Barcoding Amoebae: Comparison of SSU, ITS and COI Genes as Tools for Molecular Identification of Naked Lobose Amoebae

Elena Nassonova a,1, Alexey Smirnov b, Jose Fahrni c, and Jan Pawlowski c

aLaboratory of Cytology of Unicellular Organisms, Institute of Cytology RAS, Tikhoretsky ave. 4, 194064 St. Petersburg, Russia
bFaculty of Biology and Soil Sciences, St. Petersburg State University, Universitetskaja nab. 7/9, 199034 St. Petersburg, Russia
cMolecular Systematics Group, Department of Zoology and Animal Biology, University of Geneva, Sciences III CH-1211, Geneva 4, Switzerland

Submitted March 14, 2009; Accepted July 30, 2009
Monitoring Editor: C. Graham Clark

Morphological identification of naked lobose amoebae has always been a problem, hence the development of reliable molecular tools for species distinction is a priority for amoebae systematics. Previous studies based on SSU rDNA sequences provided a backbone for the phylogeny of Amoebozoa but were of little help for the species distinctions in this group. On one hand, the SSU rDNA sequences were rather conserved between closely related species; on the other hand, the intraspecies polymorphism of the SSU gene obscured species identification. In the present study, a 3' fragment of the SSU, a complete ITS1-5.8S-ITS2 block and a 5' fragment of COI gene were cloned and sequenced for six Vannella morphospecies, of which V. simplex was represented by six different isolates. SSU rDNA and ITS were found to be inappropriate for species differentiation, while distinctive and homogenous COI sequences were obtained for each well-defined morphospecies. Moreover, a number of distinct COI genotypes have been identified among V. simplex isolates. This suggests that COI may be a good candidate for DNA barcoding of amoebae, but further studies are necessary to confirm the accurateness of the COI gene as a barcode in other gymnamoebae, and to understand the taxonomic meaning of COI variations.

© 2009 Elsevier GmbH. All rights reserved.

Key words: amoeba; systematics; identification; phylogeny; molecular biology; DNA barcoding.

Introduction

Naked lobose amoebae (gymnamoebae) are a common group of protists inhabiting all marine, freshwater and terrestrial biotopes. Their high numerical abundance suggests that they are among the main controllers of bacterial populations; they are the primary consumers of attached bacteria unavailable to other micrograzers (Arndt 1993; Bass and Bischoff 2001; Butler and Rogers 1996). Among them are dangerous pathogens (Acanthamoeba and Balamuthia), causing fatal infections in animals and humans and acting as...
vectors of various pathogenic organisms (Horn and Wagner 2004; Rodriguez-Zaragoza 1994; Rowbotham 1980).

Yet, because of a lack of informative morphological characters, the biodiversity of gymnamoebae remains poorly known. Species identification of amoeboid protists is still among the most difficult tasks because of the low number of morphological characters available, which are hard to formalize and describe (Boeve 1953; Smirnov and Goodkov 1999). This is especially true for amoebae with a generally simple morphology, such as the widely distributed family Vannellidae, that neither show distinct light-microscopic features in most species, nor reliable ultrastructural differences (Page 1980, 1988, 1991; Page and Blakey 1979; Smirnov 2002). The development of novel tools for more precise identification of such species is a priority.

Molecular systematics offered new possibilities to improve species recognition in amoebae. However, the first molecular studies of gymnamoebae were not really concerned with this problem, focusing mainly on the phylogenetic position and the relationships within gymnamoebae and Amoebozoa (Bolivar et al. 2001; Fahrni et al. 2003; Peglar et al. 2003; Smirnov et al. 2005). All of these studies were based on SSU rDNA, which is generally known to be very conservative at the species level. Indeed, an extensive study of 21 species of the family Vannellidae based on complete SSU rDNA sequence comparisons showed very small inter-specific sequence divergences. Moreover, in a few cases, the polymorphism of molecular clones obtained from the same strain exceeded the inter-specific divergence (Smirnov et al. 2007). The resolution of species boundaries improved only slightly with ITS rDNA sequences (Dyková et al. 2005).

Several years ago, the mitochondrial gene coding for cytochrome c oxidase subunit I (COI) was proposed as an ideal DNA barcode for living organisms (Hebert et al. 2003). Indeed, several studies have shown a successful application of COI for species distinction in Lepidoptera (Hajibabaei et al. 2006; Hebert et al. 2004a), birds (Hebert et al. 2004b), flowering plants (Kress et al. 2005), spiders (Barrett and Hebert 2005) and other organisms. In protists, the COI was positively tested as barcode in red algae (Robba et al. 2006; Saunders 2005), ciliates (Barth et al. 2006; Chantangsi et al. 2007) and diatoms (Evans et al. 2007).

