Available online at www.sciencedirect.com

ScienceDirect

European Journal of Protistology 66 (2018) 26–35

European Journal of PROTISTOLOGY

www.elsevier.com/locate/ejop

Nuclear Group I introns with homing endonuclease genes in *Acanthamoeba* genotype T4

Daniele Corsaro*, Danielle Venditti

CHLAREAS, 12, rue du Maconnais, F-54500 Vandoeuvre-lès-Nancy, France

Received 12 June 2018; received in revised form 12 July 2018; accepted 12 July 2018
Available online 18 July 2018

Abstract

Various strains belonging to three *Acanthamoeba* species, *A. griffini* (genotype T3), *A. lenticulata* (T5), and *A. jacobsi* (T15), have group I introns in their 18S rRNA genes. Group I introns are self-splicing ribozymes that can spread among host lineages either through an intron-encoded endonuclease at the DNA level, or by reverse splicing during the RNA cycle. In *Acanthamoeba*, introns belong to the subclass IC1, they are located at one out four positions within the rRNA, show low identity values and all lack open reading frames to encode for an endonuclease. Uncharacterized introns from strains of another genotype, T4 (*A. castellanii* complex), resemble those of genotype T3, and at least one of them contains a non-functional endonuclease gene. Here, we analyzed all available data on *Acanthamoeba* 18S rDNA sequences to identify the possible presence of open reading frames that could encode endonucleases. We found a total of eight 18S rDNA sequences, all from T4 strains, that have introns containing putative non-functional endonuclease genes. Furthermore, two distinct endonucleases can be identified that are differently inserted in unrelated introns.

Keywords: *Acanthamoeba*; Group I intron; HEG; Homing endonuclease; Nuclear IC1 intron

Introduction

Acanthamoeba spp. (Amoebozoa: Centramoebida) are ubiquitous free-living amoebae, potentially pathogenic for humans and other vertebrates, either directly by causing various diseases or by transmitting microbial pathogens they can host. As some incongruences between morphological species and genetic data exist, strains are currently clustered into genotypes defined on the basis of the full sequence of the nuclear 18S rRNA gene (Corsaro et al. 2015, 2017; Gast et al. 1996; Stothard et al. 1998). By studying the full 18S rDNA of *Acanthamoeba*, Group I introns were discovered and characterized in strains belonging to three species, *A. griffini* (genotype T3), *A. lenticulata* (genotype T5) and *A. jacobsi* (genotype T15) (Corsaro et al. 2017; Gast et al. 1994; Schroeder-Diedrich et al. 1998).

Group I introns are mobile genetic elements scattered through a large variety of viral, prokaryotic, and nuclear and organellar (i.e., plastid and mitochondrial) genomes of eukaryotes, where they interrupt RNA- and protein-coding genes (Haugen et al. 2005; Hedberg and Johansen 2013). Sequence and structural variations of the conserved folded nine base-paired regions (P1 to P9) subdivide group I introns in five major classes (IA, IB, IC, ID and IE) and subclasses (Li and Zhang 2005; Michel and Westhof 1990; Suh et al. 1999). In the genomes of eukaryotes, nuclear group I introns are restricted to non-multicellular organisms. They are predominantly of IC1 and IE types and occur only in the rRNA genes. Those identified in *Acanthamoeba* spp. resulted to be

*Corresponding author.
E-mail address: corsaro@gmx.fr (D. Corsaro).

https://doi.org/10.1016/j.ejop.2018.07.002
0932-4739/

distinct members of the subclass IC1 occupying one of the four insertion sites on the 18S rRNA (Corsaro et al. 2017; Gast et al. 1994; Schroeder-Diedrich et al. 1998).

Uncharacterized nuclear introns have also been reported in strains of the genotype T4 (*A. castellanii* complex) (Liu et al. 2006; Nagyová et al. 2010). By analyzing two introns from T4 strains, we found that they are closely related to those of *A. griffini* T3 (Corsaro et al. 2017). The intron of one of these T4 strains (*Acanthamoeba* sp. KA/E4) was previously found to contain an open reading frame (ORF) for a homing endonuclease gene (HEG) with His-Cys box domain that is likely non-functional (pseudogene) (Haugen et al. 2004). We therefore conducted a systematic search in literature and GenBank databases for putative intron-containing 18S rDNA sequences of *Acanthamoeba* strains belonging to the genotype T4 in order to analyze the conserved core intron and to search for ORF. Sequences of the other *Acanthamoeba* genotypes were submitted to the same survey. Our results indicate that only *Acanthamoeba* T4 present 18S rRNA with at least two distinct ORF for endonucleases (EN) occurring in unrelated and differently inserted introns.

