



HAL
open science

Morphometrics and machine learning discrimination of the middle Eocene radiolarian species *Podocyrtis chalara*, *Podocyrtis goetheana* and their morphological intermediates

Francisco Pinto, Veronica Carlsson, Mathias Meunier, Bert Van Bocxlaer, Hammouda Elbez, Marie Cueille, Pierre Boulet, Taniel Danelian

► To cite this version:

Francisco Pinto, Veronica Carlsson, Mathias Meunier, Bert Van Bocxlaer, Hammouda Elbez, et al.. Morphometrics and machine learning discrimination of the middle Eocene radiolarian species *Podocyrtis chalara*, *Podocyrtis goetheana* and their morphological intermediates. *Marine Micropaleontology*, inPress, pp.102293. 10.1016/j.marmicro.2023.102293 . hal-04215322

HAL Id: hal-04215322

<https://hal.science/hal-04215322>

Submitted on 22 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Morphometrics and machine learning discrimination of the middle Eocene radiolarian**
2 **species *Podocyrtris chalara*, *Podocyrtris goetheana* and their morphological intermediates**

3

4 Francisco Pinto¹, Veronica Carlsson^{1, 2} *, Mathias Meunier¹, Bert Van Bocxlaer¹, Hammouda
5 Elbez², Marie Cueille¹, Pierre Boulet² and Taniel Danelian¹

6 ¹Univ. Lille, CNRS, UMR 8198, Evo-Eco-Paleo, F-59000 Lille, France.

7 ²Univ. Lille, CNRS, CRISAL – Centre de Recherche en Informatique Signal et Automatique
8 de Lille, UMR 9189, F-59000 Lille, France.

9

10 ***corresponding author: veronica.carlsson@univ-lille.fr**

11

12

13

14

15

16

17

18

19

20

21

22 **Abstract**

23 We present various approaches to distinguish the middle Eocene species *Podocyrtilis*
24 *chalara* and *Podocyrtilis goetheana*, which are end members of a trajectory of phenotypic
25 change, and their intermediate morphogroups. We constructed a set of thirteen traditional
26 morphological variables to classify the entire morphological variability encompassed by the
27 two morphospecies and their intermediates *Podocyrtilis* sp. cf. *P. chalara* and *Podocyrtilis* sp.
28 cf. *P. goetheana*. We used two methods of classification, namely Linear Discriminant
29 Analysis (LDA) and machine learning using artificial neural networks. LDA performed on the
30 morphometric data reveals a good discrimination for *P. chalara*, *P. goetheana* and *Podocyrtilis*
31 sp. cf. *P. goetheana*, but not for *Podocyrtilis* sp. cf. *P. chalara*. We used three approaches of
32 machine learning based on different neural networks: Convolutional Neural Networks
33 (CNNs) and two Spiking Neural Networks (SNNs). Each of these neural networks was trained
34 based on classified images of the two morphospecies and their morphological intermediates,
35 thus constituting a different set of input data than the morphometric dataset for LDA. The
36 neural network approaches identified the same three morphospecies recognized by LDA from
37 a dataset of traditional measurements, i.e. *P. chalara*, *P. goetheana* and *Podocyrtilis* sp. cf. *P.*
38 *goetheana*, with up to 92 % accuracy. Our results highlight the great potential and promising
39 perspectives of machine learning and neural networks in the application of image-based
40 object recognition for morphological classification, which may also contribute to more
41 objective taxonomic decisions.

42 **Keywords**

43 Morphometrics; Artificial Intelligence; Convolutional Neural Networks; Spiking Neural
44 Networks; Radiolarians; Automated identification

45

46 **1. Introduction**

47 Polycystine radiolaria are one of the oldest known Rhizarian lineages, with a fossil
48 record stretching back to the Early Cambrian (Obut and Iwata, 2000; Pouille et al., 2011) and
49 is thus of much interest for a number of evolutionary studies (e.g., Danelian and Johnson,
50 2001, Danelian et al., 2014, Renaudie and Lazarus, 2003, Tetard et al., 2017).

51 Since the early stages of the Deep Sea Drilling Program, the Cenozoic record of
52 polycystine radiolaria has allowed us to establish their evolutionary and biostratigraphic
53 significance, especially based on representatives of the Eocene genus *Podocyrtis*. (Sanfilippo
54 and Riedel, 1970; Riedel, 1971; Moore, 1972; Riedel and Sanfilippo, 1978). Calibrated
55 initially to the magnetostratigraphic time scale (Sanfilippo and Nigrini, 1998), the middle
56 Eocene tropical radiolarian zones are now tied to orbital chronology (Meunier and Danelian,
57 2022), which provides the highest resolution of temporal control possible today and allows
58 to define biostratigraphic events more accurately. Indeed, many of the middle Eocene
59 biozones are based on the evolution of the various lineages of the genus *Podocyrtis*, which
60 often relate to gradual anagenetic changes in phenotypes, as documented in several
61 evolutionary lineages since the early 1970s (Sanfilippo and Riedel 1970, 1992; Riedel, 1971).
62 For example, the bases of biozones RP14 and RP15 are defined based on anagenetic
63 phenotypic changes between the morphospecies *P. sinuosa* - *P. mitra* and *P. mitra* - *P.*
64 *chalara*, respectively.

