A new view on the morphology and phylogeny of eugregarines suggested by the evidence from the gregarine *Ancora sagittata* (Leuckart, 1860) Labbé, 1899 (Apicomplexa: Eugregarinida)

Timur G Simdyanov [Corresp., 1] , **Laure Guillou** [2, 3] , **Andrei Y Diakin** [4] , **Kirill V Mikhailov** [5, 6] , **Joseph Schrével** [7, 8] , **Vladimir V Aleoshin** [5, 6]

1 Department of Invertebrate Zoology, Faculty of Biology, Lomonosov Moscow State University, Moscow, Russian Federation

2 CNRS, UMR 7144, Laboratoire Adaptation et Diversité en Milieu Marin, Roscoff, France

3 CNRS, UMR 7144, Station Biologique de Roscoff, Sorbonne Universités, Université Pierre et Marie Curie - Paris 6, Roscoff, France

4 Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic

5 Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russian Federation

6 Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russian Federation

7 CNRS 7245, Molécules de Communication et Adaptation Moléculaire (MCAM), Paris, France

8 Sorbonne Universités, Muséum National d'Histoire Naturelle (MNHN), UMR 7245, Paris, France

Corresponding Author: Timur G Simdyanov
Email address: tgsimd@gmail.com

Background. Gregarines are a group of early branching Apicomplexa parasitizing invertebrate animals. Despite their wide distribution and relevance to the understanding the phylogenesis of apicomplexans, gregarines remain understudied: light microscopy data are insufficient for classification, and electron microscopy and molecular data are fragmentary and overlap only partially.

Methods. Scanning and transmission electron microscopy, PCR, DNA cloning and sequencing (Sanger and NGS), molecular phylogenetic analyses using ribosomal RNA genes (18S (SSU), 5.8S, and 28S (LSU) ribosomal DNAs (rDNAs)).

Results & Discussion. We present the results of an ultrastructural and molecular phylogenetic study on the marine gregarine *Ancora sagittata* from the polychaete *Capitella capitata* followed by evolutionary and taxonomic synthesis of the morphological and molecular phylogenetic evidence on eugregarines. The ultrastructure of *A. sagittata* generally corresponds to that of other eugregarines, but reveals some differences in epicytic folds (crests) and attachment apparatus to gregarines in the family Lecudinidae, where *A. sagittata* has been classified. Molecular phylogenetic trees based on SSU (18S) rDNA reveal several robust clades (superfamilies) of eugregarines, including Ancoroidea superfam. nov., which comprises two families (Ancoridae fam. nov. and Polyplicariidae) and branches separately from the Lecudinidae; thus, all representatives of Ancoroidea are here officially removed from the Lecudinidae. Analysis of sequence data also points to possible cryptic species within *A. sagittata* and the inclusion of numerous environmental sequences from anoxic habitats within the Ancoroidea. LSU (28S) rDNA phylogenies, unlike the analysis of SSU rDNA alone, recover a well-supported monophyly of the gregarines involved (eugregarines), although this conclusion is currently limited by sparse taxon sampling and the presence of fast-evolving sequences in some species. Comparative morphological analyses of gregarine teguments and attachment organelles lead us to revise their terminology. The terms "longitudinal folds" and "mucron" are restricted to archigregarines, whereas the terms "epicystic crests" and "epimerite" are proposed to describe the candidate synapomorphies of eugregarines, which,

consequently, are considered as a monophyletic group. Abolishing the suborders Aseptata and Septata, incorporating neogregarines into the Eugregarinida, and treating the major molecular phylogenetic lineages of eugregarines as superfamilies appear as the best way of reconciling recent morphological and molecular evidence. Accordingly, the diagnosis of the order Eugregarinida Léger, 1900 is updated.

1 **A new view on the morphology and phylogeny of eugregarines suggested by the evidence**

2 **from the gregarine *Ancora sagittata* (Leuckart, 1860) Labbé, 1899 (Apicomplexa:**

3 **Eugregarinida)**

4

5 Timur G. Simdyanov[1], Laure Guillou[2,3], Andrei Y. Diakin[4], Kirill V. Mikhailov[5,6], Joseph

6 Schrével[7,8], and Vladimir V. Aleoshin[5,6,9]

7

8 [1] Department of Invertebrate Zoology, Faculty of Biology, Lomonosov Moscow State

9 University, Moscow 119234, Russian Federation

10 [2] CNRS, UMR 7144, Laboratoire Adaptation et Diversité en Milieu Marin, Roscoff, France

11 [3] Sorbonne Universités, Université Pierre et Marie Curie - Paris VI, CNRS, UMR 7144, Station

12 Biologique de Roscoff, Roscoff, France

13 [4] Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech

14 Republic

15 [5] Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University,

16 Moscow, Russian Federation

17 [6] Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow,

18 Russian Federation

19 [7] CNRS 7245, Molécules de Communication et Adaptation Moléculaire (MCAM), Paris, France

20 [8] Sorbonne Universités, Muséum National d'Histoire Naturelle (MNHN), UMR 7245, Paris,

21 France

22 [9] Institute of Animal Physiology, Biochemistry and Nutrition, Borovsk, Kaluga region, Russian

23 Federation

24

25 Corresponding author:

26 Timur Simdyanov[1]

27 Leninskiye Gory 1-12, Moscow, 119234, Russian Federation

28 E-mail address: tgsimd@gmail.com

29 **Abstract**

30 **Background.** Gregarines are a group of early branching Apicomplexa parasitizing invertebrate

31 animals. Despite their wide distribution and relevance to the understanding the phylogenesis of

32 apicomplexans, gregarines remain understudied: light microscopy data are insufficient for

33 classification, and electron microscopy and molecular data are fragmentary and overlap only

34 partially.

35 **Methods.** Scanning and transmission electron microscopy, PCR, DNA cloning and sequencing

36 (Sanger and NGS), molecular phylogenetic analyses using ribosomal RNA genes (18S (SSU),

37 5.8S, and 28S (LSU) ribosomal DNAs (rDNAs)).

38 **Results & Discussion.** We present the results of an ultrastructural and molecular phylogenetic

39 study on the marine gregarine *Ancora sagittata* from the polychaete *Capitella capitata* followed

40 by evolutionary and taxonomic synthesis of the morphological and molecular phylogenetic

41 evidence on eugregarines. The ultrastructure of *A. sagittata* generally corresponds to that of other

42 eugregarines, but reveals some differences in epicytic folds (crests) and attachment apparatus to

43 gregarines in the family Lecudinidae, where *A. sagittata* has been classified. Molecular

44 phylogenetic trees based on SSU (18S) rDNA reveal several robust clades (superfamilies) of

45 eugregarines, including Ancoroidea superfam. nov., which comprises two families (Ancoridae

46 fam. nov. and Polyplicariidae) and branches separately from the Lecudinidae; thus, all

47 representatives of Ancoroidea are here officially removed from the Lecudinidae. Analysis of

48 sequence data also points to possible cryptic species within *A. sagittata* and the inclusion of

49 numerous environmental sequences from anoxic habitats within the Ancoroidea. LSU (28S)

50 rDNA phylogenies, unlike the analysis of SSU rDNA alone, recover a well-supported

51 monophyly of the gregarines involved (eugregarines), although this conclusion is currently

52 limited by sparse taxon sampling and the presence of fast-evolving sequences in some species.

53 Comparative morphological analyses of gregarine teguments and attachment organelles lead us

54 to revise their terminology. The terms "longitudinal folds" and "mucron" are restricted to

55 archigregarines, whereas the terms "epicystic crests" and "epimerite" are proposed to describe

56 the candidate synapomorphies of eugregarines, which, consequently, are considered as a

57 monophyletic group. Abolishing the suborders Aseptata and Septata, incorporating

58 neogregarines into the Eugregarinida, and treating the major molecular phylogenetic lineages of

59 eugregarines as superfamilies appear as the best way of reconciling recent morphological and

60   molecular evidence. Accordingly, the diagnosis of the order Eugregarinida Léger, 1900 is

61   updated.

62 **Introduction**

63

64 The Apicomplexa is a group of unicellular eukaryotes within the Alveolata encompassing

65 parasites of humans and animals. Some apicomplexans are well studied (e.g., human pathogens

66 such as *Plasmodium*, *Toxoplasma*, and *Cryptosporidium*), while early branching representatives

67 such as gregarines, are far less well known. Gregarines are obligate parasites of invertebrate

68 animals: various groups of worms, molluscs, arthropods (aquatic and terrestrial), echinoderms,

69 and tunicates. The large majority of gregarines are monoxenous (have a single invertebrate host)

70 and parasitize in the gut of their hosts, where they are commonly found as epicellular feeding

71 stages, the trophozoites, which are conspicuous due to their large size (usually ranging from 200

72 to 600 μm). Because of their minor economic importance, gregarines are poorly studied despite

73 their widespread distribution and relevance to the reconstruction of the evolutionary history of

74 apicomplexans.

75 The taxonomy and phylogeny of gregarines remains largely incomplete (Grassé, 1953; Levine,

76 1985; Levine, 1988; Perkins et al., 2000) due to uneven scrutiny: light microscopic data cannot

77 sustain a reliable classification, electron microscopy and molecular phylogenetic data are

78 fragmentary, and, additionally, sets of features that have been examined using different methods

79 overlap only partially (see Discussion). Gregarine orders differ by their life cycles, which include

80 sexual (gamogony) and asexual (merogony and sporogony) reproductions. Sexual reproduction

81 in gregarines (Grassé, 1953; Schrével et al., 2013) is initiated by syzygy (the association of

82 gamonts, usually two of them) and followed by the production of a surrounding gametocyst,

83 which is typical only for gregarines and likely represents a synapomorphy for the group (Frolov,

84 1991). The large majority of gregarines, which are classified within the order Eugregarinida

85 Léger, 1900, have lost merogony, while some others (former order Schizogregarinida Léger,

86 1900) retain it.

87 The most productive taxonomical scheme of the gregarines is based on Grassé's hypothesis about

88 their co-evolution with their hosts (Grassé, 1953). Grassé divided Schizogregarinida into two

89 orders: Archigregarinida Grassé, 1953, and Neogregarinida Grassé, 1953. Archigregarines

90 parasitize marine invertebrates, mainly polychaetes and sipunculids. Neogregarines parasitize

91 insects (intestine, Malpighian tubules, and fat body) and Grassé suggested that they are derived

92 from various representatives of the eugregarine family Actinocephalidae (parasites of insects), by

93   the secondary gain of merogony. The third order, the already mentioned Eugregarinida Léger,

94   1900, is the most diverse group of gregarines infecting a broad range of invertebrate hosts.

95   The current gregarine classification (e.g., Levine, 1988; Perkins et al., 2000) relies chiefly on the

96   light-microscopy of trophozoites and life cycle features (absence or presence of merogony),

97   discarding Grassé's co-evolutionary approach. It also ignores results of SEM and TEM studies,

98   which have revealed distinct differences between Grassé's gregarine orders in the structure of the

99   cortex and attachment apparatus, especially between archi- and eugregarines (Schrével, 1968;

100   Vivier, 1968; Vavra & Small, 1969; Vivier et al., 1970; Schrével, 1971; Simdyanov &

101   Kuvardina, 2007; Schrével et al., 1983; also see Discussion). As a result, a portion of the

102   archigregarines and even blastogregarines (Sporozoa *incertae sedis* after Grassé) were reassigned

103   to eugregarines (Levine, 1985; Levine, 1988), which in turn were divided into two main

104   suborders: Septata Lankester, 1885 and Aseptata Chakravarty, 1960. Aseptate eugregarines (e.g.,

105   the families Lecudinidae and Urosporidae) chiefly infect marine invertebrates and considered

106   plesiomorphic representatives of the order (Grassé, 1953; Perkins et al., 2000; Schrével &

107   Desportes, 2013b). Septate gregarines are widespread parasites of aquatic and terrestrial

108   arthropods and considered evolutionarily derived: they possess one or more light-refracting

109   septum, which separates the trophozoite into two compartments: a smaller protomerite and larger

110   deutomerite, where the nucleus is located.

111   Molecular phylogenetic studies of gregarines are limited in sampling and largely rely on small

112   subunit (SSU or 18S) ribosomal DNA (rDNA) sequences (Carreno, Martin & Barta, 1999;

113   Leander, Clopton & Keeling, 2003; Leander, Harper & Keeling, 2003; Leander et al., 2006;

114   Leander, 2007; Rueckert & Leander, 2008; Clopton, 2009; Rueckert & Leander, 2009; Rueckert,

115   Chantangsi & Leander, 2010; Rueckert & Leander, 2010; Rueckert et al., 2011; Rueckert,

116   Villette & Leander, 2011; Wakeman & Leander, 2012; Rueckert, Wakeman & Leander, 2013;

117   Wakeman & Leander, 2013a; Wakeman & Leander, 2013b; Wakeman, Heintzelman & Leander,

118   2014; Wakeman et al., 2014; Rueckert et al., 2015; Diakin, Wakeman & Valigurová, 2017).

119   Gregarines have been also detected in environmental sequence surveys from various marine and

120   freshwater samples, possibly because oocysts are stable in the environment (Rueckert et al.,

121   2011; Janouškovec et al., 2015). A large majority of these environmental sequences cannot be

122   taxonomically assigned to a specific gregarine family. Because many gregarine SSU rDNA

123   sequences are fast evolving and form long branches in molecular phylogenies, the entire group

124 and its orders are not recognized as monophyletic. This has lead to the proposal that eugregarines

125 are polyphyletic (Cavalier-Smith, 2014) and their shared key ultrastructural characteristics have

126 been acquired convergently (see Discussion for details).

127 In this work, we characterize the aseptate eugregarine *Ancora sagittata* (Leuckart, 1861) Labbé,

128 1899, an intestinal parasite of the marine polychaete worm *Capitella capitata* Fabricius, 1780, a

129 widely distributed and abundant inhabitant of oxygen-depleted substrates. *A. sagittata* has been

130 classified as a member of Lecudinidae Kamm, 1922, the largest family of marine aseptate

131 eugregarines (containing ~30 genera and >160 named species). The taxonomy of Lecudinidae is

132 nevertheless controversial and the family may not represent a natural group (Levine, 1977;

133 Levine, 1985; Levine, 1988; Perkins et al., 2000).

134 Trophozoites of *A. sagittata* have a characteristic anchor-like appearance (Labbé, 1899; Perkins

135 et al., 2000) and their structure, growth, and development were previously observed by light

136 microscopy (Cecconi, 1905; Hasselmann, 1927). The sexual reproduction of *A. sagittata* is little

137 understood (Hasselmann, 1927) and syzygy in this species has never been observed. Neither

138 ultrastructural nor sequence data are currently available for the parasite. Here, we undertook an

139 integrated study of the *A. sagittata* morphology, ultrastructure, and molecular phylogeny by

140 using ribosomal DNA: SSU (18S), 5.8S, and LSU (28S). We revealed that *A. sagittata* represents

141 a deep molecular phylogenetic lineage of eugregarines independent of the Lecudinidae in spite of

142 their morphological similarities. This finding led us to re-evaluate and reconcile ultrastructural

143 and molecular evidence for eugregarines and, relying on this combined approach, amend

144 conventional views on eugregarine phylogeny and taxonomy.

145

146 **Materials & Methods**

147

148 **Collection, isolation, and light microscopy.** Trophozoites of *Ancora sagittata* (Leuckart, 1860)

149 Labbé, 1899 were isolated from the intestine of the polychaete worms *Capitella capitata*

150 Fabricius, 1780 collected in 2006-2011 from two sites: (i) littoral of the beach of L'Aber, the

151 coastal zone of the English Channel near Station Biologique de Roscoff, Roscoff, France

152 (48°42'45"N, 4°00' 05"W) and (ii) a sublittoral habitat at White Sea Biological Station (WSBS)

153 of Lomonosov Moscow State University, Velikaya Salma Straight, Kandalaksha Gulf of White

154 Sea, Russia (66°33′12″N, 33°06′17″E).

155   The gregarines were isolated by breaking the host body and intestine with fine tip needles under

156   a stereomicroscope (Olympus SZ40, Japan, or MBS-1, LOMO, Russia). The released parasites

157   or small fragments of host gut with attached gregarines were rinsed with filtered seawater by

158   using thin glass pipettes and then photographed under Leica DM 2000, Leica DM 2500 or Leica

159   DM5000 light microscopes with Leica DFC 420 cameras (Leica Microsystems, Germany), or

160   fixed for electron microscopy, or subjected to DNA extraction.

161   **Electron microscopy:** The structure of the gregarines *A. sagittata* from WSBS was studied by

162   scanning electron microscopy (SEM) and transmission electron microscopy (TEM). For both

163   methods, the individual gregarines or small fragments of the host gut with the attached

164   gregarines were fixed with 2.5% (v/v) glutaraldehyde in 0.05 M cacodylate buffer (pH 7.4)

165   containing 1.28% (w/v) NaCl in an ice bath in the dark. The fixative was once replaced with

166   fresh fixative after 1 hour, and the total fixation time was 2 hours. The fixed samples were rinsed

167   three times with cacodylate buffer and post-fixed with 2% (w/v) OsO_4 in the cacodylate buffer

168   (ice bath, 2 hours).

169   For SEM study, the fixed gregarines *A. sagittata* were dehydrated in a graded series of ethanol,

170   transferred to an ethanol/acetone mixture (1:1, v/v), rinsed three times with 100% acetone, and

171   critical point-dried with CO_2. The samples were mounted on stubs, sputter-coated with

172   gold/palladium, and examined under a CamScan-S2 scanning electron microscope (CamScan,

173   UK).

174   For TEM study, after dehydration in a graded series of ethanol, the fixed parasites *A. sagittata*

175   were transferred to an ethanol/acetone mixture 1:1 (v/v), rinsed twice in pure acetone, and

176   embedded in Epon resin using a standard procedure. Ultrathin sections obtained using LKB-III

177   (LKB, Sweden) or Leica EM UC6 (Leica Microsystems, Germany) ultramicrotomes were

178   contrasted with uranyl acetate and lead citrate (Reynolds, 1963) and examined under a JEM-

179   100B or a JEM 1011 electron microscope (JEOL, Japan).

180   **DNA isolation, PCR, cloning and sequencing.** After thrice-repeated rinsing with filtered

181   seawater, the gregarine trophozoites *A. sagittata* were deposited into 1.5-ml microcentrifuge

182   tubes: ~20 individuals from Roscoff (10 hosts, 2009), ~20 individuals from WSBS (4 hosts,

183   2006), ~40 individuals from WSBS (all from the single host, 2010), and ~100 individuals from

184   WSBS (10 hosts, 2011). All four samples were fixed and stored in "RNAlater" reagent (Life

185   Technologies, USA).

186 The nucleotide sequences of *A. sagittata* (SSU, 5.8S, and LSU rDNAs, as well as internal

187 transcribed spacers 1 and 2 (ITS1 and ITS2, respectively) were obtained by two methods: (i)

188 PCR followed by Sanger sequencing (Roscoff and WSBS 2006 samples) and (ii) a genome

189 amplification approach (WSBS 2010 and 2011 samples).

190 For the first method, DNA extraction was performed using the "Diatom DNA Prep 200" kit

191 (Isogen, Russia). The rDNA sequences were amplified in several PCRs with different pairs of

192 primers (Fig. 1, Table 1). As revealed later, the sample WSBS 2006 was contaminated with

193 hyperparasitic microsporidians (Mikhailov, Simdyanov & Aleoshin, 2017), which predominantly

194 reacted with the primers A and B; therefore, a specific forward primer Q5A (Table 1) was

195 constructed to provide a specific gregarine PCR product. A set of overlapping fragments

196 encompassing SSU rDNA, ITS 1 and 2, 5.8S rDNA, and LSU rDNA was obtained for each

197 sample: fragments I-IV for the sample from Roscoff and fragments V-VI for the sample from the

198 White Sea (Fig. 1). All fragments were amplified with an Encyclo PCR kit (Evrogen, Russia) in

199 a total volume of 25 µl using a DNA Engine Dyad thermocycler (Bio-Rad) and the following

200 protocol: initial denaturation at 95°C for 3 min; 40 cycles of 95°C for 30 sec, 45°C (fragments I,

201 II, and V) or 53°C (fragments II, IV, and VI) for 30 sec, and 72°C for 1.5 min; and a final

202 extension at 72°C for 10 min. Only weak bands of the expected size were obtained by

203 electrophoresis in agarose gel for fragments II and IV. Therefore, small pieces of the gel were

204 sampled from those bands (using pipette tips on a trans-illuminator), followed by re-

205 amplification with the same primers, "ColorTaq PCR kit" (Syntol, Russia), a DNA Engine Dyad

206 thermocycler (Bio-Rad), and the following PCR conditions: initial denaturation at 95°C for 1

207 min; 25 cycles of 95°C for 30 sec, 53°C for 30 sec, and 72°C for 1.5 min; and a final extension at

208 72°C for 10 min. PCR products of the expected size were gel-isolated by the Cytokine DNA

209 isolation kit (Cytokine, Russia). For the fragments I, IV, and V, the PCR products were

210 sequenced directly. The fragments II, III, and VI were cloned by using the InsTAclone PCR

211 Cloning Kit (Fermentas, Lithuania) because the corresponding PCR products were

212 heterogeneous. Sequences were obtained by using the ABI PRISM BigDye Terminator v. 3.1

213 reagent kit and the Applied Biosystems 3730 DNA Analyzer for automatic sequencing. The

214 contiguous sequences of the ribosomal operons (SSU rDNA + ITS1 + 5.8S rDNA + ITS2 + LSU

215 rDNA) were assembled for the gregarine samples from Roscoff and WSBS 2006 (GenBank

216 accession numbers KX982501and KX982502, respectively).

