Parasitology Research (2019) 118:1751–1759
https://doi.org/10.1007/s00436-019-06294-x

GENETICS, EVOLUTION, AND PHYLOGENY - ORIGINAL PAPER

# Novel data from Italian *Vermamoeba vermiformis* isolates from multiple sources add to genetic diversity within the genus

**M. Montalbano Di Filippo**[1] · **F. Berrilli**[1] · **D. Di Cave**[1] · **A. Novelletto**[2]

Received: 9 November 2018 / Accepted: 14 March 2019 / Published online: 9 April 2019

## Abstract

*Vermamoeba vermiformis* represents one of the most common free-living amoebae identified in worldwide environmental surveys. We analyzed 56 water samples with varying characteristics, including temperature and the particular settings in which humans may be exposed to water, plus one corneal scraping from a keratitis patient, with the following aims: (i) to investigate the presence of *V. vermiformis*; (ii) to identify the isolate subtypes; (iii) to place the Italian isolates in the broader picture of the genetic diversity within *V. vermiformis.* Twenty-two isolates were identified upon culturing and sequencing of > 600 bp in the 18S ribosomal RNA (rRNA) gene sequence, bringing to 27 the number of sequences recovered from Italian sources. By adding deposited sequences, we assembled a dataset of 74 isolates. Three of our isolates were characterized by allelic code 7-5-1-1, never reported before, and two showed 100% identity with an uncultured eukaryote and carried the 719T>C variant. We show that the variable segments E5, E3, F, and G convey most of the information on diversity, enabling the clustering of the isolates in a replicable fashion. The presence of different strains in natural thermal waters and in distribution systems indicated heterogeneity of the amoebic populations. Also, ours and the only other sequence from human infection were mapped in different clades. Overall, we enlarged the repertoire of single nucleotide and indel variants and the list of allelic codes, proceeding one step further in the description of the diversity within the genus.

**Keywords** Amoebozoa · *Vermamoeba vermiformis* · 18S rRNA · Allelic variants · Dynamic homology

## Introduction

*Vermamoeba vermiformis* represents one of the most common free-living amoebae (FLA) identified in worldwide environmental surveys (Delafont et al. 2018). In recent years, its importance has been increasingly recognized in association with human diseases. However, contrary to other potentially pathogenic FLA, such as *Acanthamoeba* spp., *Naegleria* spp., *Balamuthia mandrillaris*, and *Sappinia diploidea,* known as primary agents of human diseases, the pathogenicity of *V. vermiformis* is still debated. It has been recovered from human samples only in a few instances, i.e., once from spinal fluid (Centeno et al. 1996) and in a small number of cases of ocular infections (Aitken et al. 1996; Inoue et al. 1998; Lorenzo-Morales et al. 2007; Niyyati et al. 2014).

The genus *Vermamoeba* has been recently established and consists of one single species, *Vermamoeba vermiformis*. Based on molecular phylogeny, Smirnov and collaborators carried out a reclassification of the naked lobose amoebae, showing that the species *Hartmannella vermiformis* did not group with the other members of the genus (Smirnov et al. 2005). Later, they proposed the genus *Vermamoeba* (*Vermamoeba,* Cavalier-Smith and Smirnov gen. n.), to contain the isolates previously identified as *H. vermiformis* (Smirnov et al. 2011).

There is still limited knowledge on the genetic variability and the evolutionary relationships within the genus. Molecular analyses of *V. vermiformis* are mostly restrained to sequencing of the 18S rRNA gene, which is used as marker for taxonomic inferences. Since the initial report (Gunderson et al. 1994), all 18S rRNA gene sequences available for this species show a high

---

Handling Editor: Julia Walochnik

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s00436-019-06294-x) contains supplementary material, which is available to authorized users.

✉ F. Berrilli
berrilli@uniroma2.it

[1] Department of Clinical Sciences and Translational Medicine, University of Rome Tor Vergata, Rome, Italy

[2] Department of Biology, University of Rome Tor Vergata, Rome, Italy

degree of conservation, and intragenus sequences divergence (full-length or near full-length) is less than 1% (Delafont et al. 2018). However, the presence of allelic variations at sites in 11 segments (A–D; E5; E3; F–J), spread over the length of the gene of 35 isolates of *V. vermiformis* ("almost complete" 18S rRNA gene sequences, exceeding 1500 bp), allowed the identification of different allelic codes, clustering into four to seven groups (see Fuerst and collaborators (http://u.osu.edu/acanthamoeba/genetic-variation-within-v-vermiformis/), last update December 2017, hereafter simply "Fuerst 2017").