The application of COI for barcoding amoebae has not yet been attempted to our knowledge. In fact, data on mitochondrial DNA of amoebae are limited to the study of mitochondrial genomes in Acanthamoeba castellani (Lonergan and Gray 1996) and in five mycetozoan species, namely in Dictyosteliida (Dictyostelium citrinum, D. discoideum, D. fasciculatum, Polysphondylium pallidum) and in the plasmodial slime mold Physarum polycephalum (Heidel and Glöckner 2008; Ogawa et al. 2000; Takano et al. 2001). The analysis of the mitochondrial genome of A. castellani demonstrated that COI and CO2 are co-transcribed from a single continuous ORF (cox1/2), however they do not exist in mature form as a COI-CO2 “fusion” protein. This suggests either an unusual translation termination mechanism acting between the cox1 and cox2 coding regions of the cox1/2 mRNA or post-translational proteolytic processing of a translated product whose synthesis continues into the cox2 coding region. In Dictyosteliida the COI gene also lacks a termination codon and is fused to the downstream CO2 gene. In Dictyostelium spp. the cox1/2 gene has group I introns. Mitochondrial genomes of A. castellani and D. fasciculatum have a modified genetic code (UGA specifies tryptophan). In contrast to D. fasciculatum other studied mycetozoans use the universal genetic code in mitochondria (Heidel and Glöckner 2008).

Here, we present the first analysis of COI sequences in a selected group of amoebae. We analyzed 55 partial sequences of the COI gene, the SSU rRNA gene and an ITS1-5.8S-ITS2 block from six morphospecies of the genus Vannella, and COI sequences from two species of the genus Cochlidiopodium. Our results confirm the capacity of COI to resolve the inter-species relationships in amoebae better than any other gene studied and prompt its further use for barcoding the gymnamoebae.

Results

The phylogenetic trees based on the partial SSU rDNA, ITS and COI gene sequences (Figs 1-3) were generally congruent and indicated that all Vannella species studied robustly group into two large clades, one containing all Vannella simplex strains, V. persistens, V. danica and Vannella sp. Geneva strain (clade SPD). Another clade consisted of Vannella arabica, V. bursella and V. calycinucleolus (clade ABC).
Figure 1. ML tree based on SSU rDNA gene sequences. PhyML; GTR 4 gamma rates, optimized number of invariant sites, gamma alpha parameter and tree topology; 100 bootstrap pseudo-replicates.

Vannella simplex (all strains)

SPD clade

Vannella persistens

Vannella danica

Vannella calycinucleolus, V. arabica and V. bursella

Vannella calycinucleolus CCAP 1589/12 AF464817
Vannella cf. nirodes ATCC 39845 AY183888

Vannella lata CCAP 1589/12 AF464817
Vannella arabica CCAP 1589/7 1
Figure 2. ML tree based on ITS1-5.8S-ITS2 sequences. PhyML; GTR 4 gamma rates, optimized number of invariant sites, gamma alpha parameter and tree topology; 100 bootstrap pseudo-replicates.
SSU rRNA Gene
The sequenced fragment was ca 840 bp in length. G+C content varied from 39.22% in *V. calycinucleolus* CCAP 1565/6 strain to 44.77% in *V. danica* 1589/17 strain. The structure of the SSU gene was similar in all studied strains; few gaps were inserted in the most variable areas of the alignment.

**Intra-strain variability:** We sequenced from 5 to 6 molecular clones of the amplified SSU rDNA fragment from every DNA sample. The results show that the sequenced fragment of SSU rDNA is always variable; the average value of intra-strain sequence divergence ranges from 0.7% to 1.9%, with the maximum value of 3.2% in *V. simplex* Malagnou strain, and no sequence divergence only in one case (in *V. simplex* L4A3 strain) (Table 2). Most of the nucleotide replacements in the SSU gene were found to be random; however, in some cases the same replacements were observed in several molecular clones. For example, in *V. danica* 2 of 6 clones have identical replacements (T-C position 441; CAG-TCT position 526; TC-CT position 741; and G-A position 747). Other noteworthy patterns of regular replacements were found in the sequences of *V. simplex* L4A3 strain and *V. simplex* Malagnou strain.