65 The *Podocyrtis* (*Lampterium*) lineage ends with the marked morphological transition
66 of *Podocyrtis chalara* to *P. goetheana*. Interestingly, the first occurrence of *P. goetheana*
67 defines the base of biozone RP16 (Sanfilippo and Nigrini, 1998); an anagenetic transition was
68 reported, but intermediate forms were never documented in detail. The absence of such a
69 documentation has implications for our understanding of evolutionary changes in this
70 *Podocyrtis* lineage, but as intermediate forms are poorly understood it also affects the

71 recognition of the base of RP16. Here we examine the well-preserved radiolarian material of
72 ODP Sites 1259 and 1260 from Demerara Rise (equatorial Atlantic Ocean), which present an
73 exceptionally expanded Eocene sedimentary sequence. As such, this material offers an
74 exceptional opportunity to study the morphological transitional forms between *P. chalara* and
75 *P. goetheana*.

76 Within the abovementioned context, the principal aim of our study is to document
77 morphological variation in the anagenetic sequence of *P. chalara* to *P. goetheana* with two
78 different but complementary approaches to test the performance of various machine learning
79 algorithms based on neural networks. To reach this objective, we first quantified
80 morphological variation in the anagenetic transition between *P. chalara* and *P. goetheana*
81 with traditional morphometrics, i.e., linear measurements, pore counts and associated ratios in
82 the framework of qualitatively recognized morphological entities. This quantification of shape
83 follows previous attempts to assess morphological changes in the *Podocyrtes (Lampterium)*
84 lineage (Sanfilippo and Riedel, 1990; Rohlf and Bookstein, 1991; Danelian and Macleod,
85 2019; Watanabe et al., 2022). Using this morphometric framework of measurement data and *a*
86 *priori* morphospecies assignments, we examined how well linear discriminant analysis allows
87 to distinguish morphospecies as a baseline to test the performance of machine learning with
88 neural networks.

89 Testing the capabilities of neural network approaches based on image recognition is a
90 daunting task, because during the past couple of years, a variety of techniques involving
91 Artificial Neural Networks (ANNs) have been developed and improved. Convolutional
92 Neural Networks (CNNs) have been specifically designed for analysis of visual data, and are
93 now commonly used for image recognition, warranting detailed examinations of their
94 performance in morphological classification, and, therewith, as a tool to inform, and
95 potentially reach more objective taxonomic decisions. Indeed, CNNs are becoming well-

96 integrated in various micropaleontological studies for automatic image recognition (i.e., Mitra
97 et al., 2019, Hsiang et al., 2019; Marchant et al., 2020; Dollfus and Beaufort, 1999; Beaufort
98 and Dollfus, 2004; Bourel et al., 2020; Itaki et al., 2020; Renaudie et al., 2018; Tetard et al.,
99 2020). Regarding Eocene radiolaria, a recent study by Carlsson et al. (2022) applied a CNN to
100 eight well-delimited morphospecies of the genus *Podocyrthis*, and documented the potential of
101 this method under the simplified scenario when no morphological intermediates are present.
102 Spiking Neural Networks (SNNs) present another type of neural network, which in addition to
103 neuronal and synaptic states, they also incorporate a time component; this is why such
104 networks can more closely mimic natural neural networks (Maass, 1997). SNNs have wide
105 applicability, including modeling of natural systems such as the central nervous system of
106 biological organisms, as well as for image analysis. Traditionally, SNNs were less accurate
107 than other neural networks, but in recent years their performance has significantly improved;
108 they are more appropriate to process spatio-temporal data, and they may use computational
109 resources more effectively (Tavanaej et al., 2019). As such, an evaluation of SNNs in image
110 recognition and biological classification seems warranted.

111 In this paper, we expand on previous work by Carlsson et al (2022) with CNNs by
112 aiming to classify stacked and segmented images of the entire spectrum of morphological
113 variation found between *P. chalara* and *P. goetheana*. As mentioned, we used morphometrics
114 to document shape variability, which we subjected together with *a priori* morphospecies
115 assignments to LDA as a baseline to study the classification performance for image-based
116 neural networks using a CNN, a Spike-timing-dependent plasticity (STDP)-based SNN and a
117 SuperSpike-based SNN. This examination allows us to evaluate the use of imaging data and
118 neural networks for automated classification in a complex case study with intermediate
119 shapes. If the neural networks perform well, we would expect correct classification for each
120 morphogroup. Additionally, the results should reflect those of the LDA analysis, if the

121 morphometric documentation of shape variation is representative of the four morphospecies.
122 Alternatively, neural networks may show differences compared to LDA. These distinctions
123 could arise if neural networks fail to perform well, possibly due to unsuccessful training with
124 the existing data. In opposite to that, the neural networks may perform better than LDA if the
125 data supporting LDA lacks crucial shape information necessary for distinguishing between
126 morphogroups based on the images. As such, we expect our study to shed light into future
127 opportunities for automated biological classification of polycystine radiolaria and the use of
128 neural networks in developing more objective taxonomic decisions.

129

130 **2. Analyzed morphological groups**

131 Plate 1 displays the entire range of morphological variability observed between *P. chalara*
132 and *P. goetheana*. As linear discriminant analysis and the supervised learning of neural
133 networks are based on *a priori* group assignments, we were required to assign this continuum
134 of variation to a number of morphological groups. Based on extensive qualitative assessments
135 and to challenge the employed classification algorithms we recognized four distinct
136 morphological groups for the purpose of the current study, without currently being concerned
137 by the paleobiological/evolutionary status of each group. These morphogroups are briefly
138 presented below.

139

140 *Podocyrtes chalara* Riedel and Sanfilippo

141 [Pl. 1, fig. A, B](#)

142 1970 *Podocyrtes (Lampterium) chalara* Riedel and Sanfilippo, p. 535, pl. 12, figs. 2, 3.