217   For the samples of *A. sagittata* 2010 and 2011 (WSBS), DNA extraction was performed using

218   the "NucleoSpin Tissue" kit (Macherey-Nagel). The corresponding complete ribosomal operon

219   sequences were obtained by whole genome amplification and high-throughput sequencing: ~1 ng

220   of DNA from each sample was amplified with the REPLI-g Midi kit (Qiagen) according to the

221   manufacturer's protocol and sequenced on an Illumina HiSeq2000 NGS platform using one

222   quarter of a lane in paired-end libraries, an estimated mean insert size of ~330 bp and a read

223   length of 100 bp. The Illumina reads were adapter-trimmed with Trimmomatic-0.30 (Lohse et

224   al., 2012), read pairs with reads shorter than 55 bp were discarded and the remaining reads were

225   assembled using SPAdes 2.5.0 (Bankevich et al., 2012) in single-cell mode (--sc) with read error

226   correction and five *k*-mer values: 21, 33, 55, 77, and 95. Contiguous sequences corresponding to

227   the ribosomal operon of *A. sagittata* (GenBank accession numbers KX982503 and KX982504)

228   were identified in the assembly using the standalone BLAST 2.2.25+ package (Altschul et al.,

229   1997).

230   In addition, we sequenced the LSU rDNA of the ciliate *Stentor coeruleus* to enhance the taxon

231   sampling of the LSU rDNA (GenBank accession number KX982500). Two overlapping

232   sequences was obtained using the same procedures as for the LSU rDNA fragments III and IV of

233   *A. sagittata* (PCR, cloning, and sequencing; see above for details and Table 1 for primers), and

234   the resulting contiguous sequence was assembled.

235   **Predicted secondary structures of ITS 2.** The structures were created by using MFOLD

236   (Zuker, 2003) under default parameters in the temperature range 5 – 37°C; it was made manually

237   because there was no suitable template for automatic modelling in the available databases. ITS 2

238   is a genetic region that could be a valuable marker for species delineation and compensatory base

239   changes (CBS) within it can be used to discriminate species (Coleman, 2000; Müller et al., 2007;

240   Coleman, 2009; Wolf et al., 2013).

241   **Molecular phylogenetic analyses.** Four nucleotide alignments were prepared: of SSU rDNA

242   (two variants: 114 and 52 sequences), LSU rDNA, and ribosomal operon (concatenated SSU,

243   5.8S, and LSU rDNA sequences). The alignments were generated in MUSCLE 3.6 (Edgar, 2004)

244   and manually adjusted with BioEdit 7.0.9.0 (Hall, 1999): gaps, columns containing few

245   nucleotides, and hypervariable regions were removed. The taxon sampling was designed as to

246   maximize the phylogenetic diversity and completeness of sequences in the alignments.

247   Representatives of heterokonts and rhizarians were used as outgroups. The final analysis

248   included 114 representative sequences (1,570 aligned sites).

249   To analyse sequences that are closely related to *A. sagittata*, including environmental entries

250   (from GenBank), we prepared the SSU rDNA alignment of 52 sequences for 1,709 sites; only

251   one sequence was included for each of 4 clusters of near-identical environmental clones (see

252   below). This analysis involved 139 additional nucleotides from hypervariable regions of the SSU

253   rDNA compared to the standard analysis. To assess the similarity of these closely related

254   sequences quantitatively, substitutions and indels were counted between each pair of the

255   sequences in their overlapping regions and their similarity indexes were calculated as ratio of

256   matching sites to the total amount of sites in the region of overlap, expressed with percentage:

257    $((a - d) / a) * 100\%$, where a = total number of sites in the region of overlap, d = number of

258   mismatches (Supplemental Tables 1 and 2). For these calculations, an original computer script,

259   Identity Counter, written by the author KVM and available by request, was applied.

260   For the LSU rDNA and ribosomal operon (concatenated SSU, 5.8S and LSU rDNAs) analyses,

261   the taxon sampling of only 50 sequences was used due to the limited availability of data for LSU

262   rDNA, and, especially, 5.8S rDNA. Therefore, the 5.8S rDNA (155 sites in the alignment) was

263   rejected from the analysis of the concatenated rRNA genes for seven sequences (*Chromera velia*,

264   *Colponema vietnamica*, *Goussia desseri*, *Stentor coeruleus*, and 3 environmental sequences:

265   Ma131 1A38, Ma131 1A45, and Ma131 1A49): these nucleotide sites were replaced with "N" in

266   the alignment. The resulting multiple alignments contained 50 sequences (2,911 sites) for the

267   LSU rDNA, and the same 50 sequences (4,636 sites) for the concatenated rDNAs (ribosomal

268   operon). Thus, both taxon sampling comprised an identical set of species, all of which were also

269   represented in the alignment of the 114 SSU rDNA sequences.

270   Maximum-likelihood (ML) analyses were performed by using RAxML 7.2.8 (Stamatakis, 2006)

271   under the GTR+Γ+I model with 8 categories of discrete gamma distribution. The procedure

272   included 100 independent runs of the ML analysis and 1,000 replicates of multiparametric

273   bootstrap. Bayesian inference (BI) analyses were computed in MrBayes 3.2.1 (Ronquist et al.,

274   2012) under the same model. The program was set to operate using the following parameters: nst

275   = 6, ngammacat = 12, rates = invgamma, covarion = yes; parameters of Metropolis Coupling

276   Marcov Chains Monte Carlo (mcmc): nchains = 4, nruns = 2, temp=0.2, ngen = 7,000,000,

277   samplefreq = 1,000, burninfrac = 0.5 (the first 50% of the 7,000 sampled trees, i.e., the first

278   3,500, were discarded in each run). The following average standard deviations of split

279   frequencies were obtained: 0.009904 for the SSU rDNA analysis, 0.001084 for the LSU rDNA

280   analysis, and 0.001113 for the ribosomal operon analysis. The calculations of bootstrap support

281   for the resulting Bayesian trees were performed by using RAxML 7.2.8 under the same

282   parameters as for the ML analyses (see above).

283

284   **Results**

285

286   **Light and scanning electron microscopy (Fig. 2).**

287   Thirty specimens of *Capitella capitata* from Roscoff (English Channel, France) and twenty-five

288   specimens from the White Sea biological station (WSBS, Russia) were dissected. All were

289   infected and the number of gregarine trophozoites per host varied from several individuals up to

290   about a hundred. The parasites from both locations had the same morphology, which fitted the

291   description of *A. sagittata*: an elongated body that narrowed toward the posterior end and with a

292   rounded anterior end, without a septum, and with two lateral projections giving the cell the

293   appearance of an anchor (e.g., Perkins et al., 2000). The average dimensions were 250 μm in

294   length and 37 μm in width (n = 25). The attached trophozoites were easy to dislodge from the

295   host epithelium, and a small drop of the cytoplasm then appeared at the front of the gregarine.

296   However, most of the gregarines were already free (not attached) during the dissection, without

297   any visible damage to their forebodies. All detached gregarines demonstrated gliding motility in

298   seawater. Other stages of the life cycle were not observed.

299   SEM micrographs of the gregarine surface revealed structure typical for eugregarines (Vavra &

300   Small, 1969): epicytic folds appressed to each other (3 folds per 1 μm) and converging to the

301   apex of the cell, where a small apical papilla of 2.5 μm in diameter was sometimes observed

302   (Fig. 2, F, G). The epicytic folds branched dichotomically in the apical region and similar

303   branching was also observed at the bases of the lateral projections (Fig. 2, E).

304   **Transmission electron microscopy (TEM) (Figs. 3 - 5).**

305   Cross-sections of trophozoites of *A. sagittata* showed a typical eugregarine tegument (Vivier,

306   1968; Vivier et al., 1970; Schrével et al., 1983): the epicyte consisting of numerous folds or

307   crests (Fig. 3, A-F) formed by the trimembrane pellicle (45 nm thick) composed of the plasma

308   membrane (covered by the cell coat) and the inner membrane complex, IMC (Fig. 3, D). Cross-

309   sections of the middle of the cell revealed regularly arranged and closely packed epicytic folds

310   (crests) that were approximately 1 μm high and 375 nm wide and had finger-like shapes with

311   weak constrictions at their bases. A 13-nm-thick internal lamina (an electron-dense layer

312   undelaying the pellicle) was observed, which was thickened at the bottom of grooves between

313   the epicytic folds (up to 48-50 nm) and did not form links in the bases of the folds, which are

314   characteristic for many other eugregarines (see, e.g., Vivier, 1968; Schrével et al., 2013; also see

315   Discussion). Six to eight rippled dense structures (also called apical arcs) were present at the top

316   of the folds. No 12-nm apical filaments, characteristic for most eugregarines (Vivier, 1968;

317   Schrével et al., 2013), were detected, but electron-dense plates were found at the top of the folds

318   just beneath the IMC (Fig. 3, D). The cytoplasm in the folds contained fibrils (Fig. 3, C). The

319   folds displayed increased bulging near the bases of the lateral projections of the trophozoite cell,

320   and rare micropores were observed at the lateral surfaces of the folds in this region (Fig. 3, F). In

321   the frontal region of the cell, gaps in between the folds were present and large electron-dense

322   globules were found in the cytoplasm of the folds (Fig. 3, E). Circular filaments (~30 nm) were

323   observed just beneath the tegument (Fig. 3, C, F, G). The ecto- and endoplasm were not

324   separated distinctly from each other (Fig. 3, A, B): the thickness of the ectoplasm (the cytoplasm

325   layer free of amylopectin) varied only from 1.5 to 2.5 μm. Rounded amylopectin granules of

326   approximately 0.7-1 μm were abundant in the deeper layers of the cytoplasm.

327   The trophozoites were attached to the intestinal epithelium by a bulbous attachment apparatus

328   that was embedded in the host cell and connected to the trophozoite cell by a short stalk or neck

329   (Fig. 4, A). The anterior part of the attachment apparatus contained a large lobate vacuole filled

330   with a loose, thin fibrillar network, which could be the result of coagulation of some matter

331   during the fixation and/or embedding procedures (Fig. 4, A and B). A groove pinching a small

332   portion of the host cell was present at the base of the attachment bulb (Figs. 4, B and 5). The

333   IMC of the parasite pellicle terminated at this site and the attachment bulb was apparently

334   covered only by the single plasma membrane of the gregarine, not by the pellicle (Figs. 4, B and

335   5). A bundle of longitudinal filaments spread throughout the gregarine cell backwards from the

336   IMC terminus. The wall of a large frontal vacuole arose from the same site (Figs. 4, B and 5).

337   The cytoplasm behind the vacuole contained individual amylopectin granules. Longitudinal

338   sections of four trophozoites revealed the complex structure of the contact zone between the

339   parasite and the host cell (Figs. 4, B and 5). The cell junction was formed by two closely

340 adjacent plasma membranes of the host and parasite cells without a distinct gap between them.

341 Electron-dense areas were present on both parasite and host cell sides of the junction. The area

342 within the host cell appeared uniformly grey, whereas that of the gregarine cell was distinguished

343 into three zones: (i) a black zone immediately adjacent to the cell junction, (ii) a grey zone

344 similar to that in the host cell, and (iii) thin fibrils arising from the grey zone towards the interior

345 of the cell (Figs. 4, B and 5).

346 **Sequence diversity in *A. sagittata*.**

347 Four contiguous nucleotide sequences of *A. sagittata* were obtained (Table 1), three (Roscoff,

348 WSBS 2010, and WSBS 2011) covering complete or near-complete ribosomal operon (SSU,

349 5.8S, LSU rDNAs, and the internal transcribed spacers ITS 1 and 2), and a shorter one (WSBS

350 2006) lacking the most part of LSU rDNA (only first ~600 bp of it were amplified and

351 sequenced; Fig.1). Three of these sequences (Roscoff, WSBS 2006, and 2011; ribotype 1) were

352 near identical to one another (99.4 to 100%; Supplemental Tables 1 and 2), whereas the fourth

353 (WSBS 2010; ribotype 2) was more divergent (94.3 to 96.2% identities with three other

354 sequences). Nucleotide substitutions and indels were concentrated chiefly in the hypervariable

355 regions of the rRNA genes and in the ITSs (the ITSs contained ~40% of total mismatches).

356 A search for compensatory base changes (CBCs) in ITS2 was performed to discriminate possible

357 cryptic species (Coleman, 2000; Müller et al., 2007; Coleman, 2009; Wolf et al., 2013). The

358 manually assembled secondary structure was tested by MFOLD in the temperature range of 5 –

359 37°C and was found to be nearly optimal. The *A. sagittata* ITS 2 (Fig. 6) appears to be one of the

360 shortest known sequences in eukaryotes (102 and 100 nucleotides in the ribotype 1 and 2,

361 respectively; helix IV is absent), however it retains universally conserved features (Schultz et al.

362 2005): a U-U mismatch in helix II and a vestige of the "UGGU" motif in helix III, modified as

363 "UGUGU" (Fig. 6). Four CBCs between the ribotypes 1 and 2 were detected: two putative in the

364 spacer stalk, one in helix I, and one at the base of helix III.

365 **Phylogenies inferred from SSU rDNA.**

366 Phylogenies of SSU rDNA (114 sequences; 1,570 sites) showed a well supported monophyly of

367 the major groups of alveolates (ciliates, dinoflagellates and their subgroups, and apicomplexans)

368 with a high Bayesian posterior probability (PP) and moderate ML bootstrap percentage (BP)

369 support (Fig. 7). The backbone of the apicomplexans was poorly resolved in both Bayesian and

370 ML analyses; nevertheless, the topologies were largely congruent with small differences in the

371   gregarine branching order. The cryptosporidians were consistently placed as the sister group of

372   all gregarines in both analyses, although with low support. Archigregarines (*Selenidium* spp.)

373   formed three branches of greatly variable lengths and were not monophyletic. Eugregarines were

374   separated from archigregarines and were monophyletic in both Bayesian and ML trees, although

375   without a cogent support (PP=0.58, BP=12%). They comprised eight well-supported subclades

376   of an uncertain branching order, five of which were recently erected as superfamilies (Clopton,

377   2009; Rueckert et al., 2011; Simdyanov & Diakin, 2013; Cavalier-Smith, 2014), namely: (i)

378   Lecudinoidea (*Veloxidium leptosynaptae* and the aseptate marine Lecudinidae (with the type

379   species *Lecudina pellucida*) and Urosporidae); (ii) Cephaloidophoroidea (septate and aseptate

380   gregarines from crustaceans); (iii) Gregarinoidea (septate gregarines from insects); (iv)

381   Stylocephaloidea (septate gregarines from insects); and (v) Actinocephaloidea (septate and some

382   aseptate gregarines from insects including neogregarines and *Monocystis agilis*; Fig. 7). Two

383   additional lineages were designated as *incertae* sedis: one (vi) was composed entirely of

384   unidentified environmental sequences including a "clone from the foraminiferan *Ammonia*

385   *beccarii*", and the other (vii) comprised the aseptate marine gregarine *Paralecudina polymorpha*

386   and related environmental sequences. The last lineage (viii), hereby named "Ancoroidea", was a

387   robust monophyletic clade that includes *A. sagittata*, *Polyplicarium* spp., and 70 environmental

388   sequences from anoxic marine habitats (Figs. 7 and 8). Two clusters were found within this clade

389   (Fig. 8): a robust clade including *A. sagittata* and related environmental sequences, and the clade

390   of *Polyplicarium* spp. and environmental relatives, which was either strongly (Fig. 7) or

391   moderately (Fig. 8) supported depending on the dataset. The environmental clade vii ("clone

392   from *Ammonia beccarii*" and relatives) displayed a certain affinity to the Ancoroidea (Figs. 7 and

393   8).

394   Aseptate gregarines (Ancoroidea, Lecudinoidea, and the clade of *Paralecudina*) were not

395   monophyletic, whereas the four other lineages (Actinocephaloidea, Cephaloidophoroidea,

396   Gregarinoidea, and Stylocephaloidea) formed a weakly supported clade of primarily septate

397   eugregarines, although some representatives of this "septate" clade are actually aseptate (marked

398   with asterisks in Fig. 7).

399   **Analyses of LSU rDNA and the ribosomal operon.**

400   All analyses of the LSU rDNA dataset (50 sequences, 2911 sites) showed topologies that were

401   congruent with the SSU rDNA result with minor exceptions (Fig. 9, A). The three sequences

402   from *A. sagittata* (Roscoff, WSBS 2010 and 2011) were monophyletic and related to a clade

403   containing *Gregarina* spp. and crustacean gregarines in both Bayesian and ML analyses (Fig. 9,

404   A). All gregarines, including a clade containing *Ascogregarina* and Neogregarinida sp. OPPPC1

405   (Fig. 9, A), were monophyletic.

406   Trees derived from the ribosomal operon dataset (alignment of 50 sequences, 4,636 sites)

407   showed the same topology as the LSU rDNA tree (Fig. 9, B) with increased supports for several

408   branches. Within the sporozoan clade, all gregarines were monophyletic and most supports were

409   similar to those in the LSU rDNA tree, however BP support for the subclade including *A.*

410   *sagittata* from Roscoff and *A. sagittata* from WSBS 2011 increases considerably (BP = 92% vs

411   86% in the LSU rDNA tree).

412

413   **Discussion**

414

415   **Ultrastructure of the cortex**.

416   The tegument of gregarines is composed of a trimembrane pellicle (plasma membrane and inner

417   membrane complex, formed by two closely adjacent cytomembranes) underlaid by the internal

418   lamina, an electron-dense layer that most likely consists of closely packed thin fibrils (Vivier,

419   1968; Vivier et al., 1970; Schrével et al., 1983; Schrével et al., 2013). The pellicle forms so-

420   called epicyte – a set of multiple narrow longitudinal folds called epicytic folds (Vivier, 1968;

421   Vavra & Small, 1969; Vivier et al., 1970; Schrével et al., 1983; Schrével et al., 2013). The

422   epicytic folds of most eugregarines, including the Lecudinidae, have a very specific structure:

423   they contain several rippled dense structures (also called apical arcs) and 12-nm apical filaments

424   within their top regions: the former are located between plasmalemma and the IMC, the latter

425   just beneath the IMC; the internal lamina usually forms links or septa in the bases of the folds

426   (Vivier, 1968; Schrével et al., 1983; Simdyanov, 1995b; Simdyanov, 2004; Simdyanov, 2009;

427   Schrével et al., 2013). Both the rippled dense structures and 12-nm apical filaments are thought

428   to be involved in gliding motility, which is characteristic of typical eugregarines; however, the

429   detailed mechanism of gliding remains unclear (Vivier, 1968; Vavra & Small, 1969; Mackenzie

430   & Walker, 1983; King, 1988; Valigurová et al., 2013). Observations of *A. sagittata* trophozoites

431   are congruent with the information available for eugregarines, with a few differences. The 12-nm

432   apical filaments were not observed in the epicytic folds, they are obscured or substituted by an

433   electron dense plate. Additional studies, including immunochemical methods, are required to

434   reveal the true composition of this structure.

435   Another interesting observation is the absence of the links of the internal lamina (see above) in

436   the bases of the epicytic folds of *A. sagittata* (Fig. 3, C, F) and, simultaneously, the accumulation

437   of large cytoplasmic inclusions within the folds and the location of micropores on the lateral

438   surfaces of the epicytic folds, while their typical location is in between the folds. These three

439   peculiarities have been reported for a few other eugregarines such as *Kamptocephalus mobilis*,

440   *Mastigorhynchus bradae*, and *Stylocephalus* spp. (Desportes, 1969; Simdyanov, 1995a) and they

441   have occurred together in each case, so they obviously correlate with one another. Thus, in

442   eugregarines with typical epicytes, the links of the internal lamina appear to act as barriers

443   between the space within the epicytic folds and the rest of the cytoplasm, but their true purpose

444   remains unclear.

445   Circular cortical filaments found in *A. sagittata* are similar to those in eugregarines with

446   peristaltic motility, such as *Monocystis* spp.*, Nematocystis magna,* and *Rhynchocystis pilosa*

447   (Miles, 1968; Warner, 1968; Vinckier, 1969; MacMillan, 1973), or bending motility, such as

448   some *Gregarina* spp. (Valigurová et al., 2013). However, neither peristaltic nor bending motility

449   has been observed in *A. sagittata*.

450   **Attachment apparatus: mucron or epimerite?**

451   The terminology of the gregarine attachment apparatuses is rather confusing. The commonly

452   used names are "mucron" and "epimerite" depending on whether the trophozoite is aseptate or

453   septate, respectively (Levine, 1971). The mucron is mostly small and can be pointed, rounded or

454   sucker-shaped, whereas the epimerite varies in size and shape from elongated to lenticular and is

455   sometimes equipped with hooks or other projections (Grassé, 1953; Schrével et al., 2013).

456   Consequently, the attachment apparatus of *A. sagittata* should be called the mucron (e.g., see:

457   Perkins et al., 2000), although it more resembles the epimerite on its ultrastructure and fate (see

458   below). Similar confusion arises about the attachment apparatus in other gregarines and stems

459   from Levine's definition (Levine, 1971) that "[the] mucron is an attachment organelle of aseptate

460   gregarines…", i.e., its applies equally to archigregarines and aseptate eugregarines as both of

461   them are aseptate. TEM data has since forced a revision of this light microscopy-driven

462   perspective revealing conspicuous differences between the attachment organelles of archi- and

463   eugregarines. The archigregarine mucron contains an apical complex and performs myzocytotic

464   feeding, i.e., intermittent sucking of nutrients through a temporary cytostome (Schrével, 1968;

465   Schrével, 1971a; Simdyanov & Kuvardina, 2007; Schrevel et al., 2016), whereas eugregarine

466   trophozoites have no apical complex (with exception of the earliest developmental stages) and do

467   not exhibit myzocytosis in their mucrons and epimerites (Schrével & Vivier, 1966; Devauchelle,

468   1968; Baudoin, 1969; Desportes, 1969; Ormierès & Daumal, 1970; Hildebrand, 1976; Ormierès,

469   1977; Tronchin & Schrével, 1977; Ouassi & Porchet-Henneré, 1978; Valigurová & Koudela,

470   2005; Valigurová et al., 2007; Valigurová, Michalková & Koudela, 2009; Schrével et al., 2013;

471   Schrével et al., 2016). Thus, there is no doubt that the mucron of archigregarines and the

472   epimerite in septate eugregarines differ in their genesis, structure, and feeding function (Table 2,

473   Fig. 10). However, the "mucron" of aseptate eugregarines (e.g., some lecudinids) is actually a

474   homologue of the epimerite when examined in detail, not of the archigregarine mucron (Table 2

475   and Fig. 10). The archigregarine mucron contains the apical complex and is covered by a

476   trimembrane pellicle, excepting a small region in front of the conoid where the IMC is absent

477   and the cytostome is intermittently opened (Schrével, 1968; Schrével, 1971a; Kuvardina &

478   Simdyanov, 2002; Simdyanov & Kuvardina, 2007). In contrast, the attachment organelles of

479   eugregarine trophozoites, both the "mucron" and epimerite, originate from the region

480   corresponding to the "cytostome site" of archigregarines (in front of the conoid where is no IMC)

481   as a progressing protuberance covered by a single plasma membrane (Table 2 and Fig. 10, F and

482   G) and their apical complex disappears in a short time after the protuberance is starting to

483   develop (Desportes, 1969; Tronchin & Schrével, 1977; Ouassi & Porchet-Henneré, 1978). The

484   archigregarine mucron forms a septate cell junction with the host cell, the main characteristic of

485   which is a conspicuous "septate" gap between the host plasma membrane and the gregarine

486   pellicle (Simdyanov & Kuvardina, 2007). In contrast, a cell junction in the eugregarine

487   attachment apparatus (both the so-called "mucron" and epimerite) forms no gap between the

488   parasite and host cell membranes, which are underlaid by electron dense areas (Figs. 5 and 10).