**Intra-morphospecies variability:** Among the studied strains of amoebae we identified five as probably belonging to the morphospecies *Vannella simplex*. Four of them were isolated at Valamo Island, one in Switzerland. In addition, we studied the CCAP 1989/3 strain, isolated in Germany and earlier recognized as a neotype of this species (Smirnov et al. 2002). The level of sequence divergence between the Valamo

Table 1. DNA sources, origin of strains and GeneBank accession numbers of sequences used in the present study.

<table>
<thead>
<tr>
<th>Species and strain name</th>
<th>Origin and habitat</th>
<th>DNA stock used</th>
<th>GeneBank numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Vannella simplex</em></td>
<td>Freshwater, pool in botanic garden, University of Bonn, Germany, by N. Hulsmann</td>
<td>Smirnov et al. (2007)</td>
<td>GQ265380 GQ265467 GQ354154</td>
</tr>
<tr>
<td>CCAP 1589/3</td>
<td></td>
<td></td>
<td>GQ265381 GQ265468 GQ354155</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GQ265382 GQ265469 GQ354156</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GQ265383 GQ265470 GQ354157</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GQ265384 GQ265471 GQ354158</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GQ265385 GQ265472 GQ265473</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GQ265474</td>
</tr>
<tr>
<td>Species and strain name</td>
<td>Origin and habitat</td>
<td>DNA stock used</td>
<td>GeneBank numbers</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td><strong>Vannella simplex</strong> L4A Cl 4/1/2</td>
<td>Freshwater, Valamo Island, subcloned three times, by A. Smirnov, 2006</td>
<td>Freshly prepared</td>
<td>GQ265404 GQ265405 GQ265406 GQ265407 GQ265408 GQ265409</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GQ265492 GQ265493 GQ265494</td>
</tr>
<tr>
<td><strong>Vannella simplex</strong> L4A3 Cl 10/1/2</td>
<td>Freshwater, Valamo Island, subcloned three times, by A. Smirnov, 2006</td>
<td>Freshly prepared</td>
<td>GQ265410 GQ265411 GQ265412 GQ265413 GQ265414 GQ265415</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GQ265495 GQ265496 GQ265497 GQ265498 GQ265499 GQ265500</td>
</tr>
<tr>
<td><strong>Vannella simplex</strong> L4C Cl 1</td>
<td>Freshwater, Valamo Island, clonal culture, by A. Smirnov, 2006</td>
<td>Freshly prepared</td>
<td>GQ265416 GQ265417 GQ265418 GQ265419 GQ265420 GQ265421</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GQ265501 GQ265502 GQ265503 GQ265504 GQ265505 GQ265506</td>
</tr>
<tr>
<td><strong>Vannella simplex</strong> Ladoga Cl 12/2</td>
<td>Freshwater, Ladoga Lake, coastal waters of Valamo Island, subcloned twice, by A. Smirnov, 2006</td>
<td>Freshly prepared</td>
<td>GQ265398 GQ265399 GQ265400 GQ265401 GQ265402 GQ265403</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GQ265404 GQ265405 GQ265406 GQ265407 GQ265408 GQ265409</td>
</tr>
<tr>
<td><strong>Vannella simplex</strong> Malagnou</td>
<td>Freshwater, pond at Malagnou Road, Geneva</td>
<td>Freshly prepared</td>
<td>GQ265422 GQ265423 GQ265424 GQ265425 GQ265426</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GQ265507 GQ265508 GQ265509 GQ265510 GQ265511</td>
</tr>
<tr>
<td><strong>Vannella simplex</strong> Geneva</td>
<td>Isolate of unknown origin, kindly provided by Dr. Bob Peck (Geneva)</td>
<td>Freshly prepared</td>
<td>GQ265428 GQ265429 GQ265430</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GQ265513 GQ265514 GQ265515</td>
</tr>
<tr>
<td><strong>Vannella persistens</strong> CCAP 1589/13</td>
<td>Soil, Sourhope Research Station, UK, by S. Brown, 1999</td>
<td></td>
<td>GQ265369 GQ265370 GQ265371 GQ265372</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GQ265460 GQ265461 GQ265462 GQ265463</td>
</tr>
</tbody>
</table>
Table 1. (continued)