143 1971 *Podocyrtes (Lampterium) chalara* Riedel and Sanfilippo: Moore, p. 743, pl. 3, figs. 5,

144 6.

- 145 1972 *Lampterium chalara* Riedel and Sanfilippo: Petrushevskaya and Kozlova, p. 543, pl.
146 32, fig. 12.
- 147 1978 *Podocyrtis (Lampterium) chalara* (Riedel and Sanfilippo): Riedel and Sanfilippo, p.
148 71, pl. 8, fig. 3, text-fig. 3.
- 149 2012 *Podocyrtis (Lampterium) chalara* Riedel and Sanfilippo: Kamikuri, p. 103, pl. 3, figs.
150 2a, 2b.
- 151 2012 *Podocyrtis (Lampterium) chalara* Riedel and Sanfilippo: Moore and Kamikuri, p. 9,
152 pl. P7, fig. 8.

153

154 Distinguishing characters: We include here forms displaying twelve or less vertically well-
155 aligned, subangular abdominal pores of similar size per horizontal row, illustrating the classic
156 morphology of *P. chalara*. Specimens of this morphogroup display less than thirteen pores on
157 the circumference of the abdomen.

158

159 *Podocyrtis* sp. cf. *P. chalara* Riedel and Sanfilippo

160 [Pl. 1](#), fig. C, D

161 1972 *Lampterium* sp. G: Petrushevskaya and Kozlova, pl. 32, fig. 10.

162 1972 *Lampterium* sp. aff. *L. goetheana*: Petrushevskaya and Kozlova, pl. 32, fig. 13.

163 2022 *Podocyrtis (Lampterium) chalara* Riedel and Sanfilippo: Meunier and Danelian, p. 21,
164 pl. 2.4.

165

166 Distinguishing characters: This morphogroup includes specimens that have a similar outline
167 and appearance as *P. chalara*, but they display vertically misaligned subangular abdominal
168 pores of different size. Specimens of this morphogroup may display vertical rows of pores
169 that are shifted to the right or left compared to the rows of pores above and below, giving a

170 twisted appearance for the rows of pores developed on the abdomen, with result a
171 honeycomb-like pore pattern. These shifts may be so extensive that the arrangement of pores
172 on the abdomen becomes chaotic, preventing the possibility to trace any apparent abdominal
173 rows or vertical alignment.

174

175 *Podocyrtis* sp. cf. *P. goetheana* (Haeckel)

176 [Pl. 1](#), figs. E - L

177 2006 *Podocyrtis (Lampterium) chalara* Riedel and Sanfilippo: Funakawa et al., p. 29, pl.
178 P9, figs. 11a, 11b.

179

180 Distinguishing characters: This morphogroup is mainly characterized by an increase in total
181 size, but with a significant reduction in the number of abdominal pores compared to both
182 variants of *P. chalara*. It differs from *P. goetheana* in that the bars of the second row of
183 abdominal pores are thicker and not always elongated, nor parallel to each other, as the
184 formation of the honeycomb-like pattern of pores becomes more apparent. This morphogroup
185 displays a high degree of morphological variability.

186

187 *Podocyrtis goetheana* (Haeckel)

188 [Pl. 1](#), figs. M - O

189 1887 *Cycladophora goetheana* Haeckel, p. 1376, pl. 65, fig. 5.

190 1970 *Podocyrtis (Lampterium) goetheana* (Haeckel): Riedel and Sanfilippo, p. 535.

191 1971 *Podocyrtis (Lampterium) goetheana* (Haeckel): Moore, p. 743, pl. 3, figs. 7, 8.

192 1972 *Lampterium* sp. aff. *L. goetheana* Petrushevskaya and Kozlova, pl. 32, fig. 14.

193 2005 *Podocyrtes (Lampterium) goetheana* (Haeckel): Nigrini et al., p. 45, pl. P5, figs. 11,
194 12.

195 2006 *Podocyrtes (Lampterium) goetheana* (Haeckel): Funakawa et al., p. 29, pl. P9, figs.
196 12a, 12b.

197 2012 *Podocyrtes (Lampterium) goetheana* (Haeckel): Kamikuri, p. 103, pl. 3, fig. 1.

198 2012 *Podocyrtes (Lampterium) goetheana* (Haeckel): Moore and Kamikuri, p. 9, pl. P7, fig.
199 9.