489   The eugregarine cell junction zone is bordered by the circular groove running along the edge of

490   the attachment site and pinching a small portion of the host cell; this structure is absent in

491   archigregarines. In archigregarines, a large mucronal vacuole is present within the mucron,

492   which is intermittently connected by a duct (cytopharynx) with the cytostome; obviously it is a

493   food vacuole (Schrével, 1968; Schrével, 1971a; Kuvardina & Simdyanov, 2002; Simdyanov &

494   Kuvardina, 2007; Schrével et al., 2016). In eugregarine trophozoites, no cytostome-

495 cytopharyngeal complex has been observed – only with the possibly exception of the earliest

496 developmental stage (Fig. 10, F1, G3). A large frontal vacuolar structure (usually flattened and

497 containing fibrillar matter) develops just beneath the cell junction region; however, the vacuole

498 can sometimes be completely replaced with a dense fibrillar zone (Tronchin & Schrével, 1977;

499 Ouassi & Porchet-Henneré, 1978). No evidence of its involvement in the eugregarine feeding has

500 been observed. Finally, archigregarines retain their mucron (together with the apical complex)

501 well into the syzygy stage (Fig. 10, D) (Kuvardina & Simdyanov, 2002). Although the fate of the

502 attachment apparatus in aseptate eugregarines is poorly studied, it is presumably the same that in

503 the septate eugregarines, whose gamonts lose their epimerite upon the detaching from the host

504 epithelium (Grassé, 1953; Devauchelle, 1968; Valigurová, Michalková & Koudela, 2009;

505 Schrével et al., 2013).

506 The structure and cytoplasmic content of developed epimerites in septate gregarines vary

507 substantially (Devauchelle, 1968; Baudoin, 1969; Desportes, 1969; Ormierès & Daumal, 1970;

508 Tronchin & Schrével, 1977; Valigurová & Koudela, 2005): numerous mitochondria, granules of

509 amylopectin, lipid drops and vacuoles, well developed arrays of ER, and numerous fibrillar

510 structures (microtubules and microfilaments) – especially in the basal region if it is shaped like a

511 neck or stalk (so-called "diamerite") as, e.g., in *Epicavus araeoceri* (Ormierès & Daumal, 1970),

512 can be present.

513 The attachment apparatus of *A. sagittata* displays the main features of a simple epimerite (Fig. 4)

514 lacking cytoplasmic organelles and inclusions (absence of mitochondria, ER, and lipid drops),

515 although, like in epimerites of septate gregarines, there are amylopectin granules and a large

516 frontal vacuole – although not flattened, but rather bulky. Detached gregarines (mature

517 gamonts?) have no epimerite, which has been apparently discarded, judging from the appearance

518 and behaviour of individuals that were artificially dislodged from the host epithelium (see

519 Results). Some other aseptate gregarines also possess complex attachment organelles that are

520 comparable to true epimerites of septate gregarines in shape, ultrastructure, and fate (absent in

521 mature gamots), e.g., *Lecudina* (*Cygnicollum*) *lankesteri*. Unlike *A. sagittata*, the epimerite of *L.*

522 *lankesteri* (Fig. 10, I and J) has a complex structure: the cytoplasm contains mitochondria,

523 inclusions, a well-developed cytoskeleton, and abundant fibrillar structures in the basal region

524 (Desportes & Théodoridès, 1986). Thus the sucker-shaped "mucrons" of many other lecudinids,

525 such as *Lecudina* sp. from the polychaete *Cirriformia* (*Audouinia*) *tentaculata* (Ouassi &

526   Porchet-Henneré, 1978; also see Fig. 10, G and H) and *L. pellucida* (Schrével & Vivier, 1966),

527   which are not deeply embedded into host cell, can be considered underdeveloped epimerites that

528   lack the main (middle) region containing cytoplasmic organelles and inclusions.

529   Taking into account the homologies of the eugregarine attachment organelles, the term "mucron"

530   should be restricted to the attachment apparatus in archigregarines, which contains the apical

531   complex and performs myzocytosis. In eugregarines, both aseptate and septate, the term

532   "epimerite" appears to be more appropriate, and is in accordance with the definitions and

533   gregarine descriptions in the "classic" literature on gregarines (e.g., Watson Kamm, 1922). This

534   terminological correction will remove ambiguity in taxonomical diagnoses and emphasize that

535   the epimerite is a shared evolutionary innovation (synapomorphy) of eugregarines. More

536   representative data are required to distinguish different types of epimerites: for example, the

537   "cephaloid" type of *Cephaloidophora* (Simdyanov, Diakin & Aleoshin, 2015) and the

538   underdeveloped epimerites of *Lecudina* spp. (above), which can be called "pseudomucrons". It

539   should be noted, that in some morphologically divergent eugregarines, e.g., *Uradiophora maetzi*

540   and representatives of the family Dactylophoridae, the epimerite was reduced and they are

541   anchored in host cells with projections of the protomerite (Ormierès & Marquès, 1976; Desportes

542   & Théodoridès, 1985).

543   **Reconciling molecular phylogenies with eugregarine morphology.**

544   In SSU rDNA phylogenies published to date, gregarines (*sensu* class Gregarinomorpha Grassé,

545   1953) have been not monophyletic. The most probable reason for that is that the SSU rDNA

546   sequences of many gregarines are highly divergent, therefore the topologies of resulting

547   phylogenetic trees are sensitive to changes in alignment site selection and taxon sampling.

548   Additionally, the presence of many long branches among gregarines and other apicomplexans

549   may lead to long branch attraction (LBA) artefacts (Bergsten, 2005). Only after careful manual

550   editing of the alignment (see supplemental raw data) and the exclusion of single gregarine

551   sequences corresponding to three extremely long branches (*Pyxinia crystalligera*, *Stenophora*

552   *robusta*, and *Trichotokara* spp.), all of the other gregarines did form a monophyletic lineage,

553   albeit weakly supported (Fig. 7). Despite their weakly supported monophyly in the SSU rDNA

554   phylogenies, all gregarines display a distinct morphological synapomorphy: the gametocyst,

555   which is an encysted syzygy (Frolov, 1991). Among other apicomplexans, only adeleid

556   coccidians have syzygy, but without the subsequent encystment into a gametocyst. Unlike SSU

557   rDNA alone, both the LSU rDNA and ribosomal operon-based phylogenies support the

558   monophyly of gregarines, although with a significantly limited taxon sampling without

559   archigregarine and some eugregarine lineages (Fig. 9). Similarly to SSU rDNA phylogenies,

560   long branch attraction can affect these tree topologies, although its negative effects are expected

561   to be lower than in SSU rDNA-alone phylogenies because the relative evolution rates of the LSU

562   rDNA in apicomplexans are more even than those of the SSU rDNA (Simdyanov, Diakin &

563   Aleoshin, 2015).

564   Neogregarines (order Neogregarinida Grassé, 1953) never form a monophyletic lineage but are

565   shuffled amongst actinocephalids (confirming Grassé's hypothesis for their origin) and should

566   therefore be included in the superfamily Actinocephaloidea. Consequently, the absence of

567   merogony should be removed from the eugregarine diagnosis. Archigregarines are paraphyletic

568   in SSUr DNA-based phylogenies, forming two or three independent lineages, which are often

569   shuffled with eugregarine clades in available molecular phylogenetic trees (e.g., Cavalier-Smith,

570   2014). However, the proposition of the independent polyphyletic origin of different eugregarine

571   lineages (Cavalier-Smith, 2014) contradicts evidence from ultrastructural studies. On the

572   contrary, relying on the morphological evidence, eugregarines appear to be a monophyletic

573   group because all their major lineages share at least two distinct morphological apomorphies

574   (Fig.11): (i) the presence of the epimerite (see above) and (ii) gliding motility apparently

575   associated with rippled dense structures (apical arcs) and 12-nm apical filaments in eugregarine

576   epicytic folds (also see: Krylov & Dobrovolskij, 1980). Archigregarines, apart from having the

577   different type of attachment apparatus (the mucron, see above), lack the eugregarine type of the

578   epicyte: they possess longitudinal pellicular folds (bulges), which are significantly larger than

579   eugregarine epicytic folds (crests) and contain neither the rippled dense structures (apical arcs)

580   nor 12-nm apical filaments (Fig. 11, A): just beneath the pellicle, both in the bulges and between

581   them, one to three layers of longitudinal subpellicular microtubules are located that have never

582   observed in the eugregarines (Schrével, 1971a; Schrével, 1971b; Simdyanov & Kuvardina, 2007;

583   Schrével & Desportes, 2013b). The archigregarine pellicular bulges are thus only a simple

584   surface sculpture, whereas the epicytic crests of eugregarines are complex organelles that

585   apparently provide the gliding motility, which is absent in archigregarines and substituted by

586   active bending motility instead (Grassé, 1953; Schrével, 1971a; Perkins et al., 2000). The failure

587   to distinguish these different cortical structures in archi- and eugregarines (e.g., Cavalier-Smith,

588   2014) is linked to the use of a misleading term "longitudinal folds", which assumes their identity

589   in both gregarine groups and has been used as the morphological evidence of the polyphyletic

590   origin of eugregarines from different archigregarine lineages (Cavalier-Smith, 2014). To

591   eliminate this ambiguity, we propose the term "epicytic crests" instead of "folds" for

592   eugregarines and the terms "longitudinal folds" or "bulges" for archigregarines. The term "crests"

593   has been already used to describe the eugregarine epicyte (Pitelka, 1963: p. 90) and corresponds

594   well to their narrow shape, compressed from the sides, in contrast to the large, gently sloped

595   pellicular folds of archigregarines.

596   The hypothesis of eugregarine polyphyly is inferred solely from ambiguous SSU rDNA-based

597   molecular phylogenies (which are low resolved and apparently affected by LBA) and assumes

598   the independent origin both of epimerite and epicytic crests in the major eugregarine lineages,

599   i.e., they are convergences (homoplasies), that appears unlikely considering their detailed

600   ultrastructural resemblance in a broad range of gregarines (see below). Therefore, following the

601   principle of Ockham's razor (minimum of assumptions), we rather consider the epimerite and

602   epicytic crests shared-derived characteristics of eugregarines (Figs. 10 and 11). Apart from the

603   Ancoroidea (*A. sagittata*), these features are widespread within all other eugregarine

604   superfamilies revealed to date: Actinocephaloidea (Baudoin, 1969; Vávra, 1969; Ormierès &

605   Daumal, 1970; Vorobyeva & Dyakin, 2011), Stylocephaloidea (Desportes, 1969), Gregarinoidea

606   (Devauchelle, 1968; Tronchin & Schrével, 1977; Dallai & Talluri, 1983; Schrével et al., 1983),

607   Cephaloidophoroidea (epimerite is understudied) (Desportes, Vivarès & Théodoridès, 1977;

608   Simdyanov, Diakin & Aleoshin, 2015), Lecudinoidea (Schrével & Vivier, 1966; Vivier, 1968;

609   Ouassi & Porchet-Henneré, 1968; Corbel, Desportes & Théodoridès, 1979; Simdyanov, 1995b;

610   Simdyanov, 2004; Simdyanov, 2009; Diakin et al., 2016). It should be noted, however, that the

611   Lecudinoidea and Actinocephaloidea, apart from typical (core) representatives, also include

612   morphologically divergent forms possessing various modifications of the epicyte structure

613   (fusion or reduction of the epicytic crests, sometimes formation of hair-like projections). Such

614   modifications are always attended by the loss of gliding motility and transition to metaboly

615   (peristaltic motility) or nonmotility. Within the Lecudinoidea, there is the family Urosporidae

616   parasitizing coelom of polychaetes and echinoderms, unlike the Lecudinidae – core

617   representatives, which are intestinal parasites, chiefly in polychaetes. Within the

618   Actinocephaloidea, there are intracellular neogregarines without epicyte (Žižka, 1978) and the

619   family Monocystidae parasitizing seminal vesicles of oligochaetes, whereas core representatives,

620   the Actinocephalidae, are intestinal parasites of insects (chiefly). The monocystids show very

621   divergent and various structure of the cortex and possess peristaltic motility (Miles, 1968;

622   Warner, 1968; Vinckier, 1969; MacMillan, 1973) that is similar with some species of the

623   Urosporidae (Dyakin and Simdyanov, 2005; Landers & Leander, 2005; Leander et al., 2006;

624   Diakin et al., 2016). However, in terms of comparative anatomy, existence of certain aberrant

625   forms does not cancel the presence of a shared bauplan in core representatives of the group: e.g.,

626   "archiannelids" and leeches within annelids when compared with the core forms as polychaetes

627   and oligochaetes – in this and many other cases, large majority of diagnostic characteristics may

628   be applied only to the "core group". The core (non-aberrant) representatives of all known

629   eugregarine lineages/ superfamilies share both epimerite (understudied in Cephaloidophoroidea)

630   and epicytic crests. Therefore, in compliance with the main principle of cladistics, we present a

631   morphology-driven hypothesis on the monophyly of eugregarines based on the presence of the

632   epimerite and epicytic crests as defining synapomorphies of the order Eugregarinida (Figs. 10

633   and 11), which may be included in its diagnosis (see below in the taxonomical subsection). This

634   hypothesis cannot be tested by the currently available molecular data but is potentially consistent

635   with it (at least, does not contradict: see Figs. 7 and 9). More robust molecular datasets are

636   therefore needed to test whether these structures represents true homologies.

637   Three cases seemingly challenge the monophyly eugregarines at the morphological level:

638   *Veloxidium leptosynaptae, Caliculium glossobalani*, and *Seledinium melongena* (Wakeman &

639   Leander, 2012; Wakeman, Heintzelman & Leander, 2014; Wakeman et al., 2014). *V.*

640   *leptosynaptae* and *C. glossobalani* broadly resemble archigregarines but SSU rDNA phylogenies

641   unambiguously place them in the eugregarine clades Lecudinoidea and Gregarinoidea,

642   respectively (Fif. 7). External morphology and ultrastructure and of *S. melongena* is somewhat

643   similar to *C. glossobalani*, but it is a sister taxon to the archigregarine *Selenidium terebellae*

644   (Fig.7). The certain morphological resemblances of all three species have been used to challenge

645   the archi- and eugregarine concepts, however, their ultrastructure provides no firm support for

646   such conclusions. *V. leptosynaptae* lacks ultrastructural data altogether and those available for *C.*

647   *glossobalani* do not reveal any key ultrastructural features of either archi- or eugregarines (see

648   Figs. 10 and 11). *C. glossobalani* lacks a genuine mucron, the associated conoid, mucronal

649   vacuole, and rhoptries and the layered arrangement of the subpellicular microtubules that is

650   characteristic for archigregarines (see above). *C. glossobalani* also lacks eugregarine epicytic

651   crests (it has only low, wide, and mildly sloping longitudinal folds resembling those in

652   archigregarines, however without microtubules) and the epimerite: its sucker-shaped attachment

653   organelle is covered by a trimembrane pellicle in detached individuals (note, however, that no

654   trophozoites attached to the host cells were examined by TEM). Thus, as yet, *V. leptosynaptae*

655   and *C. glossobalani* rather appear to be morphologically divergent eugregarines when compared

656   with the typical representatives of their phylogenetic lineages (e.g., Lecudinoidea and

657   Gregarinoidea) – possibly because they both occur in unusual habitats or hosts (compare with

658   Urosporidae and Monocystidae) – but their similarity with archigregarines is superficial and not

659   supported by ultrastructural data. The situation with *S. melongena* is somewhat similar to *C.*

660   *glossobalani*: most aforementioned key features of archi- and eugregarines were not identified

661   (Wakeman, Heintzelman & Leander, 2014).The presence of structures resembling the mucronal

662   vacuole and micronemes or small rhoptries combined with molecular phylogenetic data

663   nevertheless suggests that *S. melongena*a could be a divergent archigregarine, which has

664   undergone a morphological transformation possibly due its unusual, coelomic localization within

665   the host. Certain ultrastructural similarities between *C. glossobalani* and *S. melongena* are

666   actually caused by the "shared" absence of the defining ultrastructural features of both archi- and

667   eugregarines. Altogether, the external morphology of *V. leptosynaptae, C. glossobalani*, and *S.*

668   *melongena* reaffirms that the external morphology of gregarine trophozoites and gamonts is a

669   poor taxonomic marker susceptible to convergence. Because evidence of the key ultrastructural

670   characteristics in all three species is presently lacking, they cannot be used in evaluating

671   hypotheses on the evolutionary origin of archi- and eugregarines.

672   The dichotomy between aseptate and septate gregranies is rejected by the SSU rDNA

673   phylogenies: that is consistent with the hypothesis of Grassé, which considered some aseptate

674   forms likely derived secondarily from septate gregarines (e.g., *Paraschneideria* with young

675   septate trophozoites and aseptate gamonts and, most likely, *Ascogregarina* (former "mosquito

676   *Lankesteria*")); also there are intermediate forms between aseptate and septate gregarines, e.g.,

677   *Ganymedes* (Grassé, 1953; Schrével & Desportes, 2013b). Hence, the septum appears to be an

678   evolutionarily unstable trait, therefore the separation of the order Eugregarinida into Aseptata

679   and Septata, which is additionally not supported by available molecular data, should be

680   abolished.

681   In contrast, the separation of eugregarines into several deep lineages (superfamilies) is well

682   supported by the SSU rDNA phylogenies, although some families of gregarines are still missing

683   in these analyses (e.g., Dactylophoridae and Hirmocystidae), and others are represented by a

684   single species (e.g., Monocystidae) or are composed exclusively of environmental sequences

685   (e.g., the cluster of "*Ammonia*-like" clones). Despite these limitations, designation of the well

686   supported eugregarine clades with a superfamily rank (Actinocephaloidea, Stylocephaloidea,

687   Gregarinoidea, Cephaloidophoroidea, and Lecudinoidea) appears to be natural and has been

688   proposed repeatedly (Clopton, 2009; Rueckert et al., 2011; Simdyanov & Diakin, 2013;

689   Cavalier-Smith, 2014).

690   The morphology and host spectra of eugregarine superfamilies (Table 3) does not correlate with

691   the rates of evolution of their SSU rDNAs. The ancestral eugregarines were likely intestinal

692   parasites of marine invertebrates (similarly to archigregarines and lower coccidians), whose

693   morphology may have resembled aseptate lecudinids with weakly developed epimerites.

694   However, the Lecudinodea have highly divergent sequences, whereas some taxa with short

695   branched sequences have a complex and divergent morphology (Actinocephalidae and

696   Stylocephalidae). Consequently, the use of general morphology of trophozoites in defining

697   taxonomic levels lower than the order should be implemented with caution, because these

698   characteristics may be convergent (for example, peristaltic motility and aberrant surface

699   structures in eugregarines that occur in the host coelom – see above). The independent

700   morphological and molecular evolutions in eugregarines can be also observed in *Ancora*

701   *sagittata* and its sister group *Polyplicarium*, which have aseptate organization resembling

702   lecudinids, but are not closely related to the Lecudinoidea (Fig. 7). Since they form the firmly

703   supported separated molecular phylogenetic lineage, we formally delimit them as a new

704   superfamily Ancoroidea in the framework of Linnaean taxonomy.

705   **Molecular and morphological diversity in Ancoroidea.**

706   All environmental sequences in GenBank, which were affiliated with the Ancoroidea (Fig. 8),

707   were obtained from anoxic marine sediments, including cold methane seeps and shallow water

708   hydrothermal zones (Edgcomb et al., 2002; Stoeck & Epstein, 2003; Stoeck, Taylor & Epstein,

709   2003; Stoeck et al., 2007; Takishita et al., 2007; Santos et al., 2010; Boere et al., 2011; Garman

710   et al., 2011; Orsi et al., 2012). The geographical distribution of samples containing the ancoroid

711   sequences is wide: arctic, temperate, and tropical zones of the Atlantic and Indo-Pacific regions:

712   Greenland, North America (Vancouver (BC) and Cape Cod), the Gulf of Mexico, and Papua

713   New Guinea; however, the sequences that are closely related to *A. sagittata* were collected only

714   from the Atlantic and European Arctic. Considering that both *A. sagittata* and *Polyplicarium* spp.

715   parasitize polychaetes in the family Capitellidae and that most of related environmental

716   sequences have been retrieved from anoxic environments, in which the Capitellidae are

717   preferentially distributed, we hypothesize that all these Ancoroidea likely share the same group

718   of hosts in similar habitats.