The resulting sets of sequences represent a core group useful to categorize also shorter sequences into the same subgroups.

In this study, we analyzed samples from different kinds of sources in order to (i) investigate the presence of *V. vermiformis*; (ii) identify the isolates at the subtype level, using 18S rRNA as molecular marker; (iii) place the Italian isolates in the broader picture of the genetic diversity within *V. vermiformis,* as emerging from the set of homologous sequences retrieved from the literature.

## Materials and methods

### Study area, sources, and sampling

The present study analyzed a total of 57 samples, consisting of geothermal waters ($n = 36$), from two different geothermal springs in Central Italy (site A and site B, 700 m from each other—see Montalbano Di Filippo et al. 2017), as well as dental unit waterlines ($n = 19$ from inlet, spittoon, and handpiece), tap water ($n = 1$), and one clinical sample (corneal scraping from a symptomatic patient with suspected keratitis), collected at much wider distances in Northern and Central Italy. In particular, 2 L of geothermal and tap waters and 1 L from dental unit waterlines were collected in sterile flasks. The water samples were transported to the laboratory at ambient temperature and processed within 24–72 h. In Table 1, samples are grouped according to the source.

### Amoeba culture

The entire volume of each water sample and of the corneal scraping was centrifuged at 4000*g* for 20 min. The pellet was resuspended in 200 μl of supernatant, and 100-μl aliquots were inoculated on non-nutrient agar (prepared with sterile double distilled water) in two plates, with a lawn of heat inactivated (1 h at 60 °C) *Escherichia coli*, routinely cultured in our microbiology laboratory and resuspended in Page's amoeba saline (PAS) solution. The plates were incubated at 37 °C (45 °C for geothermal waters) and observed daily for amoebic growth, up to 21 days after inoculation, by an inverted microscope at × 200 and × 400 magnification.

### Molecular procedures

From all positive cultures, the growing amoebae were harvested from plates, transferred in Eppendorf tubes and washed two times with PBS, pH 7.4. DNA extraction was performed by using QIAamp DNA Micro Kit (Qiagen, Milan, Italy).

A PCR amplification targeted at the *V. vermiformis* 18S rRNA gene was performed with primers P-FLA-F (5′-CGCGGTAATTCCAGCTCCAATAGC-3′) and P-FLA-R (5′-CAGGTTAAGGTCTCGTTCGTTAAC-3′) (Tsvetkova et al. 2004), expected to produce an amplicon of 774 bp in reference sequence M95168. PCR was carried out in a 25 μL volume containing 12.5 μL PCR master mix 2X (Promega, Milan, Italy), 5 μL template DNA, and 0.6 mM of each primer and performed in TProfessional Basic Thermocycler (Biometra GmbH, Göttingen, Germany), with thermal profile as in Tsvetkova et al. (2004). The PCR products were visualized by electrophoresis on 1.5% agarose gels, stained by a SYBR Safe DNA gel stain (Invitrogen, Monza MB, Italy). For positive samples, PCR amplicons were purified using mi-PCR Purification Kit (Metabion GmbH, Steinkirchen, Germany) according to the manufacturer's instructions and shipped to external service for sequencing (Bio-Fab Research, Rome, Italy).

DNA chromatograms were examined using FinchTV 1.4 software (Geospiza, Inc., Seattle, WA, USA). After adjoining of the forward and reverse reads, the sequences included approximately 60% of positions readable on both strands. The sequences were pairwise aligned using Clustal Omega software (Chenna et al. 2003) to recheck variable positions. Sequences were trimmed to the shortest length with high quality in all of our samples and complete representation in the downloaded ones. This produced a valid alignment over the region corresponding to > 650 bp in the reference sequence M95168.

**Table 1** Isolation sources, number of samples examined, and results obtained by culture and molecular analysis of *V. vermiformis* in Italy

| Source origin | Samples examined Total | Cultured positive samples Number (%) | Molecular identification of *V. vermiformis* |
|---|---|---|---|
| Geothermal water (T) | 36 | 31 (86.1) | 13 |
| Dental unit waterlines (R) | 19 | 7 (36.8) | 7 |
| Tap water (S) | 1 | 1 (100) | 1 |
| Corneal scraping (C) | 1 | 1 (100) | 1 |

### Genetic variability and phylogenetic analysis

All sequences were compared with the specific *V. vermiformis* dataset which includes 34 isolates previously analyzed in "Fuerst 2017," after removing the 3′ incomplete sequence KC306529. Additional sequences were searched by similarity in GenBank with BLAST (McGinnis and Madden 2004) (www.ncbi.nih.gov/BLAST/), with those produced in-house as queries. A final GenBank search was performed for sequences matching the keywords *Vermamoeba vermiformis/ Hartmannella vermiformis* (452 sequences—last update July 2018), *Hartmannella* spp. (33 sequences), and uncultured eukaryote (65 sequences). This resulted in a set of sequences that went under closer screening for incomplete or improper labeling. In addition to the annotations, the identification was aided by calculating genetic distances with MEGA v5 (Tamura et al. 2011) using the Kimura 2-parameter substitution model (Kimura 1980).