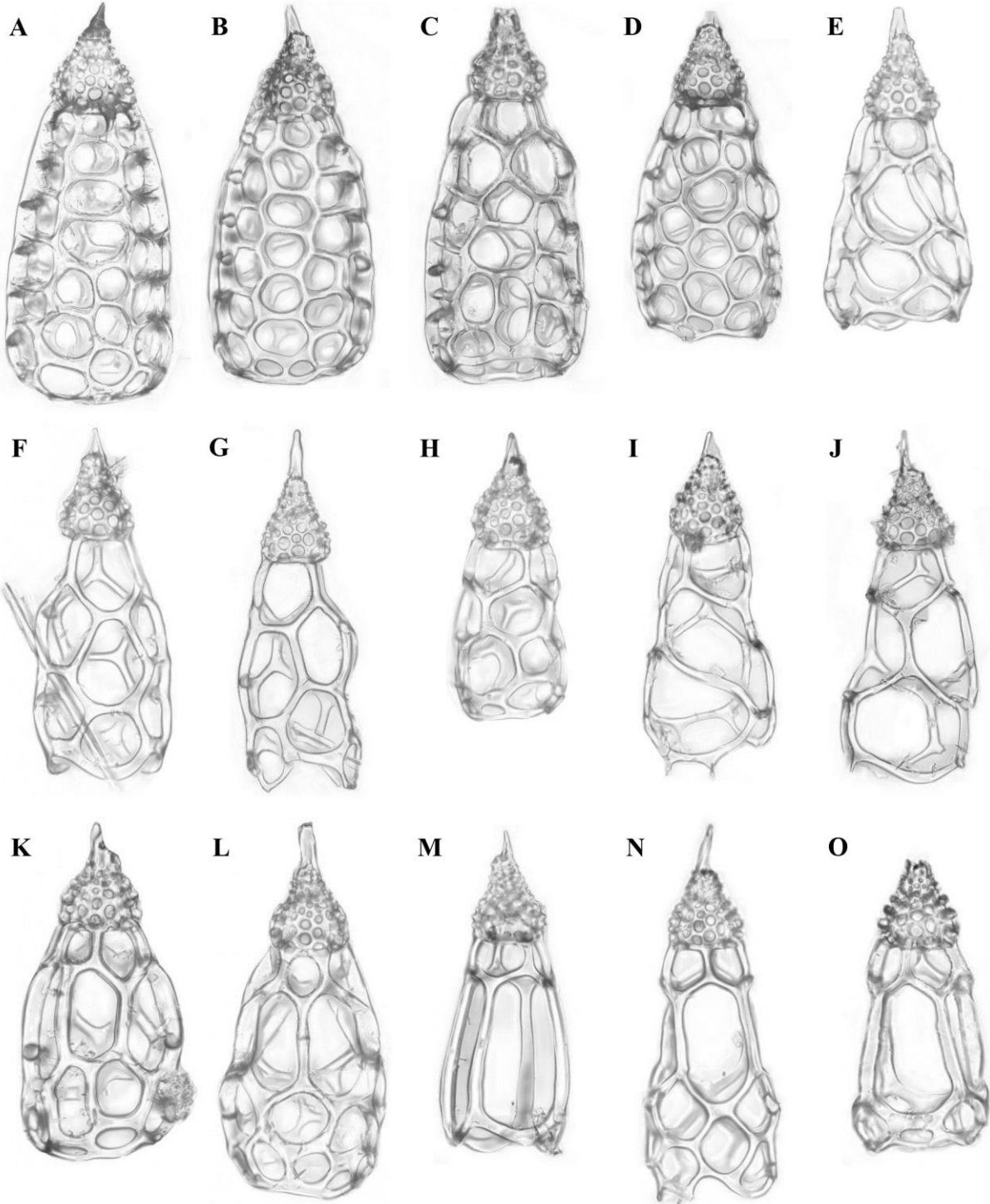
200 2022 *Podocyrtes (Lampterium) goetheana* (Haeckel): Meunier and Danelian, p. 21, pl. 2.5.

201

202 Distinguishing characters: This group includes only forms that display elongated straight bars
203 formed at the level of the second horizontal row of pores on the abdomen. This feature is
204 typical for *P. goetheana* as originally described, and, as mentioned above, the first occurrence
205 of typical *P. goetheana* defines the base of the RP16 Zone.

206

207 **Plate 1.** Composite light micrographs of *Podocyrtes* radiolaria from ODP Site 1260, processed and scaled in
208 ImageJ. (A) and (B) *Podocyrtes chalara*, samples: ODP 1260A-6R-5W, 63-65 cm and ODP 1260A-6R-5W, 20-
209 22 cm; (C) and (D) *Podocyrtes* sp. cf. *P. chalara*, samples: ODP 1260A-6R-4W, 68-70 cm and ODP 1260A-6R-
210 5W, 20-22 cm; (E) to (L) *Podocyrtes* sp. cf. *P. goetheana*, samples: ODP 1260A-6R-4W, 68-70 cm; ODP
211 1260A-6R-5W, 15-17 cm and ODP 1260A-6R-5W, 87-89 cm; (M) to (O) *Podocyrtes goetheana*, sample: ODP
212 1260A-6R-1W, 58-60 cm.



50µm

213

214

215 **3. Materials and methods**

216 **3.1. Sediment samples**

217 The material analyzed in this study consists of radiolaria obtained from an expanded
218 middle Eocene siliceous chalk sequence drilled at ODP Sites 1259 and 1260 (Leg 207,
219 Demerara Rise), located in the equatorial region of the Atlantic Ocean, 380 km offshore
220 Suriname (Erbacher et al., 2004; Danelian et al., 2005). The middle Eocene sequence is
221 particularly thick at Sites 1259 and 1260 and contains siliceous microfossils (radiolarians,
222 diatoms) of an excellent state of preservation (Danelian et al., 2007; Renaudie et al., 2010;
223 Meunier and Danelian, 2022, 2023). The part of the limestone sequence from Site 1260 that is
224 studied here is dated by orbito-chronology (Westerhold and Röhl, 2013). More specifically,
225 our samples span the interval between 41.24 Ma and 39.84 Ma. Site 1259 is dated via bio- and
226 magneto-stratigraphy and was sampled in the interval between ~39.05 and 37.70 Ma.