719   In this context, the affiliation of another species, *Ancora prolifera* Clausen, 1993, to this genus is

720   questionable because this species is a parasite of the non-capitellid polychaete *Microphthalmus*

721   *ephippiophorus* (Hesionidae). *A. prolifera* and *A. sagittata* are morphologically similar: the latter

722   also has lateral projections, however these are not located in the plane of the body axis but at an

723   angle to it, similar to lifted wings of a bird (Clausen, 1993). Clausen also observed a nucleus-like

724   structure in these projections (apart from the genuine nucleus) and therefore proposed that cell

725   division in this gregarine occurs via budding, which has never been observed in eugregarines.

726   One additional species, *Ancora lutzi* Hasselmann, 1918, was only described in a preliminary note

727   (Hasselmann, 1918) without figures and delimitation of type material. The gregarines were

728   present in two individuals of *Capitella capitata* (the same host species as *A. sagittata*) collected

729   in the bay of Manguinhos (Brazil) and distinguished from *A. sagittata* by a shorter and wider

730   body, more intense granulation in the cytoplasm, and a frontal nucleus. This species was never

731   rediscovered and was later suggested to represent a morphological variant of *A. sagittata*

732   (Watson Kamm, 1922).

733   Because the Ancoroidea is split into two distinct clusters (Fig. 8), we recognize two families

734   within this superfamily: Ancoridae fam. nov. and Polyplicariidae Cavalier-Smith, 2014. The

735   family Ancoridae is currently monotypic (single genus *Ancora*). This taxonomical rearrangement

736   removes *A. sagittata* from the family Lecudinidae. The family Polyplicariidae, apart from the

737   type genus *Polyplicarium*, likely includes at least two additional undescribed genera

738   corresponding to two environmental clusters (Fig. 8). The representatives of Ancoroidea display

739   small morphological differences from the Lecudinidae in the fine structure of the epicytic folds

740   and attachment apparatus (in *A. sagittata* described above; the ultrastructure of *Polyplicarium* is

741   not known).

742   **Putative cryptic species in *A. sagittata*.**

743 Considerable differences have been observed between two ribotypes of *A. sagittata,* WSBS 2010

744 contig (ribotype 2) and WSBS 2006, 2011, and Roscoff contigs (ribotype1). Four CBCs in ITS2

745 (Fig. 6) suggest that these ribotypes represent two distinct cryptic species (Coleman, 2000;

746 Müller et al., 2007; Coleman, 2009; Wolf et al., 2013). Although the ribotype of the *A. sagittata*

747 type material is not known, we still have annotated the sequences of ribotype 1 as belonging to

748 the type species *Ancora sagittata* (GenBank accessions KX982501 – 3) because they appear

749 more widespread then ribotype 2, the only sequence from the sample WSBS 2010, which was

750 annotated as *Ancora* cf. *sagittata*, KX982504.

751 Nine environmental sequences closely related to *A. sagittata* may belong to other cryptic species

752 within the same morphotype since, in the tree, they are flanked by the sequences obtained from

753 the same morphospecies, although belonging to the different ribotypes (Fig. 8). Two other

754 environmental sequences, M60E1D07 and M23E1H07, which form a sister branch to the *A.*

755 *sagittata* cluster, putatively belong to another species in the genus *Ancora*.

756 Hasselmann (1927) proposed parthenogenetic formation of oocysts (solitary encystment of

757 gamonts) in *A. sagittata* based on the behaviour of solitary mature gamonts due to similarities to

758 late syzygy in other gregarines (flexion of the body and circular gliding (rotation)). Although

759 solitary encystment and gametogenesis have not been described in *A. sagittata*, it is possible that

760 its morphospecies exist as parthenogenetic clones, while others likely have a regular sexual

761 cycle. This possibility could explain branch length differences within the *A. sagittata* group (Fig.

762 8) and the sympatric coexistence of two cryptic species (ribotypes) of *A. sagittata* at the WSBS.

763 Thus, the hypothesis of a cryptic species complex in *A. sagittata* should be reconsidered both in

764 terms of reproduction modes and the presence of a species complex in its host, *Capitella capitata*

765 (Grassle & Grassle, 1976).

766 **Taxonomic actions: modification of gregarine and eugregarine diagnoses and**

767 **establishment of the new superfamily Ancoroidea.**

768 **Phylum Apicomplexa** Levine, 1970

769 **Subphylum Sporozoa** Leuckart, 1879

770 **Class Gregarinomorpha** Grassé, 1953, emend.

771 **Diagnosis.** Sporozoa. Gamont coupling (syzygy) followed by encystment (formation of

772 gametocyst); progamic mitoses in both gamonts; gametogenesis and fecundation within the

773 gametocyst; anisogamy is characteristic: female gametes are non-flagellated, male gametes

774   usually flagellated, bear 1 flagellum; oocysts without sporocysts (sporozoites lie free within the

775   oocyst, not in its internal compartments). Typical representatives are epicellular intestinal

776   parasites of invertebrates, mainly Trochozoa, Arthropoda, and Deuterostomia, including lower

777   Chordata (Tunicata).

778   **Order Eugregarinida** Léger, 1900, emend.

779   **Diagnosis.** Gregarinomorpha. Typically: gliding locomotion of the gamonts likely provided by

780   epicytic crests, i.e. longitudinal pellicular folds of complex structure (rippled dense structures

781   (apical arcs) and 12-nm apical filaments within the tops of the crests, the links of internal lamina

782   in their bases); the attachment apparatus is chiefly an epimerite that develops ahead of the

783   sporozoite apical complex, which disappears in the beginning of trophozoite formation; the

784   epimerite is mostly absent in mature gamont (degenerated, retracted or discarded). A number of

785   representatives exhibit a septate morphology of the trophozoites: there are one or more fibrillar

786   septa that separate the cell into compartments – protomerite and deutomerite.

787   **Note 1.** The morphological synapomorphies of the Eugregarinida compared with the

788   plesiomorphies of Archigregarinida have been presented as diagrams in Fig. 11.

789   **Note 2.** There are a number of aberrant representatives, which lose the typical structure of the

790   attachment apparatus and epicyte. This is frequently correlated with the transition from the

791   intestinal to coelomic parasitism (e.g., Monocystidae and Urosporidae).

792   **Note 3.** The order includes several superfamilies (see below), which were erected after molecular

793   phylogenetic analyses of SSU rDNA. However, recent molecular data do not encompass the

794   complete taxonomical diversity of eugregarines, and we expect additional superfamilies to be

795   established in the future. The current composition and characteristics of the superfamilies

796   described to date are consistent with the characteristics of corresponding molecular phylogenetic

797   lineages (Table 3).

798   **Superfamily Actinocephaloidea**

799   **Note.** Since molecular data corroborate the assumption of Grassé about the origin of

800   neogregarines from actinocephalids, the superfamily must also include neogregarines.

801   Consequently, the order Neogregarinida should be abolished.

802   **Superfamily Stylocephaloidea**

803   **Superfamily Gregarinoidea**

804   **Superfamily Cephaloidophoroidea**

805 **Note.** This name was first proposed in Rueckert et al. (2011), but it was changed by Cavalier-

806 Smith (2014) into Porosporoidea because of the earlier establishment of the family Porosporidae

807 Labbé, 1899 than Cephaloidophoridae Kamm, 1922. However, considering that the guidelines of

808 the International Code of zoological nomenclature have a recommendatory (suggestive) nature

809 for superfamilies, the name Cephaloidophoroidea can be accepted and appears to be more

810 appropriate: the SSU rDNA of *Cephaloidophora communis*, the type species of this family, was

811 sequenced (Rueckert et al., 2011), unlike type species of all of the other families included in this

812 clade. Additionally, these families are more or less problematic and require revision: e.g., at the

813 last time, the family Thiriotiidae was separated from the Porosporidae based on the shape of the

814 unusual syzygy (Schrével & Desportes, 2013b); the DNA sequences of the true representatives

815 of the Porosporidae (*Porospora*, *Nematopsis*) are unavailable.

816 **Superfamily Lecudinoidea**

817 **Note.** Cavalier-Smith (2014) used the name Urosporoidea, but Lecudinoidea appears to be more

818 appropriate because the SSU rDNA sequence of the type species of the family Lecudinidae,

819 *Lecudina pellucida*, is available. In contrast, DNA sequences of the type species of the family

820 Urosporidae, *Urospora nemertis*, are unavailable. Additionally, the Urosporidae is the aberrant

821 family (see above), which also could present nomenclatural problems: while other urosporids

822 chiefly parasitize the coelom of echinoderms and polychaetes, the type species is an intestinal

823 parasite of the nemertean *Baseodiscus delineatus*, and its taxonomical position and status may be

824 questionable.

825 **Superfamily Ancoroidea**, superfam. nov.

826 **Diagnosis.** Eugregarinida. Aseptate forms parasitize marine polychaetes, mainly the family

827 Capitellidae; tightly adjacent epicytic crests; gliding motility. Molecular data: the robust SSU

828 rDNA clade.

829 **Note.** For more grounded diagnoses of the entire group and subgroups within it, additional data

830 are necessary, e.g., the ultrastructure of *Polyplicarium* spp.

831 **Family Polyplicariidae** Cavalier-Smith, 2014

832 **Diagnosis (preliminary).** Ancoroidea. Characteristics of the type genus *Polyplicarium*.

833 **Genus *Polyplicarium* Wakeman et Leander 2013.** Ovoid to elongate trophozoites with a blunt

834 anterior end. The posterior end is either blunt or tapers to a point. Longitudinal epicytic folds

835 with a density of 4–5 per 1 μm; most trophozoites also have a distinct region of wider, shallower

836   epicytic folds; gliding locomotion; other life-cycle stages are unknown. There are four named

837   species.

838   **Note.** The family likely includes at least 2 additional undescribed genera that are represented

839   only by environmental sequences.

840   **Family Ancoridae** Simdyanov, fam. nov.

841   **Diagnosis.** Ancoroidea. Monotypic, characters of the type genus *Ancora*.

842   **Genus *Ancora* Labbé, 1899.** Trophozoites and gamonts with 2 lateral projections giving them

843   appearance of an anchor. Gliding locomotion. Growing trophozoites with a bulbous epimerite.

844   Syzygy unknown. Simple gametocyst dehiscence by rupture. Oocysts ovoid. There are three

845   named species, but 2 of them are questionable.

846   **Note.** The type morphospecies *Ancora sagittata* (Leuckart, 1860) Labbé, 1899 likely is a

847   complex of cryptic sibling species.

848

849   **Conclusion**

850

851   The results of our work point to several new directions of importance to gregarine research. The

852   molecular phylogenies based on the SSU rDNA alone firmly delimit several major lineages

853   (superfamilies) in eugregarines but not their suborders (Aseptata and Septata), a finding that is

854   more consistent with Grassé's taxonomical scheme (Grassé, 1953) than with the current

855   taxonomy established by Levine and the followers (Levine, 1985; Levine 1988; Perkins et al.,

856   2000). The results also corroborate other Grassé's assumptions (Grassé, 1953): (i) the

857   polyphyletic origin of neogregarines, likely from different representatives of the eugregarine

858   family Actinocephalidae; (ii) the secondary origin of some aseptate gregarines from septate

859   ancestors; and (iii) the importance of gregarine co-evolution with their hosts. The molecular

860   evidence indicates that both the life cycle peculiarities (presence or absence of merogony) and

861   the general morphology of eugregarine trophozoites (septate or aseptate), which are broadly

862   employed in the current eugregarine taxonomy, are unreliable. However, SSU rDNA

863   phylogenies do not resolve their deeper branching and do not allow for testing the monophyly of

864   Eugregarinida, Archigregarinida, and all gregarines, possibly due to their explosive evolutionary

865   radiation and/ or rapid sequence evolution that resulted in numerous long branches in molecular

866   phylogenies suffering from long-branch attraction (LBA) artefacts.

867   The near-complete rDNA operon likely provides an increased resolution over SSU rDNA and

868   appears more resilient to LBA (Simdyanov, Diakin & Aleoshin, 2015; Fig. 8). Although neither

869   of the markers resolves deep relationships among gregarines recently, a representatively sampled

870   rDNA operon is likely to provide a more reliable test of the group's morphological evolution in

871   the future. The best strategy for the development of gregarine phylogeny (*sensu lato*) and high-

872   rank taxonomy seems to be reconciling and combining morphological evidence with

873   unambiguous molecular data such as well-resolved deep branching in the molecular phylogenetic

874   trees of gregarines – probably with the use of concatenated nuclear markers compiled from

875   transcriptomic and genomic datasets. However, until such datasets become available, we propose

876   to treat the shared ultrastructural characteristics of their epicytic crests and epimerite as

877   synapomorphies of eugregarines and, consequently, as evidence for their monophyly by

878   following the principles of cladistics. At the same time, relying on the firm molecular

879   phylogenetic support and following previous works, we also propose to abolish the suborders

880   Aseptata and Septata within the order Eugregarinida (Simdyanov & Diakin, 2013; Cavalier-

881   Smith, 2014) and accept the robust molecular phylogenetic lineages as superfamilies instead

882   (Clopton, 2009; Rueckert et al., 2011; Simdyanov & Diakin, 2013; Cavalier-Smith, 2014). On

883   the same ground, we acknowledge the abolition of the order Neogregarinida (Simdyanov &

884   Diakin, 2013; Cavalier-Smith, 2014), which apparently comprises divergent representatives of

885   the eugregarine superfamily Actinocephaloidea. Finally, following this assertion, we also

886   propose to remove the absence of merogony from the diagnostic criteria of eugregarines, despite

887   that the current gregarine taxonomy relies heavily on this characteristic (Levine, 1985; Perkins et

888   al., 2000; Adl et al., 2012). The majority of these proposals receive molecular phylogenetic and

889   ultrastructural backing and although some are more preliminary than others (the monophyly of

890   eugregarines will require thorough testing, e.g., by evidence from multigene molecular

891   phylogenetic analyses), they altogether represent a next step in a much needed revision of the

892   gregarine taxonomy and evolution.

893

894   **Acknowledgements**

895

PeerJ

898 (http://parallel.ru/cluster) to perform the phylogenetic computations. DNA sequencing (Sanger)

899 was performed at the DNA sequencing centre "Genome" (Engelhardt Institute of Molecular

900 Biology, Russian Academy of Sciences, www.genome-centre.ru). The electron microscopy

901 studies were performed at the Laboratoty of electron microscopy of Faculty of Biology,

902 Lomonosov Moscow State University, and at the Centre of Electron Miscroscopy of I.D. Papanin

903 Institute for Biology of Inland Waters, Russian Academyof Sciences.

904 The authors are grateful to Dr. Gulnara Mirzayeva (Institute of Gene Pool of Plants and Animals,

905 Uzbek Academy of Sciences, Republic of Uzbekistan) for her assistance carrying out the

906 experimental work. We thank Maria Logacheva (Belozersky Institute of Physico-Chemical

907 Biology, Russia) for the help with genomic sequencing. Many thanks are extended to Prof.

908 Isabelle Florent and Dr. Isabelle Desportes (Muséum National d'Histoire Naturelle, France) for

909 their valuable comments. We thank to Kevin Wakeman (Hokkaido University, Sapporo, Japan)

910 and Jan Janouškovec (University College London, UK) for proofreading and commenting on the

911 manuscript.

912

913 **References**

914

915 Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown M, Burki F, Dunthorn M,
916 Hampl V, Heiss A, Hoppenrath M, Lara E, leGall L, Lynn DH, McManus H, Mitchell EAD,
917 Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S, Shadwick L, Schoch C, Smirnov
918 A, Spiegel FW. 2012. The revised classification of eukaryotes. *The Journal of eukaryotic
919 microbiology* 59:429-514. DOI: 10.1111/j.1550-7408.2012.00644.x
920 Alfaro ME, Zoller S, Lutzoni F. 2003. Bayes or bootstrap? A simulation study comparing the
921   performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in
922   assessing phylogenetic confidence. *Molecular Biology and Evolution* 20:255-266. DOI:
923   10.1093/molbev/msg028
924 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped
925   BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic
926   Acids Research* 25:3389-3402
927 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko
928   SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev
929   MA, Pevzner PA. 2012. SPAdes: A new genome assembly algorithm and its applications
930   to single-cell sequencing. *Journal of Computational Biology* 19:455-477. DOI:
931   10.1089/cmb.2012.0021
932 Baudoin J. 1969. Sur l'ultrastructure de la région antérieure de la Grégarine *Ancyrophora
933   puytoraci* B. *Protistologica* 5:431-430
934 Bergsten J. 2005. A review of long-branch attraction. *Cladistics* 21:163-193. DOI:
935   10.1111/j.1096-0031.2005.00059.x

936 Boere AC, Rijpstra WI, De Lange GJ, Sinninghe Damsté JS, Coolen MJ. 2011. Preservation
937   potential of ancient plankton DNA in Pleistocene marine sediments. *Geobiology* 9:377-
938   393. DOI: 10.1111/j.1472-4669.2011.00290.x.
939 Carreno RA, Martin DS, Barta JR. 1999. *Cryptosporidium* is more closely related to the
940   gregarines than to coccidia as shown by phylogenetic analysis of apicomplexan parasites
941   inferred using small-subunit ribosomal RNA gene sequences. *Parasitology Research*
942   85:899-904
943 Cavalier-Smith T. 2014. Gregarine site-heterogeneous 18S rDNA trees, revision of gregarine
944   higher classification, and the evolutionary diversification of Sporozoa. *European Journal*
945   *of Protistology* 50:472-495. DOI: 10.1016/j.ejop.2014.07.002
946 Cecconi J. 1905. Sur l'*Anchorina sagittata* Leuck., parasite de la *Capitella capitata* O. Fabr.
947   *Archiv für Protistenkunde* 6:230-244
948 Clausen C. 1993. *Ancora prolifera* sp. n., a gregarine parasite of *Microphthalmus*
949   *ephippiophorus* Clausen (Polychaeta, Hesionidae). *Zoologica Scripta* 22:111-115
950 Clopton RE. 2009. Phylogenetic relationships, evolution, and systematic revision of the septate
951   gregarines (Apicomplexa: Eugregarinorida: Septatorina). *Comparative Parasitology*
952   76:167-190. DOI: 10.1654/4388.1
953 Coleman AW. 2000. The significance of a coincidence between evolutionary landmarks found in
954   mating affinity and a DNA sequence. *Protist* 151:1-9. DOI: 10.1078/1434-4610-00002
955 Coleman AW. 2009. Is there a molecular key to the level of "biological species" in eukaryotes?
956   A DNA guide. *Molecular Phylogenetics and Evolution* 50:197-203. DOI:
957   10.1016/j.ympev.2008.10.008
958 Corbel J-C, Desportes I, Théodoridès J. 1979. Étude de *Gonospora beloneides* (Ming.) (=
959   *Lobianchella beloneides* Ming.) (Grégarine Urosporidae), parasite coelomique d'une
960   Alciopidae (Polychaeta) et remarques sur d'autres Grégarines d'Alciopidae.
961   *Protistologica* 15:55-65
962 Dallai R, Talluri MV. 1983. Freeze-fracture study of the gregarine trophozoite: I. The top of the
963   epicyte folds. *Bolletino di Zoologia* 50. DOI: 10.1080/11250008309439448
964 Desportes I. 1969. Ultrastructure et développement des Grégarines du genre *Stylocephalus*.
965   *Annales des Sciences naturelles (Zoologie et Biologie animale, Série 12)* 11:31-96
966 Desportes I, Théodoridès J. 1985. Particularités cytologiques d'*Uradiophora maetzi* Théod. et
967   Desp. (Eugregarina, Uradiophoridae) parasite du Mysidacé bathypelagique
968   *Gnathophausia zoea* WS. *Annales des Sciences naturelles (Zoologie et Biologie animale,*
969   *Série 13)* 7:199-213
970 Desportes I, Théodoridès J. 1986. *Cygnicollum lankesteri* n. sp., Grégarine (Apicomlexa,
971   Lecudinidae) parasite des Annélides Polychètes *Laetmonice hystrix* et *L. producta*;
972   particularités de l'appareil de fixation et implications taxonomiques. *Protistologica*
973   22:47-60
974 Desportes I, Vivarès CP, Théodoridès J. 1977. Intérêt taxinomique de l'ultrastructure épicytaire
975   chez *Ganymedes* Huxley, *Porospora* Schneider et *Thiriotia* n.g., Eurégarines parasites de
976   Crustacés. *Annales des Sciences naturelles (Zoologie et Biologie animale, Série 12)*
977   19:261-277
978 Devauchelle G. 1968. Étude ultrastructurale du développement des Grégarines du *Tenebrio*
979   *molitor* L. *Protistologica* 4:313-332
980 Diakin A, Paskerova GG, Simdyanov TG, Aleoshin VV, Valigurová A. 2016. Morphology and
981   molecular phylogeny of coelomic gregarines (Apicomplexa) with different types of