<table>
<thead>
<tr>
<th>Species and strain name</th>
<th>Origin and habitat</th>
<th>DNA stock used</th>
<th>GeneBank numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vannella danica CCAP 1589/17</td>
<td>Brackish-water, Nivå Bay, sediments, Helsingør, Denmark, by A. Smirnov, 2000.</td>
<td>GQ265373, GQ265374</td>
<td>GQ265463, GQ265464, GQ265466</td>
</tr>
<tr>
<td>Vannella arabica CCAP 1589/7</td>
<td>Marine, seawater tank in Kuwait Inst. for Scientific Research, Kuwait, by F. Page, 1977.</td>
<td>GQ265392, GQ265393, GQ265394, GQ265395, GQ265396, GQ265397</td>
<td>GQ265447, GQ265448, GQ265449, GQ265450, GQ265451, GQ265452, GQ265453, GQ265454, GQ265455, GQ265456, GQ265457, GQ265458, GQ265459</td>
</tr>
<tr>
<td>Vannella (former Platyamoeba) calycinucleolus CCAP 1565/6</td>
<td>Marine, West Mersea beach, Essex, UK, by F. Page, 1972</td>
<td>GQ265363, GQ265364, GQ265365, GQ265366, GQ265367, GQ265368</td>
<td>GQ265435, GQ265436, GQ265437, GQ265438, GQ265439, GQ265440, GQ265441, GQ265442, GQ265443, GQ265444, GQ265445, GQ265446, GQ265447, GQ265448, GQ265449, GQ265450, GQ265451, GQ265452, GQ265453, GQ265454, GQ265455, GQ265456, GQ265457, GQ265458, GQ265459</td>
</tr>
<tr>
<td>Vannella (former Platyamoeba) bursella CCAP 1565/10</td>
<td>Marine, River Deben, Woodbridge, Suffolk, UK, by F. Page, 1973</td>
<td>GQ265375, GQ265376, GQ265377</td>
<td>GQ265441, GQ265442, GQ265443, GQ265444, GQ265445, GQ265446, GQ265447, GQ265448, GQ265449, GQ265450, GQ265451, GQ265452, GQ265453, GQ265454, GQ265455, GQ265456, GQ265457, GQ265458, GQ265459</td>
</tr>
<tr>
<td>Cochliopodium minutoidum Cm1</td>
<td>Freshwater, by A. Kudryavtsev</td>
<td>Freshly prepared</td>
<td>–</td>
</tr>
<tr>
<td>Cochliopodium cf. actinophorum Bor1</td>
<td>Freshwater, Borok, Russia, by A. Kudryavtsev</td>
<td>Freshly prepared</td>
<td>–</td>
</tr>
</tbody>
</table>

Subsequent numbers in a single cell are assigned to the sequences of the molecular clones originating from the same PCR product. Smirnov et al. (2007) reference in the “DNA stock used” means that the same DNA sample was used. Cultures possessing CCAP numbers were obtained in 2006 from the Culture Collection of Algae and Protozoa, Dunbeg, UK. Strains not held in public culture collections are available upon request.

Archipelago strains identified as V. simplex varied from 1.4% to 1.6%, generally being a little higher than the level of intra-strain variability (Supplementary Material Table 3). The divergence between Malagnou and CCAP 1589/3 strains and all other studied V. simplex strains was 1.6% – 1.7% and 2.0% – 2.2%, respectively. Interestingly, the Vannella simplex Geneva strain shows...
only 1.6% – 2.0% divergence from Valamo archipelago strains and 2.4% divergence from the CCAP 1589/3 strain.

**Inter-specific variability:** The sequence divergence was rather high between the species belonging to the SPD clade and the ABC clade and much lower within each clade (Supplementary Material Table 3). The inter-clade divergence typically was 9.3% – 11.4%, whereas it ranged from 0.6% to 0.9% in the ABC clade and from 1.6% and 7.5% within the SPD clade. This difference corresponds to the observation of the clear molecular signatures in the region 514 - 565 bp and 735 - 772 bp in all members of the SPD clade.

**Molecular phylogeny:** The configuration of the SSU rDNA tree (Fig. 1) was congruent with our previous data (Smirnov et al. 2007). In this tree we used as outgroups two other *Vannella* species that have been shown to branch as sister group to the species examined in this study (Smirnov et al. 2007). Using these closely related species as outgroups allowed us to preserve many semi-conserved and variable regions valuable for species distinction (see Smirnov et al. 2002) and made the relative distance between the sequences of interest and the outgroup comparable to that obtained in other studied genes. The dichotomy between the clades SPD and ABC is 100% supported, but relations within these clades are not well resolved. The clones of different species mix together with the exception of two species (*V. danica* and *V. persistens*). All *V. simplex* strains, including *Vannella simplex* Geneva strain, branch together in the weakly supported (60%) grouping. Four of six clones of *V. simplex* CCAP 1589/3 form a sister group to this clade. Two other clones branch together with other *V. simplex* strains, none of which form a distinct clade. *V. danica* branches separately with 100% support.