After alignment, the entire set was trimmed at positions 652 to 1274 of the reference, to maximize the overlap between ours and Fuerst's sequences. This procedure retained a total of 74 complete sequences, of which 22 were newly generated, 34 from "Fuerst 2017," and 18 additional from GenBank (Tables 1 and 2).

All sequences covered four out of 11 variable segments identified by "Fuerst 2017," (i.e., E5-E3-F-G at positions 662–673, 691–699,722–737, and 1193 of the reference, respectively) (Table 3).

Phylogenetic analyses were performed in a dynamic homology framework (Wheeler 1996; Wheeler 2001), using the program POY v5 (Wheeler et al. 2005), to properly account for the common occurrence of insertions/deletions. Note that this method does not rely on a prealigned set of sequences, as far as it optimizes parsimony by exploring the space of possible sites and sizes of indels. The costs of nucleotide substitutions, gap openings, and gap extensions in the alignment were adjusted to 1, 2, and 1, respectively, by means of a number of preliminary runs and our previous experience (Montalbano Di Filippo et al. 2017). The program was run under a simple, maximum parsimony search. Three sequences attributable to the genus *Echinamoeba* (AF293895, GU290114, and one available in-house), belonging to the recently robustly established lineage Tubulinea as *Vermamoeba* (Kang et al. 2017), were used as out-groups. In the run, the tree with optimal parsimony and nine additional ones with decreasing parsimony (cost) and unique topology were retained and examined. The implicit alignments were saved as FASTA files and examined, to associate branches to particular variations (the alignment of the most parsimonious tree is available in Online Resource 1, in Supplementary Material). Node support was calculated by bootstrap with 100 replicates.

## Results

From the 57 samples analyzed, 40 (70.17%) were positive in the culture examination for the presence of amoebic cells. Twenty-two out of the 40 samples (55%) turned out to be attributable to *V. vermiformis* upon 18S rRNA gene sequencing (Table 1). Together with five isolates (see Table 2) previously described by us (Montalbano Di Filippo et al. 2015), this brings to 27 the number of sequences recovered from Italian sources available for further analyses.

Based on pairwise distances, these 27 isolates could be grouped into six clusters: a first cluster included three identical sequences (11T, 14T, and 15T) recovered from hot thermal waters and are different from the others by at least three single nucleotide substitutions (SNS); a second cluster, formed by 10 identical sequences recovered from the single clinical sample and different sources of tap waters and dental units waterlines; a third cluster of two identical sequences, from two geographically distant sources of waters (4R, PUGL83F); a fourth cluster of two identical sequences from the same source sampled at different times (PUGL84F-78F); a fifth cluster of two identical sequences (13T, 17T) from the same hot thermal waters as above; and a final cluster of eight identical sequences, all from both hot springs reported in "Materials and methods" (see Online Resource 2 in Supplementary Material).

In order to assign the above sequences to subtypes, we analyzed them in the context of a total of 74 sequences, 34 of which were previously assigned to seven *V. vermiformis* subtypes.

A maximum parsimony analysis performed under dynamic homology resulted in three equally parsimonious trees (see Online Resource 3 - Supplemental Fig. 1, in Supplementary Material), to be discussed below. In all these trees, sequences were clustered based on their similarity. The corresponding implicit alignment spanned 1626 bp. In this alignment, the out-group sequences (*Echinamoeba*) formed a clade sister to all the remaining sequences in agreement with Kang et al. (2017). Two unusually long sequences (DQ123623, EU137741) were reported by "Fuerst 2017" with allelic codes 6-5-1-1 and 7-4-4-1, respectively, assigned to group 6, and shared with four additional *V. vermiformis* sequences (KP233873-76), a 983 bp insertion in between positions 1188 and 1189 of the reference. This position coincided with that reported by Thomas, V. and Greub, G. (in accession EU137741) and considered a group I intron. Moreover, KP233874 and KP233875 carried the novel arrangement TCAC-TCGCGAGG-GTGG in the variable segment F. The remaining portion of the alignment, represented in all 74 sequences, thus consisted of 643 bp (hereafter "non-intronic").