982   motility: *Urospora ovalis* and *U. travisiae* from the polychaete *Travisia forbesii*. *Protist*
983   167:279-301. DOI: http://dx.doi.org/10.1016/j.protis.2016.05.001
984 Dyakin AY, Simdyanov TG. 2005. The cortical zone of skittle-like cells of *Urospora chiridotae*,
985   a gregarine from an apode holothuria *Chiridota laevis*. *Protistology* 4:97-105
986 Diakin A, Wakeman KC, Valigurová A. 2017. Description of *Ganymedes yurii* sp. n.
987   (Ganymedidae), a new gregarine species from the Antarctic amphipod *Gondogeneia* sp.
988   (Crustacea). *Journal of Eukaryotic Microbiology* 64:56-66. DOI: 10.1111/jeu.12336
989 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
990   throughput. *Nucleic Acids Research* 35:1792-1797. DOI: 10.1093/nar/gkh340
991 Edgcomb VP, Kysela DT, Teske A, de Vera Gomez A, Sogin ML. 2002. Benthic eukaryotic
992   diversity in the Guaymas Basin hydrothermal vent environment. *Proceedings of the*
993   *National Academy of Sciences of the United States of America* 99:7658-7662. DOI:
994   10.1073/pnas.062186399
995 Frolov AO. 1991. *The world fauna of gregarines. Family Monocystidae*. Leningrad: Academy of
996   Sciences of the USSR Press.
997 Garman KM, Rubelmann H, Karlen DJ, Wu T, Garey JR. 2011. Comparison of an inactive
998   submarine spring with an active nearshore anchialine spring in Florida. *Hydrobiologia*
999   677:65-87. DOI: 10.1007/s10750-011-0740-2
1000 Ghazali M, Philippe M, Deguercy A, Gounon P, Gallo JM, Schrével J. 1989. Actin and spectrin-
1001   like (M_r = 260-240 000) proteins in gregarines. *Biology of the Cell* 67:173-184. DOI:
1002   10.1111/j.1768-322X.1989.tb00860.x
1003 Grassé P-P. 1953. Classe des Grégarinomorphes. In: Grassé P-P, ed. *Traité de Zoologie*. Paris:
1004   Masson, 550-690.
1005 Grassle J, Grassle JF. 1976. Sibling species in the marine pollution indicator *Capitella*
1006   (polychaeta). *Science* 192:567-569. DOI: 10.1126/science.1257794
1007 Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis
1008   program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41:95-98
1009 Hasselmann G. 1918. Contribuição para o estudo das gregarinas. *Ancora lutzi* sp. n. *Brasil-*
1010   *Medico* 32:249
1011 Hasselmann G. 1927. Ciclo evolutivo de *Ancora sagittata* (Leuck) 1842. *Boletim do Instituto*
1012   *Brasileiro de Sciencias* 3:34-40
1013 Hildebrand HF. 1976. Elektronenmikroskopische Untersuchungen an den Entwicklungstadien
1014   des Trophozoiten von *Didymophyes gigantea* (Sporozoa, Gregarinida). 1. Die
1015   Feinstruktur des Proto- und Epimeriten und die Beziehung zwischen Wirt und Parasit.
1016   *Zeitschrift für Parasitenkunde* 49:193-215. DOI: 10.1007/BF00380590
1017 Janouškovec J, Tikhonenkov DV, Burki F, Howe AT, Kolísko M, Mylnikov AP, Keeling PJ.
1018   2015. Factors mediating plastid dependency and the origins of parasitism in
1019   apicomplexans and their close relatives. *Proceedings of the National Academy of*
1020   *Sciences* 112:10200-10207. DOI: 10.1073/pnas.1423790112
1021 King CA. 1988. Cell motility of sporozoan protozoa. *Parasitology Today* 4:315-319. DOI:
1022   10.1016/0169-4758(88)90113-5
1023 Krylov MV, Dobrovolskij AA. 1980. Macrosystem and phylogeny of the Sporozoa. In: Krylov
1024   MV and Starobogatov YI, eds. *Principles of the construction of the macrosystem of the*
1025   *unicellular animals*. Leningrad: Academy of Sciences of the USSR Press, 62-74.
1026 Kuvardina ON, Simdyanov TG. 2002. Fine structure of syzygy in *Selenidium pennatum*
1027   (Sporozoa, Archigregarinida). *Protistology* 2:169-177

1028 Labbé A. 1899. *Sporozoa*. Berlin: R.Friedlander und Sohn.
1029 Landers SC, Leander BS. 2005. Comparative surface morphology of marine coelomic gregarines
1030   (Apicomplexa, Urosporidae): *Pterospora floridiensis* and *Pterospora schizosoma*.
1031   *Journal of Eukaryotic Microbiology* 52:23-30
1032 Leander BS. 2007. Molecular phylogeny and ultrastructure of *Selenidium serpulae*
1033   (Apicomplexa, Archigregarinia) from the calcareous tubeworm *Serpula vermicularis*
1034   (Annelida, Polychaeta, Sabellida). *Zoologica Scripta* 36:213-227. DOI: 10.1111/j.1463-
1035   6409.2007.00272.x
1036 Leander BS, Clopton RE, Keeling PG. 2003. Phylogeny of gregarines (Apicomplexa) as inferred
1037   from SSU rDNA and beta-tubulin. *International Journal of Systematic and Evolutionary*
1038   *Microbiology* 53:345-354. DOI: 10.1099/ijs.0.02284-0
1039 Leander BS, Harper JT, Keeling PG. 2003. Molecular phylogeny and surface morphology of
1040   marine aseptate gregarines (Apicomplexa): *Selenidium* spp. and *Lecudina* spp. *Journal of*
1041   *Parasitology* 89:1191-1205. DOI: 10.1645/GE-3155
1042 Leander BS, Lloyd SAJ, Marshall W, Landers SC. 2006. Phylogeny of marine gregarines
1043   (Apicomplexa) — *Pterospora*, *Lithocystis* and *Lankesteria* — and the origin(s) of
1044   coelomic parasitism. *Protist* 157:45-60. DOI: 10.1016/j.protis.2005.10.002
1045 Lepelletier F, Karpov SA, Le Panse S, Bigeard E, Skovgaard A, Jeanthon C, Guillou L. 2014.
1046   *Parvilucifera rostrata* sp. nov. (Perkinsozoa), a novel parasitoid that infects planktonic
1047   dinoflagellates. *Protist* 165:31-49. DOI: 10.1016/j.protis.2013.09.005
1048 Levine ND. 1971. Uniform terminology for the protozoan subphylum Apicomplexa. *Journal of*
1049   *Protozoology* 18:352-355. DOI: 10.1111/j.1550-7408.1971.tb03330.x
1050 Levine ND. 1977. Revision and checklist of the species (other than *Lecudina*) of the aseptate
1051   gregarine family Lecudinidae. *Journal of Protozoology* 24:41-52. DOI: 10.1111/j.1550-
1052   7408.1977.tb05279.x
1053 Levine ND. 1985. Phylum 2. Apicomplexa Levine, 1970. In: Lee JJ, Hutner SH, and Bovee EC,
1054   eds. *An Illustrated Guide To The Protozoa*. Kansas: Society of Protozoologists, 322-374.
1055 Levine ND. 1988. *The protozoan phylum Apicomlexa*. Boca Raton, FL: CRC Press.
1056 Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B. 2012. RobiNA: a user-
1057   friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids*
1058   *Research* 40:W622-W627. DOI: 10.1093/nar/gks540
1059 Mackenzie C, Walker M. 1983. Substrate contact, mucus, and Eugregarine gliding. *Journal of*
1060   *Protozoology* 30:3-8. DOI: 10.1111/j.1550-7408.1983.tb01024.x
1061 MacMillan WG. 1973. Conformation changes in the cortical region during peristaltic movements
1062   of a gregarine trophozoite. *Journal of Protozoology* 20:267-274. DOI: 10.1111/j.1550-
1063   7408.1973.tb00874.x
1064 Marquès A. 1979. *Actinocephalus dujardini* Schneider 1875. Eugrégarine parasite de *Lithobius*
1065   (Myriapoda, Chilopoda): ultrastructure de l'épimérite. *Annales des Sciences naturelles*
1066   *(Zoologie et Biologie animale, Série 13)* 1:161-168
1067 Medlin L, Elwood HJ, Stickel S, Sogin ML. 1988. The characterization of enzymatically
1068   amplified eukaryotic 16S-like rRNA-coding regions. *Gene* 71:491-499
1069 Mikhailov KV, Simdyanov TG, Aleoshin VV. 2017. Genomic survey of a hyperparasitic
1070   microsporidian *Amphiamblys* sp. (Metchnikovellidae). *Genome Biology and Evolution*
1071   9:454-457. DOI: 10.1093/gbe/evw235

1072　Miles HB. 1968. The fine structure of the epicyte of the acephaline gregarines *Monocystis*
1073　　　　*lumbrici-olidi,* and *Nematocystis magna*: observations by electron microscope. *Revista*
1074　　　　*iberica de Parasitologia* 28:455-465

1075　Müller T, Philippi N, Dandekar T, Schultz J, Wolf M. 2007. Distinguishing species. *RNA*
1076　　　　13:1469–1472. DOI: 10.1261/rna.617107

1077　Ormierès R. 1971. Une Grégarine paradoxale, *Gigaductus anchi* Tuz. et Orm., 1966:
1078　　　　Ultrastructure de la schizogonie et position systématique des Gigaductidae Filipponi
1079　　　　1948. *Protistologica* 7:261-271

1080　Ormierès R. 1977. *Pyxinia firmus* (Leger, 1892), Eugrégarine parasite du Coléoptère *Dermestes*
1081　　　　*frischi* Kugel. Étude ultrastructurale. *Zeitschrift für Parasitenkunde* 53:13-22. DOI:
1082　　　　10.1007/BF00383110

1083　Ormierès R, Daumal J. 1970. Étude ultrastructurale de la partie antérieure d'*Epicavus araeoceri*
1084　　　　Ormières et Daumal, Eugrégarine parasite du Coléoptère Anthribidae *Araeocerus*
1085　　　　*fasciculatus* de Geer. *Protistologica* 6:97-111

1086　Ormières R, Daumal J. 1970. Données ultrastructurales sur *Epicavus araeoceri* Orm. Daum.,
1087　　　　eugrégarine parasite d`*Araeocerus fasciculatus* de Geer (Coléoptere; Anthribidae).
1088　　　　*Comptes rendus hebdomadaires des séances de l'Académie des sciences, Paris (Série D)*
1089　　　　270:2451-2453

1090　Ormierès R, Marquès A. 1976. Fixation a leurs hôtes de quelques Dactylophoridae Eugrégarines
1091　　　　parasites de Myriapodes Chilopodes. *Protistologica* 12:415-424

1092　Ormierès R, Marquès A, Puisségur C. 1977. *Trichorhynchus pulcher* Schneider, 1882,
1093　　　　Eugrégarine parasite du *Scutigera coleopterata* L. Cycle, ultrastructure, systématique.
1094　　　　*Protistologica* 13:407-417

1095　Orsi W, Song YC, Hallam S, Edgcomb VP. 2012. Effect of oxygen minimum zone formation on
1096　　　　communities of marine protists. *The ISME Journal* 6:1586-1601. DOI:
1097　　　　10.1038/ismej.2012.7

1098　Ouassi MA, Porchet-Henneré E. 1978. Étude ultrastructurale de mucron d'une Grégarine du
1099　　　　genre *Lecudina*, parasite intestinal d'*Audoinia tentaculata* (Annélide Polychète) et de ses
1100　　　　rapports avec la cellule hôte. *Protistologica* 14:39-52

1101　Pawlowski J, Bolivar I, Fahrni JF, Cavalier-Smith T, Gouy M. 1996. Early origin of foraminifera
1102　　　　suggested by SSU rRNA gene sequences. *Molecular Biology and Evolution* 13:445-450

1103　Perkins FO, Barta JR, Clopton RE, Peirce MA, Upton SJ. 2000. Phylum Apicomplexa. In: Lee
1104　　　　JJ, Leedale GF, and Bradbury P, eds. *An Illustrated Guide to the Protozoa*. Lawrence, KS
1105　　　　(USA): Society of Protozoologists, 190-370.

1106　Pitelka DR. 1963. *Electron-microscopic structure of Protozoa*. London, New York: Pergamon,
1107　　　　Macmillan.

1108　Reynolds ES. 1963. The use of lead citrate at high pH as an electron opaque stain in electron
1109　　　　microscopy. *Journal of Cell Biology* 17:208–212

1110　Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L,
1111　　　　Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic
1112　　　　inference and model choice across a large model space. *Systematic Biology* 61:539-542.
1113　　　　DOI: 10.1093/sysbio/sys029

1114　Rueckert S, Chantangsi C, Leander BS. 2010. Molecular systematics of marine gregarines
1115　　　　(Apicomplexa) from North-eastern Pacific polychaetes and nemerteans, with descriptions
1116　　　　of three novel species: *Lecudina phyllochaetopteri* sp. nov., *Difficilina tubulani* sp. nov.

1117   and *Difficilina paranemertis* sp. nov. *International Journal of Systematic and*
1118   *Evolutionary Microbiology* 60:2681-2690. DOI: 10.1099/ijs.0.016436-0
1119 Rueckert S, Leander BS. 2008. Morphology and phylogenetic position of two novel marine
1120   gregarines (Apicomplexa, Eugregarinorida) from the intestines of North-eastern Pacific
1121   ascidians. *Zoologica Scripta* 37:637–645. DOI: 10.1111/j.1463-6409.2008.00346.x
1122 Rueckert S, Leander BS. 2009. Molecular phylogeny and surface morphology of marine
1123   archigregarines (Apicomplexa), *Selenidium* spp., *Filipodium phascolosomae* n. sp. and
1124   *Platyproteum* n. g. and comb. from North-Eastern Pacific peanut worms (Sipuncula).
1125   *Journal of Eukaryotic Microbiology* 56:428-439. DOI: 10.1111/j.1550-
1126   7408.2009.00422.x
1127 Rueckert S, Leander BS. 2010. Description of *Trichotokara nothriae* n. gen. et sp.
1128   (Apicomplexa, Lecudinidae) – an intestinal gregarine of *Nothria conchylega* (Polychaeta,
1129   Onuphidae). *Journal of Invertebrate Pathology* 104:172–179. DOI:
1130   10.1016/j.jip.2010.03.005
1131 Rueckert S, Simdyanov TG, Aleoshin VV, Leander BS. 2011. Identification of a divergent
1132   environmental DNA sequence clade using the phylogeny of gregarine parasites
1133   (Apicomplexa) from crustacean hosts. *PLoS ONE* 6:e18163. DOI:
1134   10.1371/journal.pone.0018163
1135 Rueckert S, Wakeman KC, Jenke-Kodama H, Leander BS. 2015. Molecular systematics of
1136   marine gregarine apicomplexans from Pacific tunicates, with descriptions of five novel
1137   species of Lankesteria. *International Journal of Systematic and Evolutionary*
1138   *Microbiology* 65:2598-2614. DOI: 10.1099/ijs.0.000300
1139 Rueckert S, Wakeman KC, Leander BS. 2013. Discovery of a diverse clade of gregarine
1140   Apicomplexans (Apicomplexa: Eugregarinorida) from Pacific eunicid and onuphid
1141   polychaetes, including descriptions of *Paralecudina* n. gen., *Trichotokara* japonica n. sp.,
1142   and *T. eunicae* n. sp. *Journal of Eukaryotic Microbiology* 60:121-136. DOI:
1143   10.1111/jeu.12015
1144 Santos HF, Cury JC, Carmo FL, Rosado AS, Peixoto RS. 2010. 18S rDNA sequences from
1145   microeukaryotes reveal oil indicators in mangrove sediment. *PLoS ONE* 5:e12437. DOI:
1146   10.1371/journal.pone.0012437
1147 Schrével J. 1968. L'ultrastructure de la région antérieure de la Grégarine *Selenidium* et son intérêt
1148   pour l'étude de la nutrition chez les Sporozoaires. *Journal de Microscopie, Paris* 7:391-
1149   410
1150 Schrével J. 1971a. Observations biologiques et ultrastructurales sur les Selenidiidae et leurs
1151   conséquences sur la systématique des Grégarinomorphes. *Journal of Protozoology*
1152   18:448-479. DOI: 10.1111/j.1550-7408.1971.tb03355.x
1153 Schrével J. 1971b. Contribution à l'étude des Selenidiidae parasites d'Annélides Polychètes. II.
1154   Ultrastructure des quelques trophozoïtes. *Protistologica* 7:101-130
1155 Schrével J, Caigneaux E, Gros D, Philippe M. 1983. The three cortical membranes of the
1156   gregarines. I. Ultrastructural organization of *Gregarina blaberae*. *Journal of Cell Science*
1157   61:151-174
1158 Schrével J, Desportes I. 2013a. Introduction: Gregarines among Apicomlexa. In: Desportes I and
1159   Schrével J, eds. *Treatise on Zoology - Anatomy, Taxonomy, Biology. The Gregarines*.
1160   Leiden: Brill, 7-24.
1161 Schrével J, Desportes I. 2013b. Marine gregarines. In: Desportes I and Schrével J, eds. *Treatise*
1162   *on Zoology - Anatomy, Taxonomy, Biology. The Gregarines*. Leiden: Brill, 197-354.

1163 Schrével J, Desportes I. 2015. Gregarines. In: Mehlhorn H, ed. *Encyclopaedia of Parasitology*.
1164   Heidelberg: Springer, 1-47.
1165 Schrével J, Desportes I, Goldstein S, Kuriyama R, Prensier G, Vávra J. 2013. Biology of
1166   gregarines and their host-parasite interactions. In: Desportes I and Schrével J, eds.
1167   *Treatise on Zoology - Anatomy, Taxonomy, Biology. The Gregarines*. Leiden: Brill, 25-
1168   195.
1169 Schrével J, Valigurová A, Prensier G, Chambouvet A, Florent I, Guillou L. 2016. Ultrastructure
1170   of *Selenidium pendula*, the Type Species of Archigregarines, and Phylogenetic Relations
1171   to Other Marine Apicomplexa. *Protist* 167: 339-368.
1172 Schrével J, Vivier E. 1966. Étude de l'ultrastructure et du rôle de la région antérieure (mucron et
1173   épimérite) de Grégarines parasites d'Annélides Polychètes. *Protistologica* 2:17-28
1174 Simdyanov TG. 1995a. Two new species of gregarines with the aberrant structure of epicyte
1175   from the White Sea. *Parazitologiya* 29:305-315 (in Russian with English summary)
1176 Simdyanov TG. 1995b. Ultrastructure of two species of gregarines of the genus *Lankesteria*
1177   (Eugregarinida: Lecudinidae). *Parazitologiya* 29:424-432 (in Russian with English
1178   Summary)
1179 Simdyanov TG. 2004. *Sphinctocystis phyllodoces* gen. n., sp. n. (Eugregarinida: Lecudinidae) - a
1180   new gregarine from *Phyllodoce citrina* (Polychaeta: Phyllodocidae). *Parazitologiya*
1181   38:322-332 (In Russian with English summary)
1182 Simdyanov TG. 2007. Class Gregarinea Dufour, 1828 - gregarines. In: Alimov AF, Krylov MV,
1183   and Frolov AO, eds. *Protists: Handbook on zoology, Part 2*. St. Petersburg: Nauka, 20-
1184   149 (In Russian with English summary).
1185 Simdyanov TG. 2009. *Difficilina cerebratuli* gen. et sp. n. (Eugregarinida: Lecudinidae) - a new
1186   gregarine species from the nemertean *Cerebratulus barentsi* (Nemertini: Cerebratulidae).
1187   *Parazitologiya* 43:273-287 (In Russian with English summary)
1188 Simdyanov TG, Diakin AY. 2013. Remarks to taxonomy of Eugregarinida (Apicomplexa) as
1189   inferred from 18S rDNA phylogenetic analysis [abstract no. P-109]. XIVth International
1190   Congress of Protistology (Abstract book). Vancouver, BC, Canada. p 134. DOI:
1191   10.13140/RG.2.2.18325.12004
1192 Simdyanov TG, Diakin AY, Aleoshin VV. 2015. Ultrastructure and 28S rDNA phylogeny of two
1193   gregarines: *Cephaloidophora* cf. *communis* and *Heliospora* cf. *longissima* with remarks
1194   on gregarine morphology and phylogenetic analysis. *Acta Protozoologica* 54:241-263.
1195   DOI: 10.4467/16890027AP.15.020.3217
1196 Simdyanov TG, Kuvardina ON. 2007. Fine structure and putative feeding mechanism of the
1197   archigregarine *Selenidium orientale* (Apicomplexa: Gregarinomorpha). *European*
1198   *Journal of Protistology* 43:17-25. DOI: 10.1016/j.ejop.2006.09.003
1199 Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
1200   thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690. DOI:
1201   10.1093/bioinformatics/btl446
1202 Stoeck T, Epstein S. 2003. Novel eukaryotic lineages inferred from small-subunit rRNA analyses
1203   of oxygen-depleted marine environments. *Applied and Environmental Microbiology*
1204   69:2657-2663. DOI: 10.1128/AEM.69.5.2657-2663.2003
1205 Stoeck T, Kasper J, Bunge J, Leslin C, Ilyin V, Epstein S. 2007. Protistan diversity in the Arctic:
1206   a case of paleoclimate shaping modern biodiversity? *PLoS ONE* 2:e728. DOI:
1207   10.1371/journal.pone.0000728

1208 Stoeck T, Taylor GT, Epstein SS. 2003. Novel eukaryotes from the permanently anoxic Cariaco
1209   Basin (Caribbean Sea). *Applied and Environmental Microbiology* 69:5656-5663. DOI:
1210   10.1128/AEM.69.9.5656-5663.2003

1211 Takishita K, Yubuki N, Kakizoe N, Inagaki Y, Maruyama T. 2007. Diversity of microbial
1212   eukaryotes in sediment at a deep-sea methane cold seep: surveys of ribosomal DNA
1213   libraries from raw sediment samples and two enrichment cultures. *Extremophiles* 11:563-
1214   576. DOI: 10.1007/s00792-007-0068-z

1215 Tronchin G, Schrével J. 1977. Chronologie des modifications ultrastructurales au cours de la
1216   croissance de *Gregarina blaberae*. *Journal of Protozoology* 24:67-82. DOI:
1217   10.1111/j.1550-7408.1977.tb05282.x

1218 Valigurová A, Hofmannová L, Koudela B, Vávra J. 2007. An ultrastructural comparison of the
1219   attachment sites between *Gregarina steini* and *Cryptosporidium muris*. *Journal of*
1220   *Eukaryotic Microbiology* 54:495-510. DOI: 10.1111/j.1550-7408.2007.00291.x

1221 Valigurová A, Koudela B. 2005. Fine structure of trophozoites of the gregarine *Leidyana*
1222   *ephestiae* (Apicomplexa: Eugregarinida) parasitic in *Ephestia kuehniella* larvae
1223   (Lepidoptera). *European Journal of Protistology* 41:209-218. DOI:
1224   10.1016/j.ejop.2005.05.005

1225 Valigurová A, Michalková V, Koudela B. 2009. Eugregarine trophozoite detachment from the
1226   host epithelium via epimerite retraction: fiction or fact? *International Journal for*
1227   *Parasitology* 39:1235-1242. DOI: 10.1016/j.ijpara.2009.04.009