**ITS1- 5.8S - ITS2 Block**

The sequenced fragment was ca 560 bp in length. The overall G+C content varied from 17.34% in *V. calycinucleolus* CCAP 1565/6 strain to 35.58% in *V. danica* 1589/17 strain. The structure of this gene block was very divergent between species belonging to the clades ABC and SPD. The sequences could be properly aligned only within each clade.

**Intra-strain variability:** Molecular clones from the ITS1-5.8S-ITS2 PCR products show the highest level of intra-strain variability compared with COI and SSU genes (Table 2). The average level of

Table 2. Intra-isolates sequence divergence in *Vannella* strains.

<table>
<thead>
<tr>
<th><em>Vannella</em> species</th>
<th>SSU</th>
<th>ITS1-5.8S-ITS2</th>
<th>COI</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Vannella calycinucleolus</em> CCAP 1565/6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td><em>Vannella persistens</em> CCAP 1589/3</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td><em>Vannella bursella</em> CCAP 1565/10</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td><em>Vannella simplex</em> CCAP 1589/3</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td><em>Vannella danica</em> CCAP 1589/17</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td><em>Vannella arabica</em> CCAP 1589/7</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td><em>Vannella simplex</em> L4A strain</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><em>Vannella simplex</em> L4A3 strain</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><em>Vannella simplex</em> L4C strain</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><em>Vannella cf. simplex</em> Malagnou strain</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><em>Vannella simplex</em> Geneva strain</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><em>Vannella simplex</em> L4A strain</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><em>Vannella simplex</em> L4A3 strain</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><em>Vannella simplex</em> L4C strain</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><em>Vannella cf. simplex</em> Malagnou strain</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><em>Vannella simplex</em> Geneva strain</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

**Table 2. Intra-isolates sequence divergence in *Vannella* strains.**
sequence divergence varied from 0.9% to 5.5%. Few clones obtained from the same amplicon were completely identical, while in some strains the level of sequence divergence reached 7.5%. As well as in SSU rRNA gene, most of the nucleotide replacements were random, but in some cases it was possible to identify similar replacements.

**Intra-morphospecies variability:** Similarly to the SSU rRNA gene, the level of sequence divergence between the Valamo archipelago strains identified as *V. simplex* was lower (0.8 – 1.1%) than between these strains and the Geneva strain (0.9 – 1.3%) and reached even higher values between all these strains and the CCAP 1589/3 strain (1.6 – 2.1%) (Supplementary Material Table 3). As well as in the SSU rRNA gene, it was impossible to establish any regular strain-specific molecular signatures due to high number of random replacements of nucleotides.

**Inter-specific variability:** Similarly to the SSU rRNA gene, the sequences of *Vannella* species belonging to the clades ABC and SPD were very different. The inter-specific divergence was rather high between species of different clades and much lower within each clade (Supplementary Material Table 3). The inter-clade divergence typically was 13.1 – 15.9%, whereas within the ABC clade it was only 0.1% – 1.3% and ranged from 0.9% to 4.1% in the SPD clade.

**Molecular phylogeny:** The configuration of the ITS tree differs slightly from that of the SSU rDNA tree. The distinction of two clades ABC and SPD is strongly supported, but the relations within each clade remain partly unresolved. Within the clade ABC, the sequences of *V. calycinucleolus* form a distinct group (100%), while the clones of two other species mix together. Within the clade SPD, there is an unresolved basal trichotomy between *V. danica*, *V. persistens* and other vannellids, among which only the *V. simplex* CCAP 1589/3 and *Vannella simplex* Geneva strains form distinct clades. The sequences of other *V. simplex* strains mix together.

**COI Gene**

The sequenced region of the COI gene was ca. 650 bp in length and had rather uniform structure among all strains studied. G+C content varied from 28.78% in *Vannella simplex* CCAP1589/3 strain to 31.83% in *V. bursella* 1565/1 strain. Among sequenced strains of *Cochliopodium* it varied from 28.11% in *C. minutoidum* to 29.87% in *Cochliopodium* cf. *actinophorum*. The differences between vannellids and *Cochliopodium* required insertion of few gaps in the alignment. The inferred translation of obtained sequences suggests that the mitochondrial genomes of gymnamoebae use a modified genetic code and translate the codon UGA as tryptophan. Moreover, in vannellids the COI RNA transcripts undergo insertional editing similar to that described in *Physarum polycephalum* mitochondria (Horton and Landweber 2000; Miller et al. 1993; Viso-mirski-Robic and Gott 1997). One putative site of C insertion at position 141 was identified in the sequenced COI fragment of all *Vannella* species. No potential edited sites were found in COI sequences of *Cochliopodium* species. No introns were found in the sequenced area of the gene.