As to diversity, the non-intronic portion of the alignment included 60 positions variable for SNS and 40 positions variable for indels. This portion of the alignment encompassed the variable segments E5-E3-F-G, amounting to 40 bp. Of the

**Table 2** List of the 74 sequences used in this study, with details on the classification and source of the isolates

| Initial classification | Isolate name | Source | Country | Accession number | Allelic code; group[a] | Database |
|---|---|---|---|---|---|---|
| *H. vermiformis* | CRIB06 | Hospital water | Switzerland | DQ123623 | 6-5-1-1; 6 | "Fuerst 2017" |
| | CRIB19 | Seine river | France | EU137741 | 7-4-1-1; 6 | |
| | Costa_Rica | | Costa Rica | AY680840 | 1-1-1-1; 1 | |
| | | Freshwater | USA | M95168 | 1-1-1-1; 1 | |
| | ECH26 | Haemolymph of *Sphaerechinus granularis* | Croatia | JQ271688 | 2-2-1-1; 2 | |
| | 4480 | Spleen of *Cyprinus carpio* | Czech Republic | DQ084363 | 2-2-1-1; 2 | |
| | 4391 | Liver of *Tinca tinca* | | DQ084364 | 2-2-1-1; 2 | |
| | TN102 | Kidney of *Oreochromis niloticus* | | DQ084365 | 2-2-1-1; 2 | |
| | 4394 | Liver of *Scardinius erythrophthalmus* | | JQ271687 | 2-2-1-1; 2 | |
| | PFG | Gills of *Perca fluviatilis* | | DQ084366 | 2-2-1-1; 2 | |
| | PM11 | *Pectinatella magnifica* | | JQ271689 | 3-3-1-1; 5 | |
| | CDC_19 | Hospital cooling tower | USA | X75513 | 2-2-1-2; ½ | |
| | OS_101 | – | Germany | X75514 | 3-2-1-2; 4 | |
| | CCAP_1534/7B | – | UK | X75515 | 3-3-3-2; 4 | |
| | KWR3 | Drinking water | Netherlands | AY502961 | 5-2-2-1; 3a | |
| | KWR2 | | | AY502960 | 5-2-2-1; 3a | |
| | KWR1 | | | AY502959 | 4-2-2-1; 3a | |
| | C3/8 | Water reservoir | Germany | AF426157 | 3-3-1-1; 4 | |
| *Hartmannella* sp. | GERF1 | Gills of *Oncorhynchus mykiss* | Germany | HM363626 | 2-2-5-1; 2 | |
| | GERF2 | | | HM363627 | 2-2-1-1; 2 | |
| | Mbc_3H | Rice field | Italy | AB425953 | 3-3-1-1; 5 | |
| | FS5 | Nitrifying bioreactor | Austria | KF697197 | 8-3-1-1; 5 | |
| | HCLCCOF001 | Ocular keratitis | Spain | EF205324 | 3-3-3-2; 4 | |
| | T33 | Intestine of *Rana catesbeiana* | USA | EF378671 | *4-2-1-1; new* | GenBank |
| *V. vermiformis* | *3T* | Geothermal springs (site A) | Central Italy | NS[b] | *5-3-2-1; 3a* | Present study |
| | *11T* | | | | *7-5-1-1; new* | |
| | *12T* | Geothermal springs (site B) | | | *5-2-2-1; 3a* | |
| | *13T* | Geothermal springs (site A) | | MK110504 | *4-2-1-1; new* | |
| | *14T* | | | NS[b] | *7-5-1-1; new* | |
| | *15T* | Geothermal springs (site B) | | MK110505 | *7-5-1-1; new* | |
| | *16T* | | | NS[b] | *5-2-2-1; 3a* | |
| | *17T* | | | | *4-2-1-1; new* | |
| | *18T* | Geothermal springs (site A) | | | *5-2-2-1; 3a* | |
| | *19T* | | | | *5-2-2-1; 3a* | |
| | *20T* | | | | *5-2-2-1; 3a* | |
| | *21T* | | | | *5-2-2-1; 3a* | |
| | *22T* | | | | *5-2-2-1; 3a* | |
| | *1R* | Dental unit waterline | Northern Italy | | *1-1-1-1; 1* | |
| | *4R* | | | | *3-2-1-1; 4* | |
| | *5R* | | | | *1-1-1-1; 1* | |
| | *6R* | | | | *1-1-1-1; 1* | |
| | *7R* | | | | *1-1-1-1; 1* | |
| | *8R* | | | | *1-1-1-1; 1* | |
| | *9R* | | | | *1-1-1-1; 1* | |
| | *2S* | Tap water | Central Italy | | *1-1-1-1; 1* | |
| | *23C* | Corneal scraping | Italy | | *1-1-1-1; 1* | |
| | LAZ9TW | Tap water | Central Italy | KP792378 | *1-1-1-1; 1* | GenBank |
| | LAZ3TW | | | KP792377 | *1-1-1-1; 1* | |