227

228 **3.2. Slide preparation**

229 A combined total of 15 samples from both sites were chosen and prepared for
230 microscopic observation using techniques described by Sanfilippo et al. (1985). A small
231 quantity (~2 cm³) of unprocessed sediment was collected from each sample and dried
232 overnight at 50°C to eliminate any residual water. After being weighed, sediment samples
233 were soaked for 2 hours in a 500 mL polypropylene beaker containing 30 mL of 30 %
234 hydrochloric acid (HCl), to dissolve their carbonate content and concentrate siliceous
235 microfossils. A few mL of HCl were added at the end to confirm the end of the reaction. The
236 residues resulting from the acid treatment was then washed by adding ~200 mL of distilled
237 water. After 2 hours of decantation, excess water was carefully removed using a pipette.
238 Residues were subsequently soaked for 2 hours in 30 mL of 10 % hydrogen peroxide (H₂O₂)

239 to remove organic matter, and subsequently washed through a 63 μm sieve using distilled
240 water. The $> 63 \mu\text{m}$ fraction was then exposed to ultrasonic waves for 10 min, then passed
241 again through the 63 μm sieve, and finally left to dry overnight at 50°C. For each sample, ~2
242 g of dried residue was carefully spread on top of a slide covered with several drops of Norland
243 Optical Adhesive 61, then topped with a coverslip and sealed by two minutes of exposure to
244 UV light.

245

246 **3.3. Microscopy and image processing**

247 The resulting slides were analyzed with a Zeiss AXIO Images A2 microscope under
248 transmitted light at $\times 10$ and $\times 20$ magnifications. All specimens recognized qualitatively as *P.*
249 *chalara*, *Podocyrtilis* sp. cf. *P. chalara*, *Podocyrtilis* sp. cf. *P. goetheana* or *P. goetheana* were
250 manually photographed using the mounted Axiocam ERc5s with Zen 3.5 (blue edition)
251 software. For each specimen, a batch of 5-10 photographs were taken at different focal points
252 to obtain a series of images, which were stacked afterwards using Helicon Focus 7.7.0
253 (HeliconSoft) to create a composite picture entirely in focus. The stacked images were
254 subsequently retouched with Paint3D to facilitate their automated segmentation. This last
255 procedure was performed with the ImageJ BioVoxxel plugin (Brocher, 2022) and the
256 AutoRadio_Segmenter plugin developed for ImageJ / Fiji (Tetard et al., 2020).

257

258 **3.4. Morphometric analyses**

259 As we aimed to subject specimens to discriminant analysis based on morphometric
260 measurements, it was essential that the documented morphological variables would
261 adequately capture shape variations present in between the four morphogroups. We first
262 designed a set of variables that would allow to compare with the variables used by Watanabe

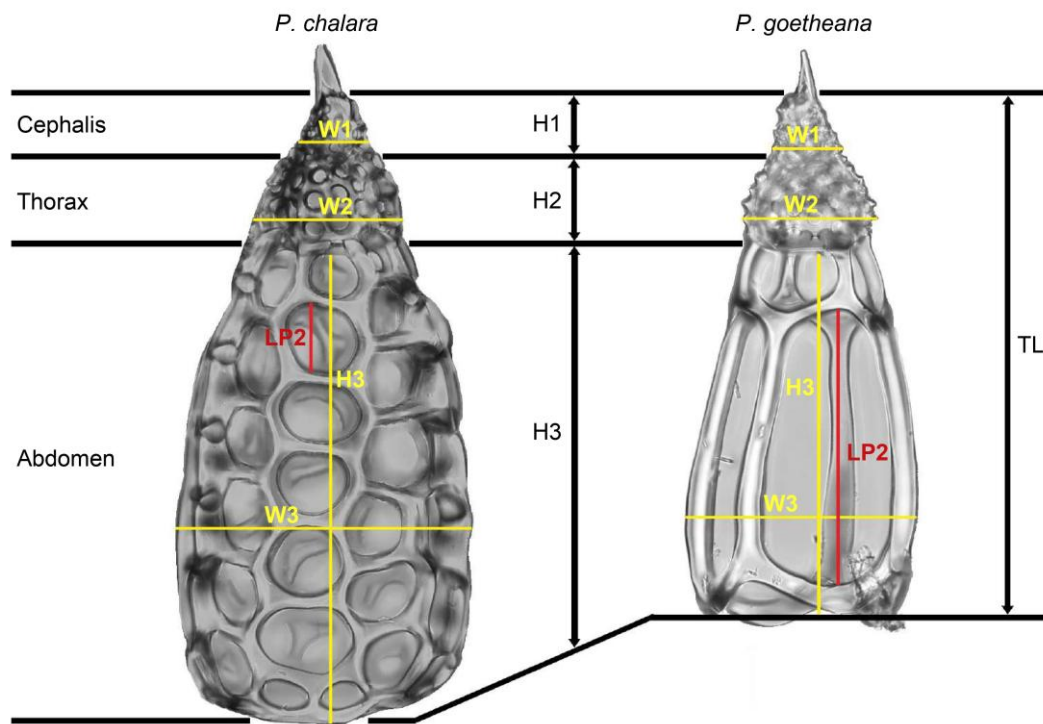
263 et al. (2022) on the specimens of the *Lampterium* lineage from the Pacific Ocean. A subset of
264 seven of these variables was retained and supplemented with six newly proposed variables to
265 result in a set of thirteen morphological variables that document well the morphological
266 variation between the four morphogroups (Figure 1). The seven morphological variables
267 proposed by Watanabe et al. (2022) are:

- 268 - W1: Maximal width of the cephalis
- 269 - W2: Maximal width of the thorax
- 270 - W3: Maximal width of the abdomen
- 271 - H1: Maximal height of the cephalis
- 272 - H2: Maximal height of the thorax
- 273 - H3: Maximal height of the abdomen
- 274 - TL: Total length or height of the specimen

275 and our six additional variables are:

- 276 - LP2: Maximum length of the second abdominal pore along along the axis used to
277 measure H3 (with the first pore being the one closest to the thorax-abdomen border)
- 278 - NPV: Number of abdominal pores aligned vertically along the axis used to measure
279 H3 (on the front-facing side of the skeleton)
- 280 - NPH: Number of abdominal pores aligned horizontally behind the axis of W3 (on the
281 front-facing side of the skeleton)
- 282 - R1: Maximum length of the second pore of the abdomen / H3
- 283 - R2: Number of abdominal pores aligned vertically behind the axis of H3 / H3
- 284 - R3: Number of abdominal pores aligned horizontally behind the axis of W3 / W3

285



286

287 **Figure 1.** Schematic representation of the eight skeletal variables measured for all specimens outlined in Table
 288 1. Five additional variables related to pore counts and ratios are not illustrated. Abbreviations: W1: maximum
 289 width of the cephalis; W2: maximum width of the thorax; W3: maximum width of the abdomen; H1: maximum
 290 height of the cephalis without the apical horn; H2: maximum height of the thorax; H3: maximum height of the
 291 abdomen; TL: total length or height of the specimen without the apical horn; LP2: maximum length of the
 292 second abdominal pore.

293

294 Quantifications of these variables (**Table S1** in **Supplementary materials**) were directly
 295 performed on a dataset of 214 photographs/specimens from samples 1260A-6R-1W, 58-60
 296 cm, 1260A-6R-4W, 68-70 cm, 1260A-6R-5W, 15-17 cm and 1260A-6R-5W, 87-89 cm,
 297 outlined in **Table 1**, using the image processing and analysis software ImageJ (Schneider et
 298 al., 2012). All of these data were tabulated and then imported in the statistical environment R
 299 (v. 4.1.3; R Core Team 2022) for subsequent LDA using the packages MASS (v. 7.3-60;
 300 Ripley et al., 2013) and vegan (v. 2.6-4; Oksanen et al., 2013). This analysis is a constrained

301 ordination procedure that uses a linear combination of coefficients to maximize the distance
 302 between *a priori* defined groups, while minimizing the distance within each group (Venables
 303 and Ripley, 2002). As morphometric variables were expressed in both metric units and as
 304 ratios, the LDA was performed on data that were transformed to have the mean at zero- and
 305 one-unit standard deviation (z transformation). After subjecting the whole dataset to LDA we
 306 performed cross-validation by 100 replicates of randomly assigning 80 % of the data to a
 307 training dataset and the remainder to a testing set to evaluate the classification success of the
 308 LDA.

309

310 **Table 1.** Number of specimens analyzed per morphogroup and per sampled core interval.

Samples	<i>P. chalara</i>	<i>Podocyrtris</i> sp. cf. <i>P. chalara</i>	<i>Podocyrtris</i> sp. cf. <i>P. goetheana</i>	<i>P. goetheana</i>	Total number of measured specimens per sample
1260A 6R 1W 58-60cm	0	0	2	29	31
1260A 6R 4W 68-70cm	18	19	37	0	74
1260A 6R 5W 15-17cm	19	12	7	0	38
1260A 6R 5W 87-89 cm	34	9	28	0	71

311

312

313 **3.55. Artificial Neural Networks**

314 Artificial Neural Networks or simply neural networks are machine learning algorithms
 315 designed to simulate the decision-making processes of the human brain by analyzing and
 316 exploiting patterns in data (Yang and Yang, 2014). Prior to analysis, the data given to a
 317 network is split into two parts, one for training and one for testing, usually in a 80 to 20 ratio.
 318 The first set of data is used to train the neural network, so that to enable it to learn recognizing

319 features and patterns present in the data, whereas the second set of data is used to test the
320 performance of the network to classify cases based on the previously trained capabilities of
321 recognition.

322 A type of ANNs that has been specifically designed to analyze visual data are CNNs,
323 commonly used for image recognition. They are designed to analyze visual data by
324 considering the color values of each pixel and by identifying patterns within images (Hijazi et
325 al., 2015). CNNs utilize a process known as convolution. Convolution can be described as a
326 linear operation to decompose the input image by sliding small windows known as filters or
327 kernels over the input image to construct layers that each obtain certain features. The
328 convolutional layers in a CNN modify gradually the image parameters, such as weights or
329 bias, to learn and recognize specific patterns or objects in the images. By adjusting these
330 parameters through training, the network aims to correctly classify the output given a
331 particular input. When a CNN has multiple layers, typically more than three, the procedure is
332 referred to as deep learning, as each layer enables the recognition of more and more advanced
333 features in an image. As mentioned, SNNs consider additionally the time factor, alike
334 biological neurons, which use discrete spikes to compute and transmit information, instead of
335 characterizing neurons by a single, static continuous-valued activation.

336 The hyperparameters (i.e. weights) of the neural network analyzed in this study were
337 chosen randomly; more specifically a value between -1 and 1 was chosen for the VGG16 and
338 SuperSpike-based networks, while a value network between 0 and 1 was chosen for the
339 STDP-based network neural. We analyzed two sets of stacked and segmented images. The
340 first set contained images with *a priori* assignments to the four morphogroups represented in
341 section 2; for the second set assignments were altered based on the results of LDA. In each
342 case, we performed ten runs per type of neural network used, with 20 epochs for each
343 network, except for the STDP-based network that was run with 100 epochs. An epoch simply

344 means how many passes it goes through the training set and updates parameters based on each
345 pass. The following neural networks were used to perform runs:

346 - Visual Geometry Group 16 (VGG16). This 16-layer deep CNN was used for its
347 simplicity. For our analyses, we use transfer learning, meaning that the first 15 layers
348 were already pre-trained based on a large-scale image dataset from ImageNet, and we
349 only trained the last layer specifically using our data and PyTorch (Paszke et al.,
350 2019). More information about VGG16 is provided by Simonyan and Zisserman
351 (2015).

