Nucleic Acids Res*., **9**, 5507–5520.
24 Nielsen, H. and Engberg, J. (1985) *Nucleic Acids Res*., **13**, 7445–7455.
25 Mercure, S., Montplaisir, S. and Lemay, G. (1993) *Nucleic Acids Res*., **21**, 6020–6027.
26 Boucher, H., Mercure, S., Montplaisir, S. and Lemay, G. (1996) *Gene*, **180**, 189–196.
27 Pawlowski, J., Bolivar, I., Fahrni, J. and Zaninetti, L. (1994) *Micropaleontology*, **40**, 373–377.
28 Egger, K. N., Osmond, G. and Goodier, J.L. (1995) *Biochim. Biophys. Acta*, **1261**, 275–278.
29 Tadanori, A., Yamada, T., Yamashita, M. and Morooka, Y. (1994) *Gene*, **145**, 139–144.
30 Decatur, W.A., Einvik, C., Johanson, S. and Vogt, V.M. (1995) *EMBO J*., **14,** 4558–4658.