**Intra-strain variability:** We have sequenced six to eight clones from the amplified fragment obtained from every DNA sample, each originating from a single clonal strain of amoebae. The common level of intra-strain variability of COI gene was 0.1%. Only in one case, in *V. persistens*, it reached 0.5%, and in five strains we observed no intra-strain variability (Table 2). Among the studied *V. simplex* strains, in *V. simplex* L4A3 two molecular clones had regular replacements of a single nucleotide (A-G, position 33, T-C, position 300 and 393 and A-T position 419). Of these replacements, three were found to be synonymic at the amino acid level, while the latter one resulted in the replacement of one amino acid residue. This may reflect the presence of paralogs of the COI gene in the genome of this strain.

**Intra-morphospecies variability:** Data provided in the Supplementary Material Table 4 indicate that the level of variability among the strains identified as *V. simplex* is 0.2 – 0.8% in the case of the strains isolated at Valamo archipelago; it rises to a maximum of 1.8% including the strain isolated in Malagnou and up to a maximum 5.1% when the CCAP 1589/3 strain was added. The sequenced fragment of the COI gene showed a certain polymorphism even among the isolates originating from closely related sites at the Valamo archipelago. For example, *V. simplex* Ladoga strain CI12/2 from the coastal waters of the archipelago had regular differences from the three other Valamo strains isolated from an inner lake in three positions (A-G, positions 96; T-C in the position 311 and A-C position 335). But all these replacements were found to be synonymic with other Valamo Island strains of this species at the amino acid level. It must be stressed that the numerical analysis of sequence divergence in this case is somehow misleading because all studied
sequences of all other studied nucleotide replacements compared with the sequences of all other studied Vannella species.

**Inter-specific variability:** Variations between the different Vannella morphospecies always exceeded the intra-specific ones, varying from 7.3% and reaching 21.6% in some cases (Supplementary Table 2). The exception was the pair of most closely related species – V. bursella and V. arabica, in which the level of sequence divergence was only 0.9%. However, these two species differ clearly in six positions (21, 125, 222, 258, 496 and 608) easily observed in the alignment. The level of sequence divergence between the species belonging to the same clade (0.9 – 7.8% in the case of clade ABC and 4.3 – 12.6% in the case of clade SPD) was always lower than between those belonging to different clades (20.1 – 21.7%).

**Molecular phylogeny:** The COI tree (Fig. 3) shows clear separation of all examined strains, including the strains identified as belonging to the morphospecies V. simplex. As in the SSU rDNA and ITS trees, there is a clear dichotomy between the SPD clade and ABC clade. Each of the two clades is strongly supported as well as their dichotomy (100% bootstrap). Within clade SPD, all strains of V. simplex isolated from the Valamo archipelago group together, whereas the V. simplex Ladoga strain forms an isolated clade. The support for the entire group of V. simplex strains isolated from the Valamo archipelago is low. However, the clade of all V. simplex strains, including the Malagnou strain and the CCAP 1589/3 strain is supported by 98% bootstrap value. There is also 100% support for the clade including all V. simplex, V. persistens, V. danica strains and the Vannella simplex Geneva strain.

Inside clade ABC, each species-specific clade was highly supported, with bootstrap values ranging from 83% for V. arabica, 99% for V. bursella to 90% for V. calycinucleolus. As in the ITS tree, V. calycinucleolus formed a sister group to V. arabica and V. bursella, a grouping supported by 100% bootstrap value.

**Discussion**

Currently it is clear that morphological identification is not always adequate for naked lobose amoebae. Some of the morphologically indistinguishable isolates appear to be genetically very different, as best exemplified by analysis of the V. simplex and V. danica species groups (Smirnov et al. 2002, 2007). Vice versa, a number of amoebae species, well distinguishable morphologically, have only very few sequence differences (Smirnov et al. 2007). Until now, the molecular diversity of amoebae has been studied in only few genera (Acanthamoeba, Vannella, Nebela) and is based almost exclusively on SSU rRNA gene sequences (Lara et al. 2008; Liu et al. 2006; Smirnov et al. 2007; Stothard et al. 1998). This gene has been shown to be very useful for inferring phylogenetic relations within Amoebida (Bolivar et al. 2001; reviewed in Pawlowski 2008; Pawlowski and Burki 2009; Smirnov et al. 2005). However, its use to recognize species in lobose amoebae is limited due to its low rate of evolution and intra-specific polymorphism (Smirnov et al. 2007). Our study confirms these limitations. As shown in Figure 1, the analysis of SSU rDNA cannot distinguish between species in the ABC clade (V. arabica+V. bursella and V. calycinucleolus) and does not resolve the relations between different strains of V. simplex, which are blurred by high intra-strain polymorphism reaching up to 3.2%.