Materials and Methods

Retrieval of *Acanthamoeba* 18S rDNA sequences with introns

Acanthamoeba 18S rDNA sequences that possibly contain introns, with special emphasis for T4 strains, were retrieved from GenBank on the basis of published works. Further searches were conducted in the databases focusing on unusually long sequences. Once the introns and endonucleases of *Acanthamoeba* were identified, BlastN and TBlastN were used to search for putative close relatives, by using as query portions of the conserved core intron or the deduced proteins, respectively.

18S rDNA phylogeny

For each identified intron-bearing *Acanthamoeba* T4 strain, the intron was excluded from the 18S rDNA sequence and then close relatives were searched for with Blast. Retrieved sequences were aligned with selected reference strains of T4 and members of T3 using MAFFT and manually refined using BioEdit. Maximum Likelihood (ML, 1000 replicates) tree and pair-wise similarity values were inferred as described previously, using T3 as outgroup (Corsaro et al. 2015).

Intron analysis and phylogeny

The conserved core paired helices P3–P7 were identified manually on the basis of previous works (Corsaro et al. 2017; Michel and Westhof 1990) and Mfold was used to infer the remaining portions of the molecules. For intron phylogeny, the secondary structure was used to align the conserved core sequences of *Acanthamoeba* introns with those of selected HEG-containing introns. A subset of them are IE introns that were used as outgroup. Maximum likelihood (ML), distance (NJ) and maximum parsimony (MP) (1000 replicates) trees were inferred as described previously (Corsaro et al. 2017). In parallel, relationships between the introns of *Acanthamoeba* were analyzed based on full-length sequences without HEG insertions, with the Unweighted Pair-Group Method with Arithmetic Mean (UPGMA), and with MP (1000 replicates), using MEGA7 (Kumar et al. 2016).

Endonuclease identification and phylogeny

Nuclear introns from all *Acanthamoeba* genotypes were submitted to ORF search in their entire length using BioEdit. Once ORF was identified, the amino acid (aa) sequences were deduced according to standard code. Frameshift mutations and premature stop codons were corrected by comparing sequences with those from closely related functional EN, using the ATG start codon and the coding sequence for the His-Cys box domain as a guide. The obtained proteins were identified by comparative analyses of functional domains via the Conserved Domain Database (Marchler-Bauer et al. 2017).

The EN from nuclear introns of other amoebae, fungi and algae, were selected from the previous study of Haugen et al. (2004), along with some additional sequences recovered by similarity searches, and aligned with the EN from *Acanthamoeba* introns by using COBALT (Papadopoulos and Agarwala 2007). Phylogenetic trees (1000 replicates) were inferred with protein ML (proML, JTT + Γ model) using PHYLIP, and distance (minimum evolution) and maximum parsimony (MP), using MEGA7 (Kumar et al. 2016).

Results

Origin and 18S rDNA phylogeny of *Acanthamoeba* T4 strains with introns

A total of eight strains of genotype T4 with an intron in their 18S rDNA were found in the literature and GenBank. The strain VN14 was isolated from a salt cave, located in Slovakia (Nagyová et al. 2010). In that study, sixteen environmental *Acanthamoeba* strains were genotyped by GTSA.B1 sequencing and a T3 strain with intron (lacking however HEG) also was recovered. The other seven T4 strains originated all from South Korea, and they were isolated either from human keratitis samples (KA/E4, KA/E5, and KA/E21) (Jeong et al. 2007; Xuan et al. 2007) or from oceanic sediments (KA/MSS2, KA/MSS6, KA/MSS7, and KA/MSG23) (Liu et al. 2006).

Molecular phylogeny (Fig. 1) indicates that these strains form four distinct lineages. The Korean isolates from ocean sediments cluster into two separate groups according to