**Table 2** (continued)

| Initial classification | Isolate name | Source | Country | Accession number | Allelic code; group[a] | Database |
|---|---|---|---|---|---|---|
| | PUGL83F | Fountain water | South Italy | KP792383 | *3-2-1-1; 4* | |
| | PUGL84F | | | KP792384 | *8-3-1-1; 5* | |
| | PUGL78F | | | KP792381 | *8-3-1-1; 5* | |
| | BA/SWH1/IR | Park | Iran | KP233873 | *6-2-1-1; new* | |
| | BA/SWH2/IR | | | KP233874 | *6-2-new-1; new* | |
| | BA/SWH3/IR | Swimming pool | | KP233875 | *6-2-new-1; new* | |
| | BA/SWH4/IR | | | KP233876 | *6-4-4-1; 6* | |
| | | Drinking water | France | KT290031 | *1-1-1-1; 1* | |
| | MG1 | Freshwater | South Korea | KU519742 | 2-2-1-1; 2 | "Fuerst 2017" |
| | BCP-EM3VF21-2 | Mixed with green algae | USA | KT185625 | 2-2-1-1; 2 | |
| | JG10.99 | | | JQ519505 | 2-2-1-1; 2 | |
| | FP9 | Swimming pool | Pakistan | MF112024 | *3-2-1-1; 4* | GenBank |
| | SSH12 | Soil | Spain (Canary Islands) | KU746978 | *3-2-1-1; 4* | |
| | SSH23 | | | KU746981 | *3-2-1-1; new* | |
| | SSH16 | | | KU746979 | *3-2-1-1; new* | |
| | SSH13 | | | KU746977 | *3-2-1-1; new* | |
| | SSH19 | | | KU746980 | *3-2-1-1; new* | |
| | CCAP 1534/16 | Dental unit waterline | England | KC161965 | 5-2-1-1; 3b | "Fuerst 2017" |
| | CCAP 1534/17 | | | KC188996 | 5-2-1-1; 3b | |
| | Wang17 | Mixed with mass culture of *Scenedesmus* sp. | China | KY476315 | 5-2-1-1; 3b | |
| Uncultured marine eukaryote | M2_18C04 | Environmental | Denmark | DQ103830 | 1-1-1-1; 1 | |
| | SIF_2F3 | | Norway | EF526997 | 1-1-1-1; 1 | |
| Uncultured amoeba | Cher4_1E_58 | | Ukraine | JN020235 | 3-3-1-1; 5 | |
| Uncultured eukaryote | HRT3hrs | Wastewater | Thailand | LC222939 | *4-2-1-1; new* | GenBank |
| Amoebozoa sp. | LPG1 | Water pond | USA | KT892702 | 3-2-1-1; 4 | "Fuerst 2017" |
| *Auxenochlorella pyrenoidosa* | HIT9 | | China | MF040792 | 3-3-1-1; 5 | |

[a] Sequence allelic variations and subgroup classification according to "Fuerst 2017." Allelic variations determined in the present study and the corresponding inferred groups are in italic

[b] Not submitted as redundant

60 SNS variable sites overall, 14 fell within these variable segments, confirming a contribution to the overall diversity above the average ("Fuerst 2017"). The pairwise identity across this portion ranged between 100 and 96.7%.

As to the intron, 167 SNS were observed, plus 11 indels amounting to 17 positions. The minimum identity among the six sequences was 83%. Notably, the same sequences displayed an identity above 99% in their non-intronic portion.

Sequence analysis of our 27 isolates in the context of the entire dataset (Fig. 1) revealed:

(i) Three identical isolates (11T, 14T, and 15T; top of Fig. 1) from hot thermal waters were novel. These sequences were characterized by the new allelic code 7-5-1-1, never reported before; this novel sequence type was deposited in GenBank under accession number MK110505 (isolate 15T).

(ii) Ten isolates (1R, 5R, 6R, 7R, 8R, 9R, 2S, 23C, and KP792377-8) from different sources, with allelic code 1-1-1-1, showed 100% identity with sequences AY680840 and M95168 and were thus assigned to group 1. This cluster included three additional sequences, sharing the code 1-1-1-1 but diverging at up to eight positions, two of which (DQ10830, EF526997) were already assigned to group 1 (see Online Resource 3 - Supplemental Fig. 2, in Supplementary Material).