1228 Valigurová A, Vaškovicová N, Musilová N, Schrével J. 2013. The enigma of eugregarine
1229   epicytic folds: where gliding motility originates? *Frontiers in Zoology* 10:1-28. DOI:
1230   10.1186/1742-9994-10-57

1231 Van der Auwera G, Chapelle S, De Wachter R. 1994. Structure of the large ribosomal subunit
1232   RNA of *Phytophtora megasperma*, and phylogeny of the oomycetes. *FEBS Letters*
1233   338:133-136. DOI: 10.1016/0014-5793(94)80350-1

1234 Vávra J. 1969. *Lankesteria barretti* n.sp. (Eugregarinida, Diplocystidae), a parasite of the
1235   mosquito *Aedes triseriatus* (Say) and a review of the genus *Lankesteria* Mingazzini.
1236   *Journal of Protozoology* 16:546-570. DOI: 10.1111/j.1550-7408.1969.tb02314.x

1237 Vávra J, Small EB. 1969. Scanning electron microscopy of gregarines (Protozoa, Sporozoa) and
1238   its contribution to the theory of gregarine movement. *Journal of Protozoology* 16:745-
1239   757. DOI: 10.1111/j.1550-7408.1969.tb02338.x

1240 Vinckier D. 1969. Organisation ultrastructurale corticale de quelques Monocystidées parasites du
1241   ver Oligochète *Lumbricus terrestris* L. *Protistologica* 5:505-517

1242 Vivier E. 1968. L'organisation ultrastructurale corticale de la Grégarine *Lecudina pellucida*; ses
1243   rapports avec l' alimentation et la locomotion. *Journal of Protozoology* 15:230-246. DOI:
1244   10.1111/j.1550-7408.1968.tb02115.x

1245 Vivier E, Devauchelle G, Petitprez A, Porchet-Henneré E, Prensier G, Schrével J, Vinckier D.
1246   1970. Observations de Cytologie comparée chez les Sporozoaires. I. - Les structures
1247   superficielles chez les formes végétatives. *Protistologica* 6:127-150

1248 Vorobyeva IG, Dyakin AY. 2011. The fine structure of the cortical zone in the gregarine
1249   *Bothriopsides histrio* (Eugregarinida: Actinocephalidae). *Parazitologiya* 45:220-233

1250 Wakeman KC, Heintzelman MB, Leander BS. 2014. Comparative ultrastructure and molecular
1251   phylogeny of *Selenidium melongena* n. sp. and *S. terebellae* Ray 1930 demonstrate niche
1252   partitioning in marine gregarine parasites (Apicomplexa). *Protist* 165:493-511. DOI:
1253   10.1016/j.protis.2014.05.007

1254 Wakeman KC, Leander BS. 2012. Molecular phylogeny of pacific archigregarines
1255   (Apicomplexa), including descriptions of *Veloxidium leptosynaptae* n. gen., n. sp., from
1256   the sea cucumber *Leptosynapta clarki* (Echinodermata), and two new species of
1257   *Selenidium*. *Journal of Eukaryotic Microbiology* 59:232-245. DOI: 10.1111/j.1550-
1258   7408.2012.00616.x
1259 Wakeman KC, Leander BS. 2013a. Identity of environmental DNA sequences using descriptions
1260   of four novel marine gregarine parasites, *Polyplicarium* n. gen. (Apicomplexa), from
1261   capitellid polychaetes. *Marine Biodiversity* 43:133-147. DOI: 10.1007/s12526-012-0140-
1262   5
1263 Wakeman KC, Leander BS. 2013b. Molecular phylogeny of marine gregarine parasites
1264   (Apicomplexa) from tube-forming polychaetes (Sabellariidae, Cirratulidae, and
1265   Serpulidae), including descriptions of two new species of *Selenidium*. *Journal of
1266   Eukaryotic Microbiology* 60:514-525. DOI: 10.1111/jeu.12059
1267 Wakeman KC, Reimer JD, Jenke-Kodama H, Leander BS. 2014. Molecular phylogeny and
1268   ultrastructure of *Caliculium glossobalani* n. gen. et sp. (Apicomplexa) from a Pacific
1269   *Glossobalanus minutus* (Hemichordata) confounds the relationships between marine and
1270   terrestrial gregarines. *Journal of Eukaryotic Microbiology* 61:343-353. DOI:
1271   10.1111/jeu.12114
1272 Warner FD. 1968. The fine structure of *Rhynchocystis pilosa* (Sporozoa, Eugregarinida). *Journal
1273   of Protozoology* 15:59-73. DOI: 10.1111/j.1550-7408.1968.tb02090.x
1274 Watson Kamm ME. 1922. Studies on gregaines II. Synopsis of the polycystid gregarines of the
1275   world, excluding those from the Myriapoda, Orthoptera, and Coleoptera. *Illinois
1276   Biological Monographs* 7:1-102
1277 Wolf M, Chen S, Song J, Ankenbrand M, Müller T. 2013. Compensatory base changes in ITS2
1278   secondary structures correlate with the biological species concept despite intragenomic
1279   variability in ITS2 sequences – a proof of concept. *PLoS ONE* 8:e66726. DOI:
1280   10.1371/journal.pone.0066726
1281 Žižka Z. 1978. Fine structure of the neogregarine *Farinocystis tribolii* Weiser, 1953. Syzygy and
1282   gametes formation. *Protistologica* 14:209-215
1283 Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic
1284   Acids Research* 31:3406-3415

1285 Table 1. Main characteristics of the sequences obtained in this study.

Sample name, obtained resulting sequence and its accession number	Characteristics of the PCR-amplified fragments or NGS-obtained contigs	Method of sequencing and PCR primers (if applicable): forward (F) and reverse (R)
Ancora sagittata from Roscoff 2009, contig of 4,853 bp long: part of SSU rDNA (1713 bp), complete ITS1, complete 5.8S rDNA, complete ITS2, and part of LSU rDNA (2767 bp); KX982501	(I) SSU rDNA (part); 1,567 bp	Sanger (direct sequencing of the PCR product) A[1] (F) 5'- GTATCTGGTTGATCCTGCCAGT -3' r71 (R) 5'- GCGACGGGCGGTGTGTAC -3'
	(II) SSU rDNA (part), ITS1, 5.8S rDNA, ITS2, and LSU rDNA (part); 941 bp	Sanger (after cloning) d6 (F) 5'- CCGTTCTTAGTTGGTGG -3' 28r3[2] (R) 5'-CCTTGGTCCGTGTTTCAAGAC-3'
	(III) LSU rDNA (part); 1,748 bp	Sanger (after cloning) 28d1[2] (F) 5'-ACCCGCTGAAYTTAAGCATAT-3' 28r7[2] (R) 5'- GCCAATCCTTWTCCCGAAGTTAC -3'
	(IV) LSU rDNA (part); 1,608 bp	Sanger (direct sequencing of the PCR product) 28d5[2] (F) 5'- CCGCTAAGGAGTGTGTAACAAC -3' 28r11[2] (R) 5'-GTCTAAACCCAGCTCACGTTCCCT-3'
Ancora sagittata from WSBS 2006, contig of 2,634 bp long: part of SSU rDNA (1696 bp), complete ITS1, complete 5.8S rDNA, complete ITS2, and part of LSU rDNA (first 564 bp); KX982502	(V) SSU rDNA (part); 1,663 bp	Sanger (direct sequencing of the PCR product) Q5A (F) 5'- GATTAAGCCATGCATGTCT -3' B[1] (R) 5'- GATCCTTCTGCAGGTTCACCTAC -3'
	(VI) SSU rDNA (part), ITS1, 5.8S rDNA, ITS2, and LSU rDNA (part); 1,096 bp	Sanger (after cloning) d71 (F) 5'-GTCCCTGCCCTTTGTACACACCGCCCG-3' 28r3[2] (R) 5'- CCTTGGTCCGTGTTTCAAGAC -3'
Ancora sagittata from WSBS 2010, contig of complete ribosomal operon (5,973 bp) KX982504	complete sequences of SSU rDNA (1737 bp), ITS1, 5.8S, rDNA, ITS2, LSU rDNA (3169 bp), and parts of ETSs	NGS (Illumina HiSeq 2000)
Ancora sagittata from WSBS 2011 contig of complete ribosomal operon (5,973 bp); KX982503	complete sequences of SSU rDNA (1737 bp), ITS1, 5.8S, rDNA, ITS2, LSU rDNA (3170 bp), and parts of ETSs	NGS (Illumina HiSeq 2000)
Stentor coeruleus, contig of 3,064 bp long: LSU rDNA, partial sequence; KX982500	LSU rDNA (part); 1,728 bp	Sanger (direct sequencing of the PCR product) 28d1[2] (F) 5'- ACCCGCTGAAYTTAAGCATAT -3' 28r7[2] (R) 5'- GCCAATCCTTWTCCCGAAGTTAC -3'
	LSU rDNA (part); 1958 bp	Sanger (after cloning) 28d5[2] (F) 5'- CCGCTAAGGAGTGTGTAACAAC -3' 28r13[2] (R) 5'- DYWRGCYGCGTTCTTCATCG -3'

1286 [1] The primer sequences were based on: Medlin et al., 1988.
1287 [2] The primer sequences were based on: Van der Auwera, Chapelle & De Wachter, 1994.

1288   Table 2. Comparison of the key features of gregarine attachment organelles.

	Mucron of archigregarines (*Selenidium*)	"Mucron" of aseptate eugregarines	Epimerite of septate eugregarines
Shape	Knob-like	Sucker-shaped or dome-shaped	Various, usually – a well-developed frontal protuberance of the cell of diverse shape
Tegument structure in the region of the junction with the host cell	The tegument of mucron is trimembrane pellicle excepting small region in front of conoid, where a cytostome and duct of mucron vacuole is intermittently formed, IMC is absent and there is just a single plasma membrane.	The IMC of the pellicle terminates at the edge of the cell junction zone, so the tegument of the attachment organelle is represented only by a single plasma membrane.	The IMC of the pellicle terminates at the edge of the cell junction zone, so the tegument of the attachment organelle is represented only by a single plasma membrane.
Cell junction between host and parasite	Septate cell junction; no peculiar structures on the edge of the junction zone.	Two closely adjacent plasma membranes (of host and parasite) forming high electron density zone. A circular groove in the gregarine tegument (plasma membrane) runs along the edge of the region of the cell junction (where the IMC terminates) and pinches a small portion of the host cell.	Two closely adjacent plasma membranes (of host and parasite) forming high electron density zone. A circular groove in the gregarine tegument (plasma membrane) runs along the edge of the region of the cell junction (where the IMC terminates) and pinches a small portion of the host cell.
Cytoplasm organelles	Apical complex (conoid, apical polar ring(s), rhoptries) and mucronal (food) vacuole.	Frontal region of "mucron" contains fibrillar zone or large vacuole with fibrillar content adjoining the cell junction zone; no mitochondria were observed; longitudinal actin-like fibrillar structures (filaments) are well developed.	Frontal region of epimerite contains a large flattened frontal vacuole with fibrillar content adjoining the cell junction zone; it is built from ER vesicles; mitochondria, often numerous and arranged in a layer, are located beneath this vacuole during growth and development of the trophozoite; different inclusions (lipid globules, amylopectin granules); longitudinal fibrillar structures (microfilaments and microtubules) can be well developed within the stalk of epimerite (if present).
Functioning	Attachment and feeding by myzocytosis: formation of temporary cytostome-cytopharingeal complex consisting of mucronal vacuole with the duct running through the conoid.	Attachment; feeding is questionable, no myzocytosis observed.	Attachment; feeding is questionable, no myzocytosis observed.
Fate	Mucron of archigregarines persists for long time after	Unknown; most likely is to retract/ condense: frontal part of	When trophozoite transforms into mature gamont, the epimerite is to

	trophozoite detachment and retains apical complex (conoid at least) till (including) stage of syzygy.	mature detached gamonts is covered by trimembrane pellicle, but not by a single plasma membrane.	break off or to retract/ condense.
Studied species	*Selenidium pendula, S. hollandei, S. orientale, S. pennatum*	*Lecudina* sp. from *Cirriformia tentaculata, ,L. pellucida, Lankesteria levinei, Difficilina cerebratuli*	*Didymophyes gigantea, Epicavus araeoceri, Gregarina* spp., *Leidyana ephestiae, Pyxinia firmus, Stylocephalus africanus.*
References	Schrével, 1968; Schrével, 1971a; Kuvardina & Simdyanov, 2002; Simdyanov & Kuvardina, 2007; Schrével et al., 2016	Schrével & Vivier, 1966; Ouassi & Porchet-Henneré, 1978; Simdyanov, 1995b; Simdyanov, 2009	Grassé, 1953; Devauchelle, 1968; Baudoin, 1969; Desportes, 1969; Ormières & Daumal, 1970; Hildebrand, 1976; Ormierès, 1977; Tronchin & Schrével, 1977; Marquès, 1979; Ghazali et al., 1989; Valigurová & Koudela, 2005; Valigurová et al., 2007; Valigurová, Michalková & Koudela, 2009; Schrével et al., 2013

1289

1290   Table 3. Characteristics of the main phylogenetic lineages of eugregarines[1].

Lineage and main representatives	Main characteristics	Hosts
Actinocephaloidea (short branch)	Morphologically diverse group, but well-supported with SSU rDNA phylogenies; possible morphological synapomorphy: biconical or bipyramidal oocysts[2]; also frontal syzygy are characteristic for the majority of the representatives.	Chiefly insects, but also earthworms.
Actinocephalidae[3]	Septate, typically with well-developed protruded epimerit often bearing hooks or other projections, or secondary aseptate (e.g., *Ascogregarina*, *Paraschneideria*); gliding motility and typical epicyte; epimerite discarding in mature gamonts; syzygy frontal; oocysts chiefly biconical or bipyramidal (sometimes spiny), sometimes crescent (e.g., *Menospora*).	Insects (intestine).
Monocystidae	Aberrant aseptate gregarines without pronounced epimerite (no valid TEM data); peristaltic motility (metaboly), aberrant epicyte (variously modified up to full loss); syzygy frontal or lateral; oocysts biconical.	Earthworms (seminal vesicles and coelom).
"Neogregarines"	Aseptate forms (sometimes intracellular) without pronounced epimerite; gliding motility and typical epicyte are absent in studied representatives; syzygy frontal (including intracellular species (Žižka, 1978)); oocysts biconical or bipyramidal (sometimes spiny).	Insects (intestine, Malpighian tubules, and fat body - for intracellular species).
Stylocephaloidea (short branch) Stylocephalidae	Septate gregarines likely related to Actinocephaloidea: trophozoite and syzygy morphology similar to the family Actinocephalidae, but epimerite is always elongate, without projections; oocysts purse-shaped.	Insects (intestine).
Gregarinoidea (long branch)	Chiefly septate (excepting *Caliculium glossobalani*). Possible synapomorphy: gametocysts with sporoducts (tubular projections for the releasing of oocysts); the other non-sequenced gregarines having them (e.g., *Gigaductus* having merogony like neogregarines (Ormierès, 1971)) are probably members of this lineage (Simdyanov, 2007; Schrével & Desportes, 2015).	Chiefly insects (intestine).
Gregarinidae[4]	Septate with bulbous epimerite retracted or condensed in mature gamonts, gliding motility and typical epicyte; early syzygy of caudo-frontal type; gametocysts with sporoducts, oocysts barrel-like	Insects (intestine).
Leidyanidae	Similar to Gregarinidae, but with late syzygy (just before gametocyst formation).	Insects (intestine).
Caliculium glossobalani	Weird marine aseptate gregarine superficially resembling *Selenidium*, but possessing neither bending motility nor the key ultrastructural features of archigregarines. Molecular data place it within Gregarinoidea.	*Glossobalanus minutus* (Hemichordata), intestine.
Cephaloidophoroidea (extremely long branch)	Septate and aseptate forms, intestinal parasites in crustaceans, robust clade in molecular phylogenetic trees with multiple distinct signatures in SSU rDNA sequences; no obvious morphological synapomorphies.	Crustaceans (intestine).
Cephaloidophoridae	Septate, with small epimerite (cephaloid) separated by septa persisting in mature gamonts, gliding motility and typical epicyte; syzygy caudo-frontal; oocysts ovoid or spherical with equatorial	Crustaceans (intestine).

suture or crest.

|		
Uradiophoridae	Septate, with small epimerite persisting in mature gamonts, gliding motility and typical epicyte; syzygy caudo-frontal; oocysts spherical with equatorial crest or radial projections.	Crustaceans (intestine).
Thiriotiidae	Aseptate, epimerite appears absent (no TEM data), gliding motility and typical epicyte; syzygy of unusual type (head-to-side); oocysts unknown.	Crustaceans (intestine).
Ganymedidae	Aseptate, epimerite appears absent (no TEM data), gliding motility and typical epicytic folds, syzygy caudo-frontal; oocysts unknown.	Crustaceans (intestine).
Lecudinoidea (long branch)	Chiefly aseptate forms without obvious morphological synapomorphies, but robust clade in molecular phylogenetic trees with nice multiple signatures in SSU rDNA sequences.	Broad range of various aquatic (chiefly marine) invertebrates.
Urosporidae[5]	Aseptate with weakly developed attachment apparatus (no TEM data); motility can be modified from gliding to peristaltic or loss of motility; epicyt from typical to aberrant; syzygy mainly lateral or frontal; oocysts heteropolar with funnel on the one pole and tail-like projection(s) on the other pole.	Chiefly coelom of echinoderms and polychaetes.
Ancoroidea (moderately long branch)	Robust clade in molecular phylogenetic trees (SSU rDNA); the external morphology of known representatives is similar to Lecudinidae; ultrastructure is understudied.	Capitellid polychaetes (intestine).
Ancoridae	Aseptate with two lateral projections and bulbous epimerite thought to be discarded in mature gamonts; gliding motility and typical epicyte, but apical filaments are probably modified; syzygy unknown; oocysts ovoid.	Capitellid polychaetes (intestine).
Polyplicariidae	Aseptate; attachment apparatus unknown; gliding motility and epicyte crests (no TEM data).	Capitellid polychaetes (intestine).
"*Ammonia*-like" environmental SSU rDNA sequences (moderate length branch)	Identified only with molecular data (SSU rDNA). Putative gregarines, expected to be aseptate, possibly a part of the current Lecudinidae.	Unknown.

1291   [1] Morphological characteristics were taken mainly from Grassé, 1953 and Perkins et al., 2000.
1292   [2] Oocysts = sporocysts or spores in Grassé, 1953; Schrével & Desportes, 2013a; Schrével et al.,
1293   2013.
1294   [3] *sensu lato*, i.e., including Sphaerocystidae and other related minor families separated by
1295   Levine.
1296   [4] *sensu lato*, including Blabericolidae Clopton, 2009.
1297   [5] *sensu lato*, including Gonosporidae Schrével and Desportes, 2013
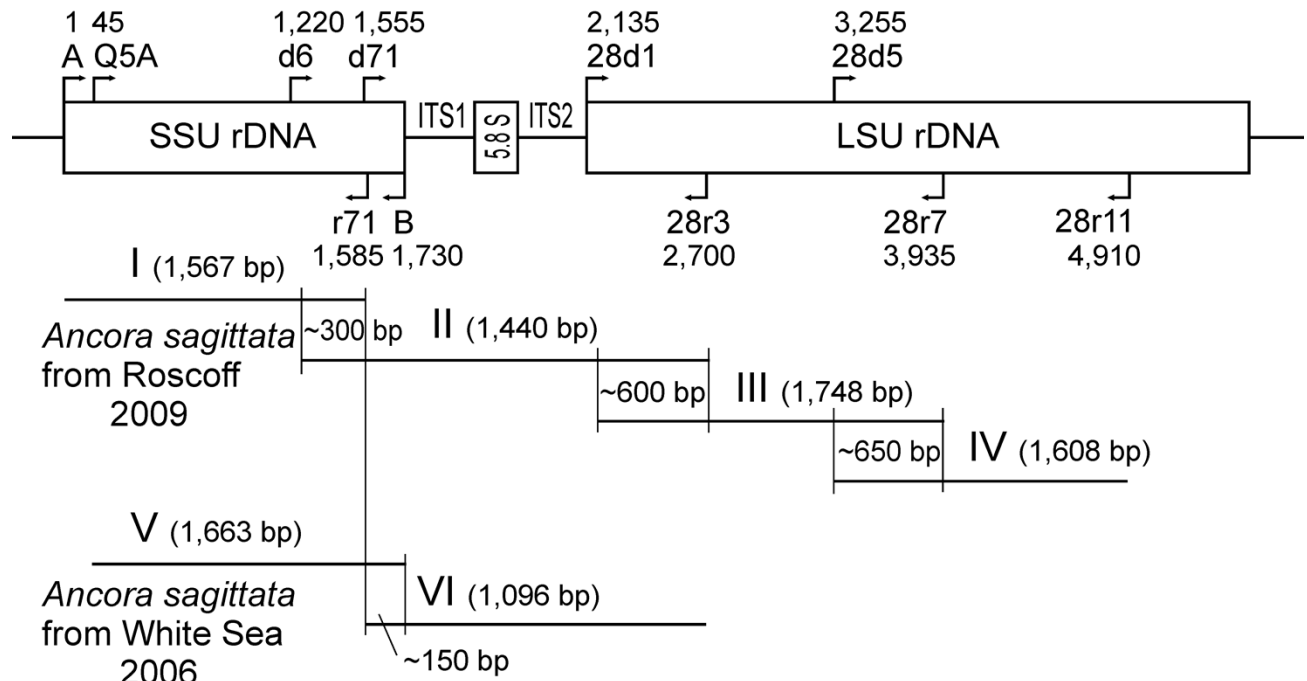
Figure 1: Strategy used to obtain contigs of the ribosomal operon from the samples of *A. sagittata* Roscoff 2009 and *A. sagittata* WSBS 2006.