The ITS region provides much better resolution of relationships in the genus Vannella (Fig. 2). This is in agreement with the study of Dyková et al. (2005), who showed that 12 out of 14 strains of Vannella could be distinguished using the ITS, although the relations among these strains differed from those obtained using SSU rDNA. In the genus Acanthamoeba, ITS1 variation was tenfold higher than in the SSU rDNA, but the ITS1 analysis corroborated the Acanthamoeba sequence type system established based on the SSU data (Köhlsler et al. 2006). Similarly, there is no major contradiction between the ITS and SSU rDNA trees in our study (Figs 1-2), suggesting that both loci bear comparable phylogenetic signal. Nevertheless, not all relations between Vannella species could be resolved using ITS (Fig. 2). In particular, V. arabica and V. bursella, as well as different strains of V. simplex, remain indistinguishable in the ITS trees. In the latter case, this is partly due to high intra-strain polymorphism, reaching up to 5.3% in some V. simplex isolates. Although we know nothing about the ITS polymorphism in other gymnamoebae, the heterogeneity of the ITS sequences is well documented in other groups of organisms and is usually given as the argument against using this locus as DNA barcode in animals (Blaxter 2004) and some groups of land plants (Chase et al. 2005).
In comparison to the SSU rDNA and ITS, the COI offers the highest resolution of relations among Vannella species. All examined morphospecies, including V. arabica and V. bursella, can be distinguished in the COI tree, usually with strong support (Fig. 3). The COI also allows distinction of some strains of V. simplex (Ladoga, Malagnou), indistinguishable using SSU rDNA or ITS. Although the sequence divergence between these strains is relatively low (0.6 – 1.7%), we did not observe the intra-strain heterogeneity that blurred their distinction in SSU rDNA and ITS analyses. The few paralogous copies of COI that have been found in our study were easily distinguishable and probably corresponded to mitochondrial genes transferred to the nucleus (numts; Gray et al. 2004).

The undeniable advantages of COI, in terms of high phylogenetic resolution, low level of polymorphism, and congruence with rDNA-based phylogenies, makes it an ideal barcode for the Vannellidae, and possibly for many other amoebazoans. Although the database for COI of Amoebozoa is very limited, the ease with which we have obtained new COI sequences for Cochliopodium and Acanthamoeba (presented in this paper) as well as for Amoeba proteus and Trichosphaerium sieboldi (unpublished data) suggests that it may be possible to sequence the COI for a much broader range of Amoebozoa. Apparently the classical Folmer’s primers work quite well for amoebae, in contrast to what was observed in other groups of protists for which COI has been tested. In diatoms, the COI primers used for amplification of raphid pennates do not work for centric diatoms (Evans et al. 2007).

In ciliates, specific COI primers have been developed for each of the examined genera, including Paramecium (Barth et al. 2006), Tetrahymena (Chantangsi et al. 2007; Lynn and Strüder-Kypke 2006) and Cyclidium (Guggiari and Peck 2008).

If the universality of COI primers is confirmed for a wider range of amoebozoan taxa, it would make this gene a very promising tool for studying the molecular ecology of amoebae. Indeed, the SSU rDNA-based environmental DNA surveys of eukaryotic diversity usually report very few amoebozoan sequences (Berney et al. 2004), even with a multi-primer approach (Stoeck et al. 2006). This can be due to various difficulties related to the amplification of amoebozoan rDNA genes, such as their unusual length, high level of sequence divergence, presence of introns or possibly limited amount of DNA in some species (Bolivar et al. 2001; Pawlowski 2008). Although there is no published environmental study using the COI gene, we can predict that such a study would provide us with a much more realistic view on the natural diversity of amoebae.

The only major problem with using COI as a barcode for lobose amoebae is the apparent lack of a clearly defined “barcoding gap”. This term denotes a significant difference between intra-specific and inter-specific divergence, which is considered as an essential feature of a good barcoding gene (Waugh 2007). In the case of the Vannellidae, two morphospecies (V. arabica and V. bursella) are separated by 0.9% divergence, which is less than the maximum divergence within the V. simplex complex. If we consider V. arabica and V. bursella as separate species based on their phenotypic differences, then we should also consider in the same way each of the morphologically indistinguishable V. simplex strains which differ by more than 0.9%. Because we do not have any well-defined species concept for amoebae, it may be preferable to consider the distinctive COI phylotypes as MOTU (Molecular Operational Taxonomic Unit) (Floyd et al. 2002) unless their taxonomic meaning is established.