(iii) Two identical isolates (4R, KP792383) with allelic code 3-2-1-1 were also identical to KT892702, assigned to group 4—"Fuerst 2017", and to two additional *V. vermiformis* sequences (MF112024, KU746978).

(iv) Two identical isolates (KP792384, KP792381) shared allelic code 8-3-1-1 but differed by other six substitutions from KF697197, a variant sequence within group 5—"Fuerst 2017."

**Table 3** Variable nucleotides over the sequence of the *V. vermiformis* 18S rRNA and the corresponding genetic types assigned to each segment (E5-E3-F-G) as reported by "Fuerst 2017." By combining the different genetic types, it is possible to identify multiple allelic combinations (allelic variants), condensed into a 4-digit allelic code (Reference sequence M95168: allelic code 1-1-1-1)

| | Variable segments | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | E5 | | | | | | E3 | | | | | F | | | | | | | | G | |
| | Nucleotide positions in M95168 | | | | | | | | | | | | | | | | | | | | |
| | | 664 | 667 | 669 | 670 | 671 | | 693 | 694 | 695 | 697 | | 724 | 725 | 726 | 728 | 730 | 733 | 735 | | 1193 |
| Genetic types | 1 | T | A | T | G | G | 1 | T | C | A | T | 1 | A | T | – | G | G | – | T | 1 | C |
| | 2 | T | A | C | A | G | 2 | T | T | G | T | 2 | G | T | – | G | G | – | C | 2 | – |
| | 3 | C | A | C | A | G | 3 | C | T | G | T | 3 | A | T | – | C | G | – | T | | |
| | 4 | C | A | C | A | A | 4 | T | C | G | A | 4 | C | C | G | G | G | C | G | | |
| | 5 | C | A | T | A | A | 5 | T | C | G | T | 5 | A | T | – | G | A | – | T | | |
| | 6 | C | T | C | G | G | | | | | | | | | | | | | | | |
| | 7 | C | A | C | G | G | | | | | | | | | | | | | | | |
| | 8 | C | A | T | A | G | | | | | | | | | | | | | | | |
| | 9 | C | C | C | G | G | | | | | | | | | | | | | | | |

(v) Two identical isolates from hot thermal waters (13T, 17T) with allelic code 4-2-1-1, not previously reported by "Fuerst 2017," showed 100% identity with an uncultured eukaryote isolate HRT3 from Bangkok (LC222939). All these sequences shared the 719T>C variant. This novel sequence type was deposited in GenBank under accession number MK110504 (isolate 13T).

(vi) Eight identical isolates (3T, 12T, 16T, 18T, 19T, 20T, 21T, 22T) with allelic code 5-2-2-1 and all from hot thermal waters exhibited also 100% identity with the *H. vermiformis* isolate KWR2 (AY502960) and were thus assigned to group 3a—"Fuerst 2017." Within the same cluster, a single nucleotide substitution (807A>G) differentiated AY502961. Another single nucleotide substitution, within the variable segment E5 (669C), differentiated AY502959 and generated the allelic code 4-2-2-1.

## Discussion

According to the recent classification of Amoebozoa, the species *Vermamoeba vermiformis* is now related with *Echinamoeba* spp. within the clade Echinamoebidia (Cavalier-Smith et al. 2016; Kang et al. 2017). This amoeba has been recovered at a global scale in natural freshwater environments, including thermal springs, man-made environments, and engineered water systems (water treatment, distribution systems, and tap water installations). In contrast with this common occurrence, little is known about the prevalence of amoebic keratitis caused by *Vermamoeba* worldwide and its epidemiology at a local scale.

Here, we report 22 new isolates attributable to *V. vermiformis* from different sources in Italy and characterized as different types on the basis of the DNA sequence in the central portion of the 18S rRNA gene, as described by "Fuerst 2017." Though this region encompasses only four of the 11 variable segments reported so far for the entire 18S rRNA, we notice that they can include a large part of the diversity information to cluster the isolates in a replicable fashion. By assembling a dataset of 74 overlapping sequences, we proceeded one step further in the description of the pattern of similarities among the sequences, summarized in the tree of Fig. 1.

There are strong and weak features in this chart, which deserve discussion.

As to the strong features, though we used a reduced span of the sequenced segment and a different tree building method, the 34 sequences derived from "Fuerst 2017" were clustered as in the original report, with the exception of a group 4 sequence (AF426157) mapping in an otherwise group 5 cluster and a sequence of uncertain group assignment (X77513) here mapping in a group 4 cluster. This shows that the variable segments E5, E3, F, and G convey most of the information on diversity. Accordingly, only in three cases in each of our clusters corresponding to groups 2, 4/5, and 3b, respectively, sequences identical in this study were instead distinguishable based on the complete sequence (see tree in "Fuerst 2017").