Upper part: schematic ribosomal operon with approximate positions of the forward and reverse primers. Lower part: the amplified fragments of ribosomal DNA aligned with the ribosomal operon (above). Numbers indicate the length of the overlapping regions. Roman numerals denote the amplified fragments.
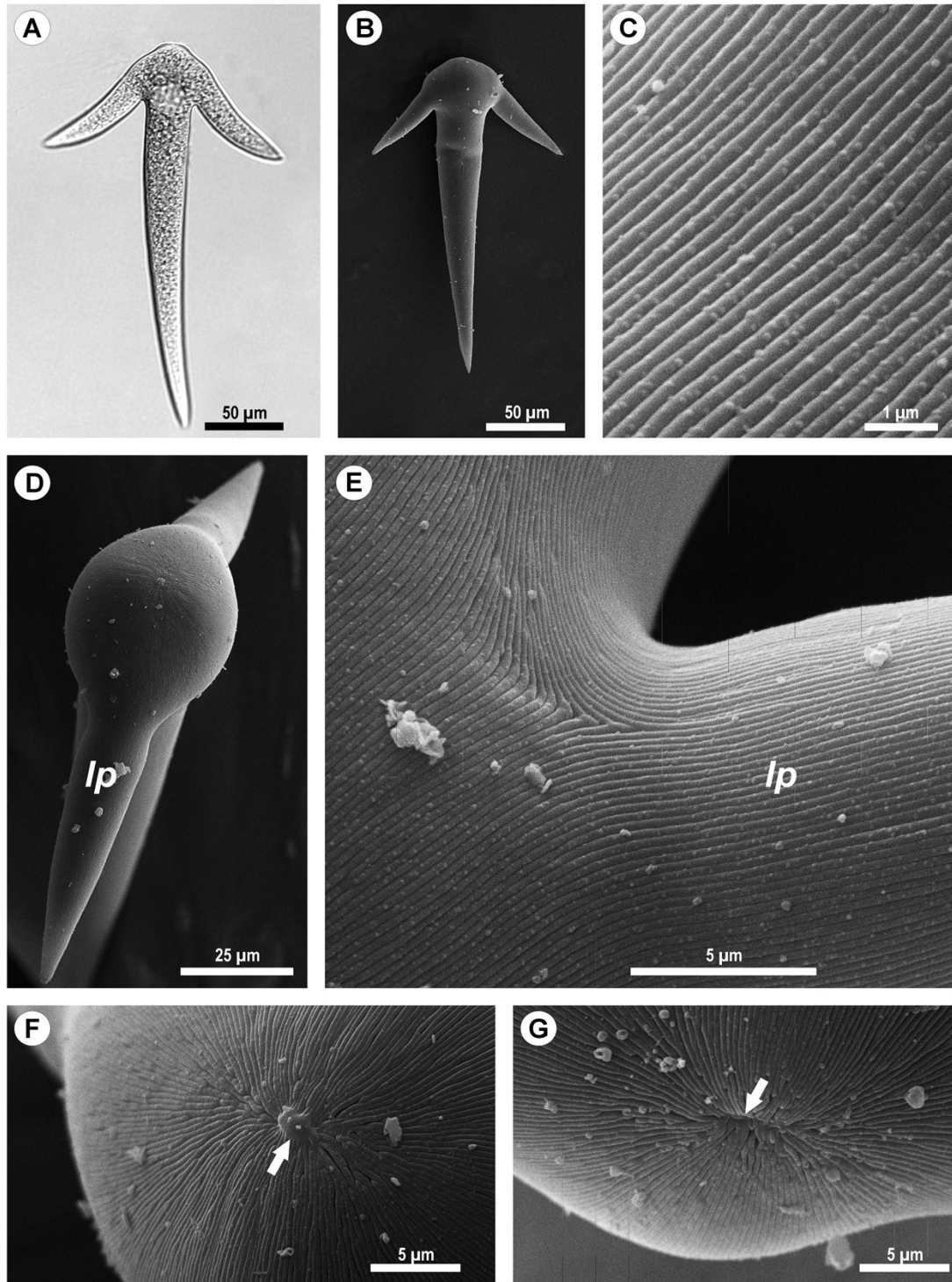
1306
1307 **Figure 2: Light (A) and scanning electron microscopy (B–G) of *Ancora sagittata*.**
1308 (A and B) General view of the gregarine; (C) Epicyte; (D) View of the gregarine from the apical
1309 pole of the cell; (E) Epicytic folds at the base of the lateral projections (*lp*); (F, G) Apical pole of
1310 *Ancora sagittata* (arrows) with (F) and without the apical papilla (G). *lp*, lateral projections of
1311 the cell.
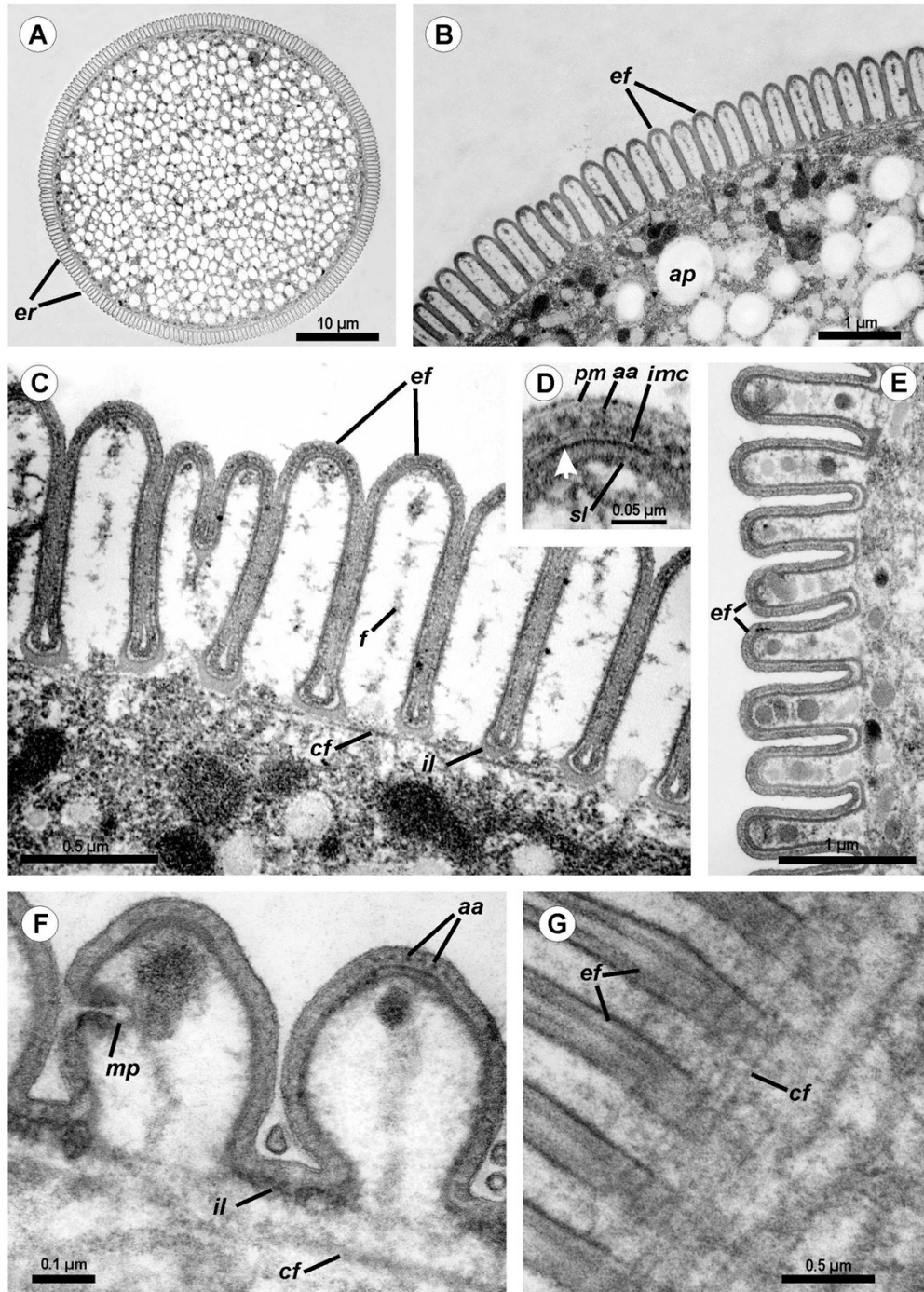
1312
1313 **Figure 3: Transmission electron microscopy of *Ancora sagittata*.**
1314 (A–C) Cross sections in the middle of the cell show epicytic folds (*ef*) with fibrils (*f*) inside,
1315 internal lamina (*il*), circular cortical filaments (*cf*), and granules of amylopectin (*ap*); (D) The top
1316 of the epicytic fold reveals a structure of the pellicle consisting of the plasma membrane (*pm*)
1317 and the internal membrane complex (*imc*) with rippled dense structures (= apical arcs, *aa*) and an
1318 electron dense plate (arrow); (E) Epicytic folds of the frontal zone of the cell with electron dense
1319 globules inside the folds; (F) Cross section at the level of the lateral projections: a micropore
1320 (*mp*) and circular cortical filaments (*cf*) are visible; (G) Tangential section of the cortex in the
1321 posterior region of the trophozoite reveals circular cortical filaments (*cf*).
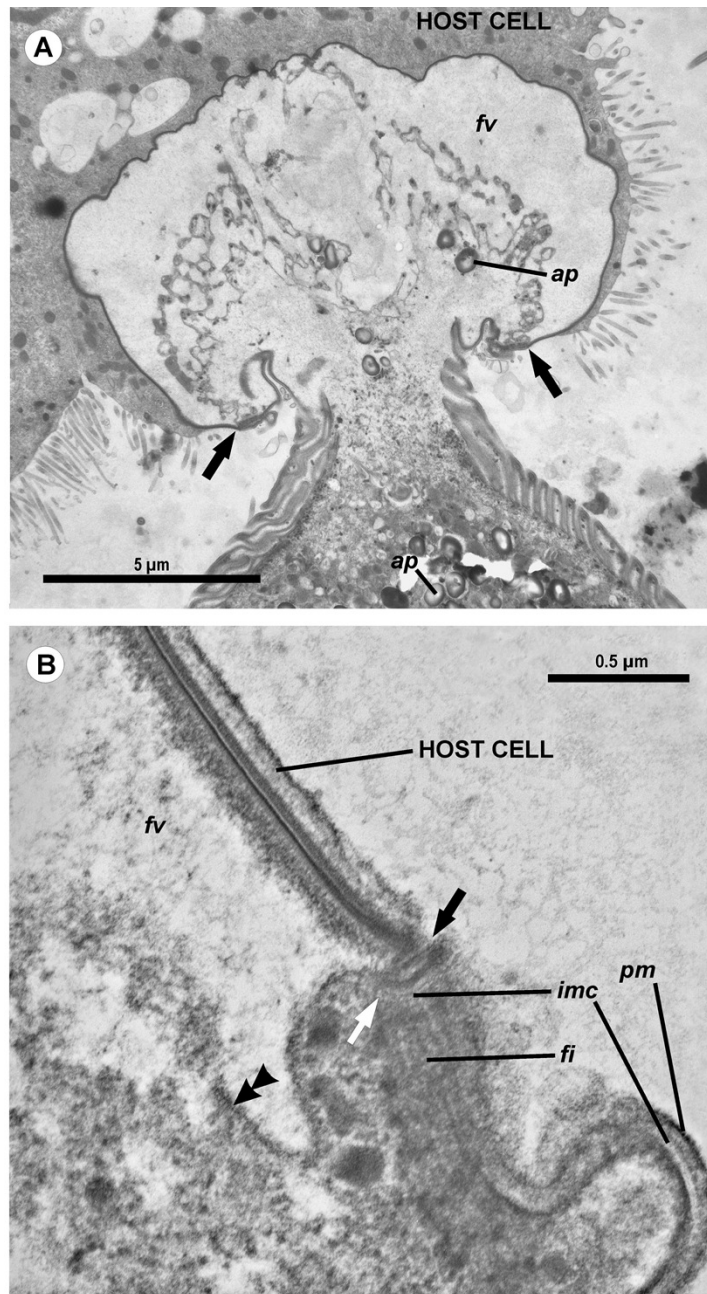
1322
1323
1324   **Figure 4: Transmission electron microscopy of the attachment apparatus of *Ancora***
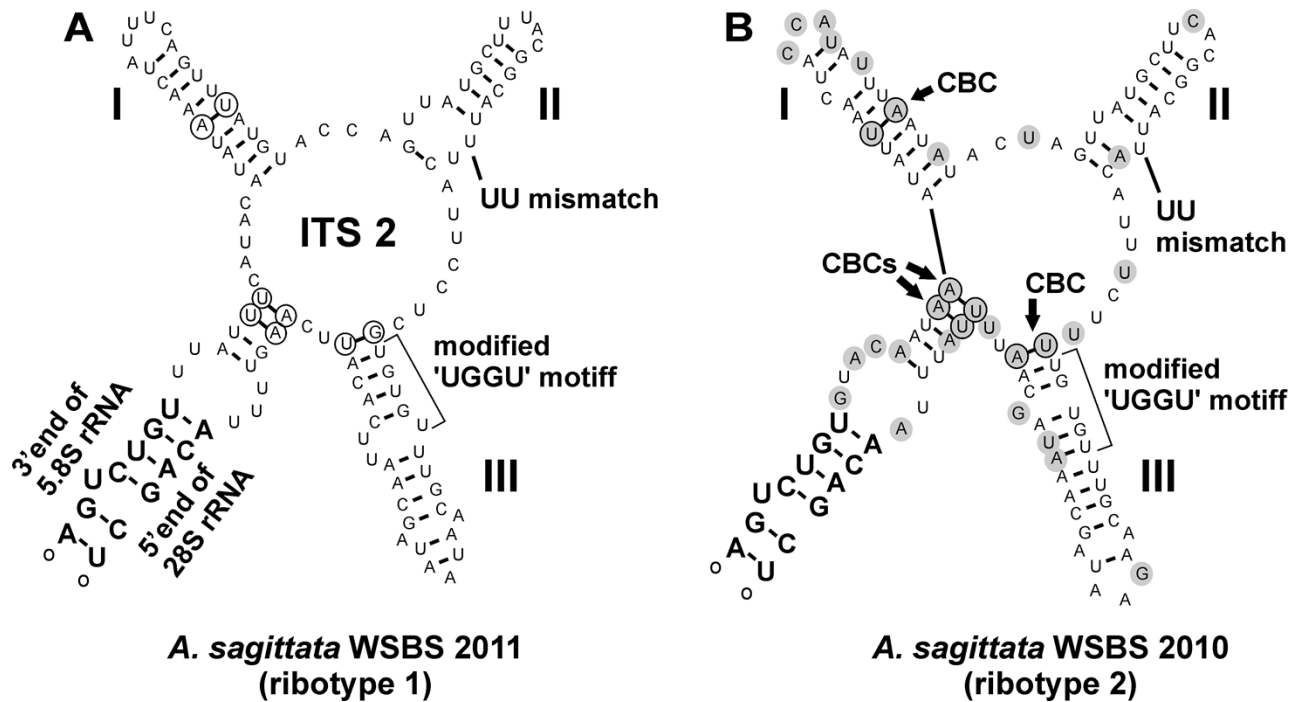1325   ***sagittata.***
1326   (A) Longitudinal section of the gregarine forebody embedded in a host cell shows a large frontal
1327   vacuole (*fv*) and amylopectin granules (*ap*) within the attachment organelle and the main part of
1328   the cell; the black arrows indicate the base of the contact zone (circular groove, see B); (B) The
1329   base of the contact zone between the gregarine and host cell under a higher magnification:
1330   gregarine cell forms a circular groove (black arrow) pinching the host cell; the rear wall of the
1331   frontal vacuole (double arrow) arises from this area; parallel filaments (*fi*) arise from the groove
1332   zone backward; the white arrow indicates the terminus of the internal membrane complex (*imc*)
1333   of the pellicle; *pm* is the plasma membrane of the gregarine cell.
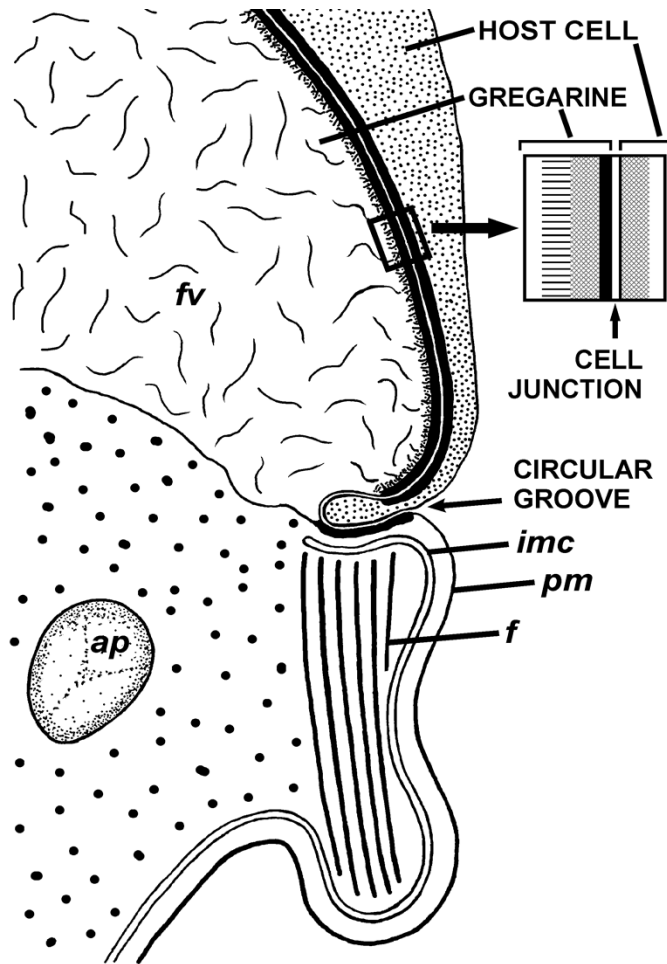
1334
1335
1336 **Figure 5: Diagram of the contact between the gregarine and the host cell as inferred from**
1337 **TEM micrographs.**
1338 Abbreviations are the same as in Figure 4.

Figure 6: Predicted secondary structures of ITS 2 transcripts of two *Ancora sagittata* ribotypes demonstrating differences between them.

(A) Ribotype 1; (B) Ribotype2. Nucleotide substitutions and insertions in the ribotype 2 are highlighted in grey. Nucleotides involved in compensatory base changes are encircled.
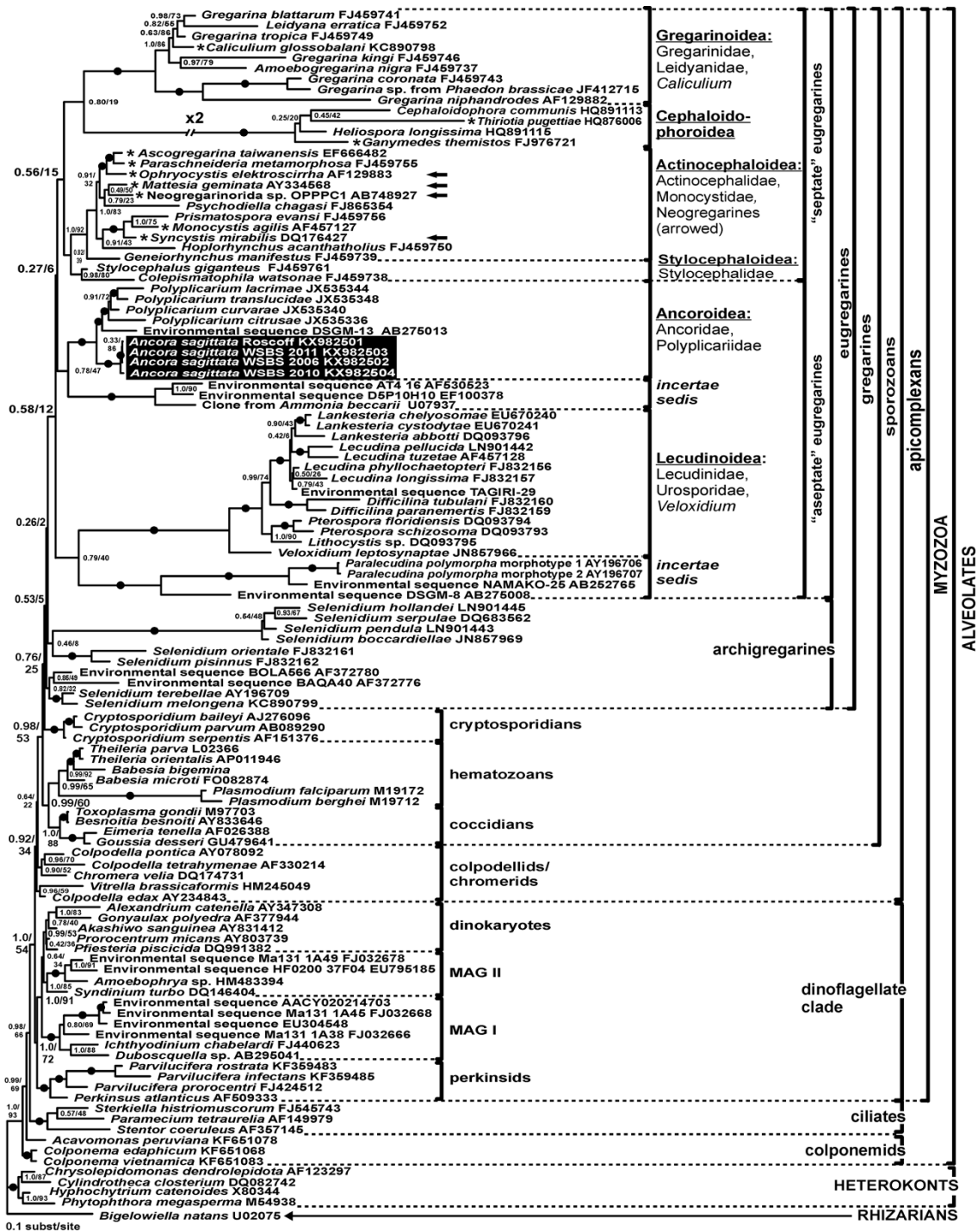
1346

Figure 7: Bayesian inference tree of alveolates calculated by using the GTR+Γ+I model from the dataset of 114 SSU rDNA sequences (1,570 sites).