To conclude, in view of our study, the COI gene appears to be the best candidate for barcoding lobose amoebae. However, as for any gene used for species identification, its general application will require better definition of species boundaries in amoebae and better understanding of morphological and ecological characteristics of amoebae COI types.

**Methods**

For the present study we have chosen six species of amoebae of the family Vannellidae, belonging both to the genus Vannella and the former genus Platymoeba. These six species represent two independent clades in the SSU rDNA tree (Smirnov et al. 2007). To prevent misidentification problems, we used the same DNA stocks, which were prepared in 2006 for the study of the 18S rRNA phylogeny of vannellid amoebae (op. cit.). To study the polymorphism of genes inside a single amoeba morphospecies, DNA was extracted from four strains of Vannella simplex isolated at Valamo Island (North-Western Russia) during the year 2006 (Table 1), Malagnou pond (Geneva) and from a strain of vannellid amoeba maintained at the University of Geneva. All these strains were cloned, and most were subcloned once or twice to make sure that we really work with a population recently derived from a single cell. The samples of DNA of Cochliopodium minutoidum and Cochliopodium cf. actinophororum (both were the clonal cultures) used as an outgroup for the COI tree were kindly provided by Dr. A. Kudryavtsev.
Amoebae were maintained on NN agar plates (Page 1988), feeding on accompanying bacteria. DNA was extracted using guanidine thiocyanate followed by isopropyl alcohol precipitation (Maniatis et al. 1982). To collect cells, amoebae and food bacteria were gently scraped from the agar surface with a disposable plastic scraper to form an aggregate, to the top of which ca.100 µl of guanidine thiocyanate was added, mixed briefly, and immediately transferred by Pasteur pipette to an Eppendorf tube. The same method was used for DNA preparation in 2006.

The 3’ fragment of the SSU rRNA gene was amplified using primers s12.2 (5’-GAT YAG AYA CCG TCG TAG TC-3’) and sB (5’-TGA TCC TTC TGCC AGG TTC ACC TAC -3’). The ITS region was amplified using primers Gfor (5’-GGG ATC CGT TTC CGT AGG TGA ACC TGC -3’) and Grev. A 5’ fragment of COI gene ca 650 bp in length was amplified using forward primer LCO1490f 5’ GGT CAA CAA ATC ATA AAG ATA TGG G 3’ and reverse primer HCO2198r 5’ TAA ACT TCA GGG T-3’) (Folmer et al. 1994). The PCR conditions were described elsewhere (Fahrni et al. 2003). Positive amplification products were purified on agarose gels with a MinElute Gel extraction kit (Qiagen), ligated into pCR4-TOPO vector (TOPO TA cloning Kit, Invitrogen) and cloned in OneShot TOP10 ultracompetent cells (Invitrogen). To estimate the intra-strain polymorphism, from six to ten clones were sequenced for each strain was calculated as an average of the divergence values obtained using BioEdit (Hall 1999). Maximum likelihood phylogenetic trees were obtained using PhyML software (Guindon and Gascuel 2003), using GTR+G+I model suggested by Modeltest (Posada and Crandall 1998), with four gamma rate categories and the number of invariant sites, gamma alpha parameter and tree topology optimized by PhyML. The non-parametric bootstrap analysis was performed with 1000 bootstrap pseudo-replicates. Sequence divergence was calculated using the observed divergence matrix generated by PhyloWin program (Galtier et al. 1996); since we had multiple sequences per strain, the inter-specific divergence for each pair of strain was calculated as an average of the divergence calculated for every molecular clone of one strains against every clone of another strain.

Acknowledgements

This work was supported by the Swiss National Science Foundation SCOPES project IZ73A0111064, RFBR 06-04-49387 grant and G. & A. Claraz Donation. We thank Dr. R. Peck and two anonymous reviewers for comments and corrections.

Appendix A. Supporting Information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.dsr.2009.08.007.

References


Smirnov AV, Nasonova ES, Holzmann M, Pawlowski J (2002) Morphological, ecological and molecular studies of...
Vannella simplex Wohlfarth-Bottermann 1960 (Lobosea, Gymnamoebia), with a new diagnosis of this species. Protist 153:367–377


Available online at www.sciencedirect.com