By searching GenBank, we also expanded the set of sequences characterized by a ~1000 bp insertion. Our reconstruction placed the inserted DNA at an homologous position

**Fig. 1** Maximum parsimony cladogram obtained with POY on 74 *Vermamoeba* (*Hartmannella*) sequences, out-grouped with three *Echinamoeba* sequences. Branches are fixed-length, irrespective of the number of substitution(s) corresponding to each of them. Sequences from the literature are identified by their accession number; the reference M95168 is highlighted. Sequences obtained in this work are bolded and dotted. When available, the nomenclature of genotypes according to "Fuerst 2017" is boxed. For branches defining the main clades, the corresponding mutational events, or the type at the variable segments E5-E3-F-G according to "Fuerst 2017," are reported. Bootstrap support values are italicized within boxes at the corresponding nodes. Note the basal position of sequences 11T, 14T, and 15T and the match between genotype classification and topological consistency

in the six sequences and revealed an internal identity of 83% or higher. These results point strongly to a unique event (monophyly) for the insertion, followed by an accelerated divergence which contrasts sharply with the strong conservation of the non-intronic portion.

The six sequences above, together with three novel sequences (11T, 14T, and 15T), reported here for the first time, share nucleotide states with our out-groups, i.e., the 664(C)-669(C) and the 694(C)-695(G)-696(C) arrangements in E5 and E3, respectively. These variants locate the corresponding sequences in basal branches of all the equally parsimonious trees (top in Fig. 1 and Online Resource 3 - Supplemental Fig. 1, in Supplementary Material). At the opposite end, we find branches with sequences of groups 3b and 3a. The trees also include robust clades defined by substitutions 719T>C and 934A>T, not represented in the "Fuerst 2017" dataset.

One weak aspect of the tree is the low support for branches in between these extremes. In particular, the relative order of codes 1-1-1-1, 2-2-1-1, 3-2-1-1, and 3-3-1-1, associated with groups 1, 2, 4, and 4/5, respectively, remains unresolved. In the tree by “Fuerst 2017,” too, nodes basal to these clusters received inconclusive support.

This survey, carried out on a diverse array of potential sources, recovered *V. vermiformis* from a remarkable proportion of samples, in line with previous reports in the literature (Delafont et al. 2018). As to the diversity of the samples here analyzed, the 27 Italian isolates were characterized for the first time at the level of sequence type and mapped to different clades of the tree. In particular, isolates from both geographically close thermal springs were mapped to three different clusters of sequences. The presence of different strains of *Vermamoeba* in these small natural areas could indicate environmental stability of the sites, allowing to preserve genetic diversity of the amoebic populations.

Also, the isolates from one clinical sample and from a variety of water distribution systems (e.g., public fountains, dental units, and taps/showers), derived from a geographically widespread sampling, display diversity. Ours and the only other *Vermamoeba* sequence from human infection, included in the “Fuerst 2017” dataset, were mapped in different clades of the tree. Isolates recovered from dental units and tap water, both from managed waters, also confirmed multiple sequence types and hence heterogeneity of the amoebic populations.

Finally, identical sequences were recovered from independent stations located at > 800 km from each other. In the context of the entire dataset, with few exceptions, the types defined by the sequences here analyzed are distributed across huge distances, often transcontinental. Confinement to specific natural or man-made environments remains to be investigated by means of specific sampling designs.

In conclusion, our dataset revealed a novel quota of diversity which enlarged the repertoire of single nucleotide and indel variants and the list of allelic codes. Our experimental procedure cannot exclude sequence heterogeneity among the multiple copies of the *V. vermiformis* 18S rRNA gene (Kuiper et al. 2006). However, the background noise in sequence profiles of isolates with and without the intronic insertion points to a negligible quota of variant copies, if any.

While sequences previously assigned to each of the seven groups by “Fuerst 2017” are largely clustered together in our tree, we prompt caution in simply equating an allelic code with group affiliation. In fact, codes at E5, E3, and F often consist in a variation at a single position, mostly a transition, i.e., an event with a non-negligible probability of recurrence. We thus recommend using a complete set of variants as possible within the available sequenced segment(s) to determine the similarity between isolates and their assignment to the same/different clades.

**Funding** This study was partially funded by the University of Rome Tor Vergata, Italy (Ricerca Scientifica di Ateneo 2016 “Mission: Sustainability”).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable standards.