352 - STDP-based Spiking Neural Network (STDP-Network). This network contains
353 convolutional and pooling layers that learn the features from the data using a Spike
354 Timing Dependent Plasticity (STDP), a learning algorithm inspired by natural
355 neurons. STDP adapts the synaptic connections between the neurons based on the
356 timing of the spikes to transmit information (Masquelier and Thorpe, 2007) in an
357 unsupervised way, i.e., without *a priori* group assignments. This SNN is then
358 combined with a Support Vector Machine (SVM) for classification in the STDP-based
359 network using the *a priori* group assignments (Cortes and Vapnik, 1995). To train the
360 STDP-Network, we used the CSNN-simulator (Falez, 2019).

361 - SuperSpike-based Spiking Neural Network (SuperSpike-Network). This SNN is
362 trained using a nonlinear voltage-based three-factor learning rule capable of training
363 multilayer networks called the SuperSpike (Zenke and Ganguli, 2018), which is a
364 supervised global learning rule similar to deep learning. We used the Norse simulator
365 for our analyses (Pehle and Pedersen, 2021).

366

367 All the neural network simulations were conducted on the cluster “grouille” of the Grid’5000
368 test bed (Balouek et al., 2013) using two Nvidia A100-PCIE-40GB GPUs, an AMD EPYC

369 7452 32 core CPU (Zen 2, 2 CPUs/node), and 128GB of RAM. For each type of neural
 370 network, we averaged the assignment accuracies obtained over the ten replicate runs to gain
 371 robust insight into performance.

372 The training of the neural networks are expected to be better when a large, data-rich training
 373 set is used. Because our two analyzed datasets (Tables 2 and 3) are composed of 428 and 514
 374 original images, respectively, we considered it necessary to augment the data available for
 375 ANN training. Therefore, we performed the following data augmentation procedures:

- 376 — Rotate the images by a randomized angle between -15 and 15 degrees and keep all copies.
- 377 — Randomly choose images that would be flipped from left to right and keep both copies.
- 378 — Rescale random images with values between 1 and 1.3 (with 1 being the default scale
 379 value).

380 The total number of images in each dataset was enhanced to >1000 images via data
 381 augmentation.

382

Table 2. List of images included (prior to augmentation) in each of the classes for the four- class dataset used for analysis withVGG16with.

Samples	<i>P. chalara</i>	<i>Podocyrtris</i> sp. cf. <i>P. chalara</i>	<i>Podocyrtris</i> sp. cf. <i>P. goetheana</i>	<i>P. goetheana</i>	Total number of analyzed specimens per sample
1260A 6R 4W 68-70cm	0	19	34	3	56
1260A 6R 4W 119-121cm	0	13	2	1	16
1260A 6R 5W 15-17cm	0	12	7	0	19
1260A 6R 5W 63-65cm	22	9	15	0	46
1260A 6R 5W 87-89 cm	33	12	18	14	77
1260A 6R 6W 20-22 cm	12	18	9	0	39
1260A 6R 6W 57-59 cm	19	14	2	0	35
1260A 7R 1W 22-24 cm	24	15	0	0	39
1260A 7R 1W 69-71 cm	33	11	0	0	44

1260A 7R 1W 121-123 cm	31	11	0	0	42
1260A 7R 2W 19-21 cm	0	13	0	0	13
1260A 8R 3W 65-67cm	0	1	0	0	1
TOTAL	174	148	87	18	427

Table 3. List of images included (prior to augmentation) in each of the classes for the three class three dataset used for analysis with VGG16, STDP- based SNN and SuperSpike-based SNN.

Samples	<i>P. chalara</i> + <i>Podocyrtis</i> sp. cf. <i>P. chalara</i>	<i>Podocyrtis</i> sp. cf. <i>P. goetheana</i>	<i>P. goetheana</i>	Total number of analyzed specimens per sample
1259A 17R 1W 54-56cm	0	0	1	1
1259A 18R 1W 53-55cm	0	0	41	41
1259A 18R 2W 53-55cm	0	0	36	36
1260A 6R 1W 58-60cm	0	1	26	27
1260A 6R 4W 68-70cm	19	29	0	48
1260A 6R 4W 119- 121cm	13	2	1	16
1260A 6R 5W 15-17cm	12	5	0	17
1260A 6R 5W 63-65cm	31	15	0	46
1260A 6R 5W 87-89cm	43	28	0	71
1260A 6R 6W 20-22cm	29	8	0	37
1260A 6R 6W 57-59cm	34	2	0	36
1260A 7R 1W 22-24cm	39	0	0	39
1260A 7R 1W 69-71cm	44	0	0	44
1260A 7R 1W 121- 123cm	42	0	0	42
1260A 7R 2W 19-21cm	13	0	0	13
TOTAL	319	90	105	514