1349 Numbers at the nodes indicate Bayesian posterior probabilities / ML bootstrap percentage. Black
1350 dots on the branches indicate Bayesian posterior probabilities and bootstrap percentages of at
1351 least 0.95 and 95%, respectively. The newly obtained sequences of *Ancora sagittata* are
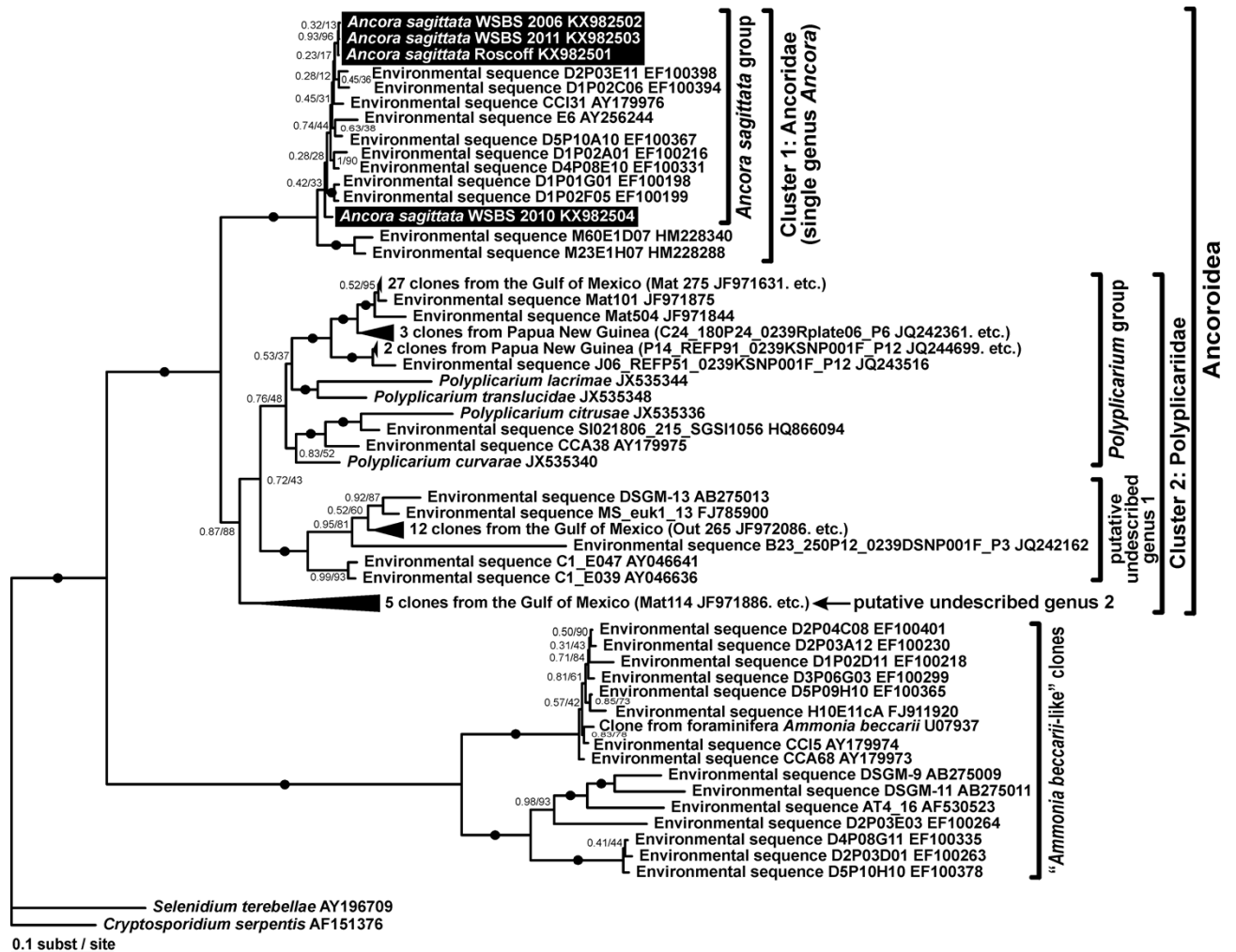1352 highlighted in by a black rectangle. Asterisks indicate aseptate gregarines within the "septate"
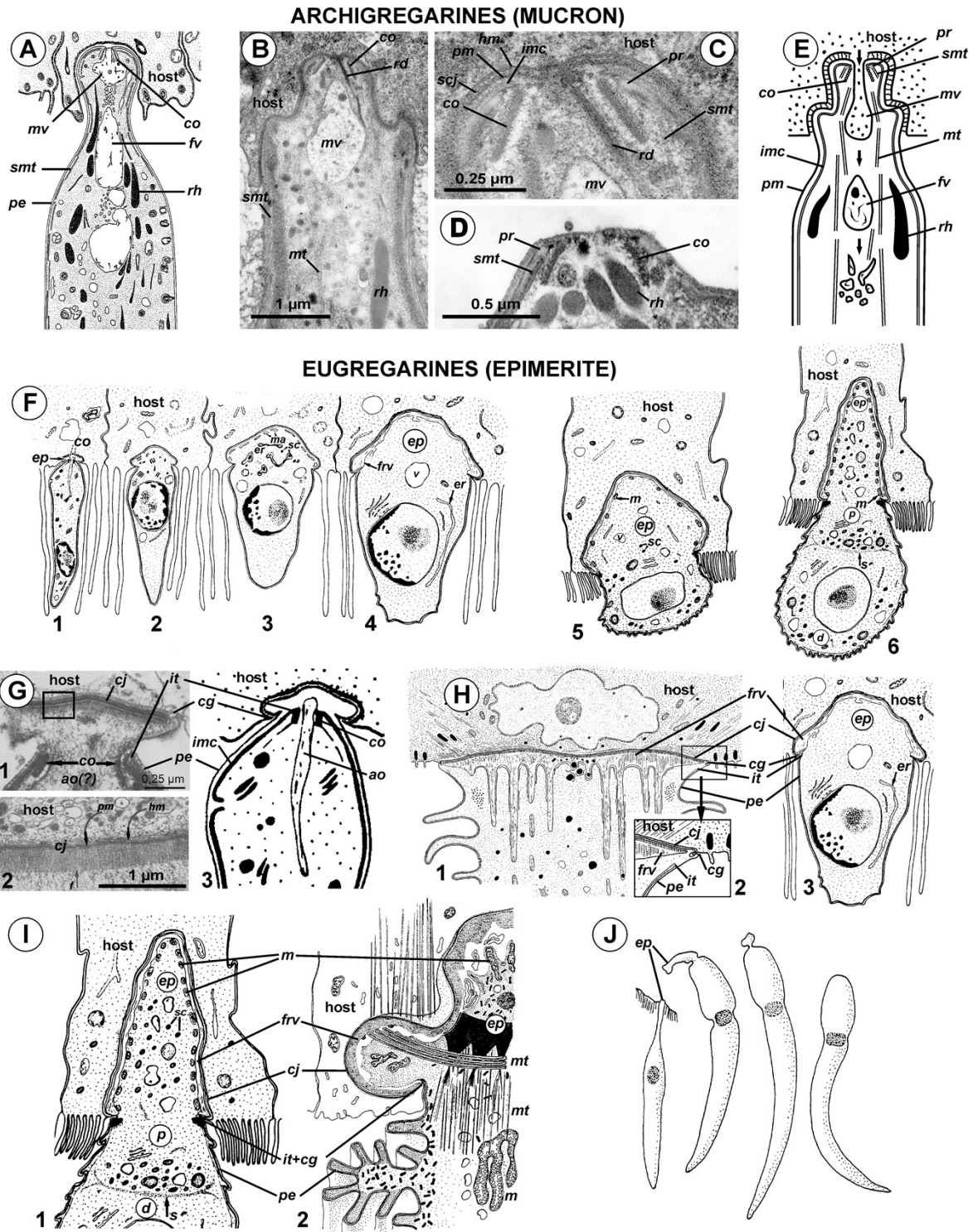1353 clade; arrows indicate neogregarines.

1354
1355
1356 **Figure 8: Bayesian inference tree of *Ancora sagittata* and related sequences obtained by**
1357 **using the GTR+Γ+I model from the dataset of 52 SSU rDNA sequences (1,709 sites).**
1358 Numbers at the nodes indicate Bayesian posterior probabilities / ML bootstrap percentage. Black
1359 dots on the branches indicate Bayesian posterior probabilities and bootstrap percentages of at
1360 least 0.95 and 95%, respectively. The newly obtained sequences of *Ancora sagittata* are
1361 highlighted by black rectangles. Black triangles indicate clusters of near-identical sequences
1362 (identity of 99% or more), each of which was represented by a single representative.

Figure 9: Bayesian inference trees of the alveolates obtained by using the GTR+Γ+I model and 50 sequences.

(A) LSU rDNA dataset (2,911 sites); (B) Ribosomal operon dataset (4,636 sites). Numbers at the nodes indicate Bayesian posterior probabilities / ML bootstrap percentages. Black dots on the branches indicate Bayesian posterior probabilities and bootstrap percentages of at least 95 and 95%, respectively. The newly obtained sequences of *Ancora sagittata* are highlighted by black rectangles. Accession numbers in (B) are arranged in following order: SSU rDNA, 5.8S (if available), LSU rDNA. The sequences of *Babesia bigemina* were obtained from the Sanger Institute genome project (www.sanger.ac.uk/Projects/B_bigemina/). Asterisks mark partial LSU rDNA sequences of small size (300-700 bp).

ARCHIGREGARINES (MUCRON)


EUGREGARINES (EPIMERITE)

1374
1377   (A) Drawing of the apical part of a *Selenidium hollandei* cell; (B) Ultrastructure of the apical part
1378   of an *S. orientale* cell, a longitudinal section; (C) The frontal region of the mucron under a higher
1379   magnification; (D) Mucron of the gamont (syzygy partner) of *S. pennatum*, a longitudinal
1380   section; (E) Predicted myzocytotic feeding in *Selenidium*; the mucron is embedded in the host

1381   cell and contains well-developed apical complex consisting of the conoid (*co*), polar ring (*pr*)
1382   giving rise to subpellicular microtubules (*smt*), rhoptries (*rh*) with rhoptry ducts (*rd*), and a large
1383   mucronal vacuole (*mv*); the tegument of the mucron comprises a trimembrane pellicle (*pe*)
1384   consisting of the plasma membrane (*pm*) and internal membrane complex, IMC (*imc*), with the
1385   exception of a small region in front of the conoid, a "cytostome site", where the IMC is absent
1386   and only single plasma membrane is present; the cytostome is intermittently opened in this
1387   region to myzocytosis: at first, food comes through the duct (temporary cytopharynx) in the
1388   newly formed mucronal vacuole (*mv*), which then becomes a food vacuole (*fv*) and is transported
1389   into the cell along microtubules (*mt*) for digestion; the parasite-host contact is mediated by the
1390   septate cell junction (*scj*) with a characteristic wide gap between the plasma membranes (*pm* and
1391   *hm*, respectively). (D) The mucron with the apical complex persists for a long time into the
1392   syzygy; the mucronal food vacuole is absent because the syzygy is a non-feeding stage.
1393   (F) Development of trophozoite of the septate gregarine *Gregarina blaberae* (scheme):
1394   (1), epimerite (*ep*) develops as a bulb in front of the apical complex consisting of the conoid and
1395   axial organelle (*ao*), which is likely a homologue of mucronal vacuole (also see (G, 3)); the IMC
1396   terminates near the apical part of conoid (similarly to mature *Selenidium*), therefore the
1397   developing epimerite is covered only by a single plasma membrane, not by the pellicle; (2–6),
1398   the apical complex disappears, the epimerite is growing; a large flattened frontal vacuole (*frv*)
1399   arising from the layer of membrane alveoli (*ma*) of endoplasmic-reticulum (*er*) origin, numerous
1400   mitochondria (*m*), granules of storage carbohydrate amylopectin (*sc*), lipid drops (*ld*), and
1401   vacuoles (*v*) are present in the epimerite cytoplasm; (6), finally, protomerite (*p*) and deutomerite
1402   (*d*) are separated by the septum (*s*). (G) Comparison of developing attachment organelles in the
1403   youngest trophozoites of the aseptate gregarine *Lecudina* sp. from the polychaete *Cirriformia*
1404   (Syn. *Audouinia*) *tentaculata*: ((1 and 2); (2) shows the details of the cell junction) and *G.*
1405   *blaberae* ((3), the magnified fragment of (F, 1)): both organelles develop ahead of the conoid in
1406   the same way and are covered by a single plasma membrane; the cell junction (*cj*) between the
1407   parasite and host cells is, unlike *Selenidium*, formed by two closely adjacent plasma membranes
1408   (parasite and host); an electron-dense fibrillar zone adjoins the cell junction in the gregarine cell
1409   (arrow); the cell junction is bordered by the circular groove (*cg*) pinching a small portion of the
1410   host cell; the IMC terminates (*it*) at the apical part of the conoid. (H) Comparison of the
1411   "mucron" of a well-developed trophozoite of the same *Lecudina* sp. ((1 and 2); the magnified
1412   fragment of (1) marked by the rectangle) and underdeveloped epimerite (*ep*) of a growing
1413   trophozoite of *G. blaberae* ((3), stage (4) from (F), magnified): the IMC terminates (*ie*) at the
1414   base of the attachment organelle (it marks the former apex of the sporozoite mucron), the cell
1415   junction consists of two closely adjacent plasma membranes bordered by the circular groove (*cg*)
1416   pinching a small part of the host cell, a large flattened frontal vacuole (*frv*) with fibrillar content
1417   develops just beneath the region of cell junction. (I) Comparison of the developing epimerite of
1418   an older trophozoite of *G. blaberae* ((1), stage (6) from (F), magnified) and the attachment
1419   organelle of *Lecudina* (Syn. *Cygnicollum*) *lankesteri* (2); (*m*), mitochondria. (J) A trophozoite
1420   and mature gamonts of *L. lankesteri*: losing of the epimerite.
1421   (A) is reprinted from: Schrével J. 1968. L'ultrastructure de la région antérieure de la Grégarine
1422   *Selenidium* et son interêt pour l'étude de la nutrition chez les Sporozoaires. *Journal de*
1423   *Microscopie, Paris* 7: 391-410 (© 1968 Société Française de Microscopie Electronique, Paris),
1424   with permission from the Journal de Microscopie et Biology Cellulaire published by Société
1425   Française de Microscopie Electronique, Paris (Apr 24, 2017); (B, C, and E) are reprinted from:
1426   Simdyanov TG & Kuvardina ON. 2007. Fine structure and putative feeding mechanism of the

1427   archigregarine *Selenidium orientale* (Apicomplexa: Gregarinomorpha). *European Journal of*
1428   *Protistology* 43:17-25 (© 2007 Elsevier), with permission from Elsevier (license number:
1429   4091531279186, Apr 17, 2017); (D) is reprinted from: Kuvardina ON & Simdyanov TG. 2002.
1430   Fine structure of syzygy in *Selenidium pennatum* (Sporozoa, Archigregarinida). *Protistology*
1431   2:169-177 (© 2002 by Russia, Protistology), with permission from the journal Protistology (Apr
1432   19, 2017); (F, G 3, H 3, and I1) are reprinted from: Tronchin G & Schrével J. 1977. Chronologie
1433   des modifications ultrastructurales au cours de la croissance de *Gregarina blaberae*. *Journal of*
1434   *Protozoology* 24:67-82 (© 1977 Society of Protozoologists, © John Wiley and Sons), with
1435   permission from John Wiley and Sons (license number: 4091540950763, Apr 17, 2017); (G 1, G
1436   2, and H 1) are reprinted from: Ouassi MA, Porchet-Henneré E. 1978. Étude ultrastructurale de
1437   mucron d'une Grégarine du genre *Lecudina*, parasite intestinal d'*Audoinia tentaculata* (Annélide
1438   Polychète) et de ses rapports avec la cellule hôte. *Protistologica* 14:39-52 (© 1978 Elsevier),
1439   with permission from Elsevier #RP016388; (I 2 and J) are reprinted from: Desportes I,
1440   Théodoridès J. 1986. *Cygnicollum lankesteri* n. sp., Grégarine (Apicomlexa, Lecudinidae)
1441   parasite des Annélides Polychètes *Laetmonice hystrix* et *L. producta*; particularités de l'appareil
1442   de fixation et implications taxonomiques. *Protistologica* 22:47-60 (© 1986 Elsevier), with
1443   permission from Elsevier #RP016388.
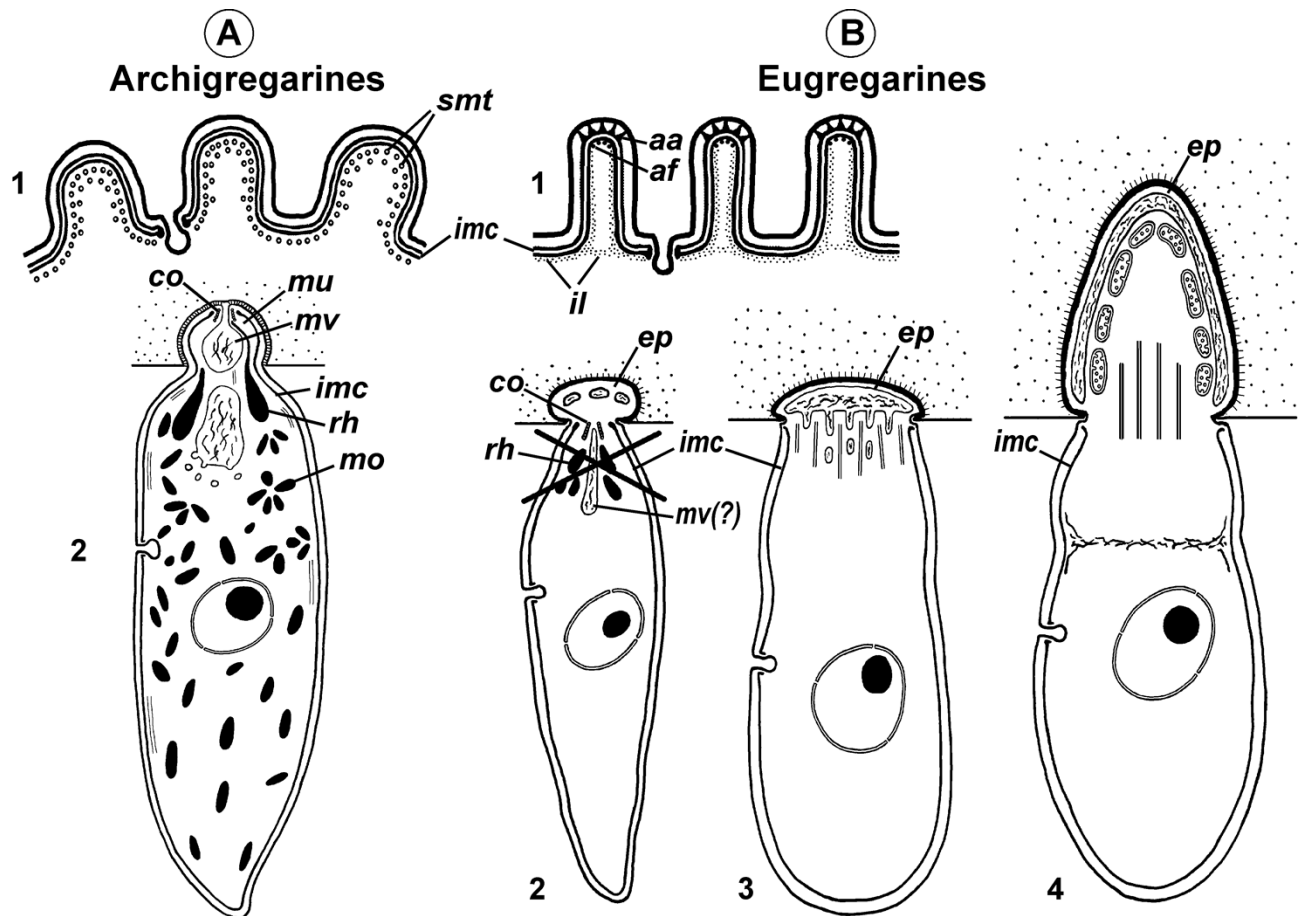
1444
1445
1446 **Figure 11: Comparison of archigregarine (A) and eugregarine (B) cell organization with**
1447 **their main diagnostic characteristics (candidate synapomorphies).**
1448 (A, 1 and B, 1) Cross sections of the cortex of a typical representatives showing regularly
1449 arranged longitudinal subpellicular microtubules (*smt*) in archigregarine longitudinal folds vs.
1450 ripple dense structures (apical arcs (*aa*)) and 12-nm filaments (apical filaments (*af*)) closely
1451 adjacent to the inner membrane complex of the pellicle (*imc*) within the tops of eugregarine
1452 epicytic crests; typically, internal lamina (*il*) forms links in the bases of the epicytic crests. (A, 2)
1453 Archigregarine trophozoite showing a mucron (*mu*) with an apical complex (conoid (*co*) and
1454 rhoptries (*rh*)) and mucronal food vacuole (*mv*) performing myzocytosis (the cell junction type
1455 between the host and parasite cells is septate junction); the cytoplasm is rich in microneme-like
1456 organelles (*mo*). (B, 2) Formation of the epimerite (*ep*) in eugregarines: a protuberance of the
1457 gregarine cell emerging ahead of the degrading apical complex. (B, 3) Epimerite (so-called
1458 "mucron") of some aseptate gregarines *Lecudina* spp. without the apical complex and with a
1459 large flat frontal vacuole and microtubules in the base. (B, 4) Epimerite of septate gregarines
1460 with the same structures and with mitochondria. In eugregarines, the cell junction between the
1461 host and parasite is formed by two closely adjacent plasma membranes and there is no
1462 myzocytosis (or perhaps only in the earliest developmental stages before the reduction of the
1463 apical complex).