## References

Aitken D, Hay J, Kinnear FB, Kirkness CM, Lee WR, Seal DV (1996) Amebic keratitis in a wearer of disposable contact lenses due to a mixed *Vahlkampfia* and *Hartmannella* infection. Ophthalmology 103:485–494

Cavalier-Smith T, Chao EE, Lewis R (2016) 187-gene phylogeny of protozoan phylum Amoebozoa reveals a new class (Cutosea) of deep-branching, ultrastructurally unique, enveloped marine Lobosa and clarifies amoeba evolution. Mol Phylogenet Evol 99:275–296

Centeno M, Rivera F, Cerva L, Tsutsumi V, Gallegos E, Calderón A, Ortiz R, Bonilla P, Ramirez E, Suarez G (1996) *Hartmannella vermiformis* isolated from the cerebrospinal fluid of a young male patient with meningoencephalitis and bronchopneumonia. Arch Med Res 27:579–586

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31:3497–3500

Delafont V, Rodier MH, Maisonneuve E, Cateau E (2018) *Vermamoeba vermiformis*: a free-living amoeba of interest. Microb Ecol 76:991–1001. https://doi.org/10.1007/s00248-018-1199-8

Gunderson JH, Goss SJ, Sogin ML (1994) The sequence of the *Hartmannella vermiformis* small subunit rRNA coding region. J Eukaryot Microbiol 41:481–482

Inoue T, Asari S, Tahara K, Hayashi K, Kiritoshi A, Shimomura Y (1998) *Acanthamoeba* keratitis with symbiosis of *Hartmannella* ameba. Am J Ophthalmol 125:721–723

Kang S, Tice AK, Spiegel FW, Silberman JD, Pánek T, Cepicka I, Kostka M, Kosakyan A, Alcântara DMC, Roger AJ, Shadwick LL, Smirnov A, Kudryavtsev A, Lahr DJG, Brown MW (2017) Between a pod and a hard test: the deep evolution of amoebae. Mol Biol Evol 34:2258–2270

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kuiper MW, Valster RM, Wullings BA, Boonstra H, Smidt H, van der Kooij D (2006) Quantitative detection of the free-living amoeba *Hartmannella vermiformis* in surface water by using real-time PCR. Appl Environ Microbiol 72:5750–5756

Lorenzo-Morales J, Martínez-Carretero E, Batista N, Alvarez-Marín J, Bahaya Y, Walochnik J, Valladares B (2007) Early diagnosis of amoebic keratitis due to a mixed infection with *Acanthamoeba* and *Hartmannella*. Parasitol Res 102:167–169

McGinnis S, Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 32:W20–W25

Montalbano Di Filippo M, Santoro M, Lovreglio P, Monno R, Capolongo C, Calia C, Fumarola L, D’Alfonso R, Berrilli F, Di Cave D (2015)

Isolation and molecular characterization of free-living amoebae from different water sources in Italy. Int J Environ Res Public Health 12:3417–3427

Montalbano Di Filippo M, Novelletto A, Di Cave D, Berrilli F (2017) Identification and phylogenetic position of *Naegleria* spp. from geothermal springs in Italy. Exp Parasitol 183:143–149

Niyyati M, Rahimi F, Lasejerdi Z, Rezaeian M (2014) Potentially pathogenic free–living amoebae in contact lenses of the asymptomatic contact lens wearers. Iranian J Parasitol 9:14–19

Smirnov A, Nassonova E, Berney C, Fahrni J, Bolivar I, Pawlowski J (2005) Molecular phylogeny and classification of the lobose amoebae. Protist 156:129–142

Smirnov AV, Chao E, Nassonova ES, Cavalier-Smith T (2011) A revised classification of naked lobose amoebae (Amoebozoa: lobosa). Protist 162:545–570

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Tsvetkova N, Schild M, Panaiotov S, Kurdova-Mintcheva R, Gottstein B, Walochnik J, Aspock H, Lucas MS, Muller N (2004) The identification of free-living environmental isolates of amoebae from Bulgaria. Parasitol Res 92:405–413

Wheeler WC (1996) Optimization alignment: the end of multiple sequence alignment in phylogenetics. Cladistics 12:1–9

Wheeler WC (2001) Homology and the optimization of DNA sequence data. Cladistics 17:S3–S11

Wheeler WC, Aagesen L, Arango CP, Faivovich J, Grant T, D'Haese C, Janies D, Smith WL, Varón A, Giribet G (2005) Dynamic homology and phylogenetic systematics: a unified approach using POY. American Museum of Natural History, New York