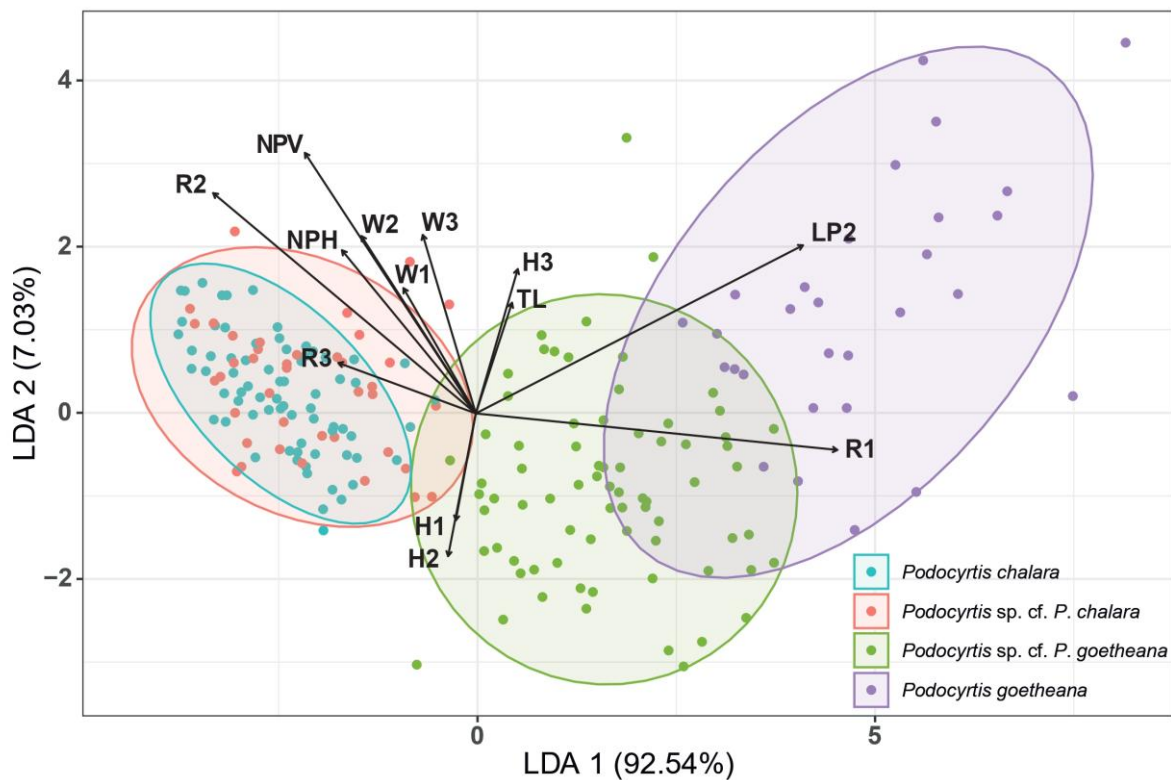
383

384 4. Results

385 4.1. Morphometrics and linear discriminant analysis

386 The LDA performed on the matrix of our 13 morphometric variables, i.e., measurements, pore
387 counts and ratios, represented >99 % of the variation on the first two axes and clustered
388 *Podocyrtis goetheana* and *Podocyrtis* sp. cf. *P. goetheana* successfully (Fig. 2).
389 Comparatively, specimens belonging to *P. chalara* were regularly confused with *Podocyrtis*

390 sp. cf. *P. chalara* and vice versa, resulting overall in 73.5 ± 6.1 % (mean \pm sd) correct
 391 assignments (Table 4). These two latter morphogroups overlap completely on the LDA plot,
 392 whereas *P. Podocyrtis* sp. cf. *P. goetheana* and *P. goetheana* are relatively well-separated
 393 from *P. chalara* and *Podocyrtis* sp. cf. *P. chalara*, although they share limited overlap with
 394 each other (Figure 2). The first axis of the LDA mainly represents size variations of the
 395 second row of abdominal pores. The morphological variables that contribute the most to
 396 discriminate the morphogroups on the first LDA axis correspond to the maximum length of
 397 the second abdominal pore (LP2), and the maximum length of the second pore of the
 398 abdomen/maximum height of the abdomen (R1).



399
 400 **Figure 2.** Scatter plot of the two first axes of the Linear Discriminant Analysis (LDA) conducted on the 13
 401 variables that constituted the morphometric data using four *a priori* identified morphogroups (indicated in the
 402 legend; ellipses correspond to the 95 % confidence intervals for each morphogroup). Abbreviations: W1:
 403 maximum width of the cephalis without the apical horn; W2: maximum width of the thorax; W3: maximum
 404 width of the abdomen; H1: maximum height of the cephalis; H2: maximum height of the thorax; H3: maximum
 405 height of the abdomen; TL: total length or height of the specimen without the apical horn; LP2: maximum length

406 of the second abdominal pore; NPV: number of abdominal pores aligned vertically; NPH: number of abdominal
 407 pores aligned horizontally; R1: maximum length of the second pore of the abdomen/H3; R2: number of
 408 abdominal pores aligned vertically/H3; R3: number of abdominal pores aligned horizontally/W3.

409

410 Upon considering three morphogroups achieved by lumping *P. chalara* and *Podocyrtis* sp. cf.
 411 *P. chalara*, the LDA ordination is highly similar to that obtained in the four-group analysis,
 412 with all variation represented on the first two axes (~92.87 % on axis 1; **Figure S1**).
 413 Classification results improved substantially, with 94.5 ± 3.1 % of correctly identified
 414 specimens in the three-group LDA (Table S2). Misclassification mainly occurred between *P.*
 415 *goetheana* and *Podocyrtis* sp. cf. *P. goetheana*, occasionally also between *P. chalara* and
 416 *Podocyrtis* sp. cf. *P. goetheana*, but never between *P. chalara* and *P. goetheana*.

417 **Table 4.** Average confusion matrix of the *Podocyrtis* morphotypes based on linear discriminant analysis (LDA)
 418 performed on the matrix of our 13 morphometric variables (.i.e., measurements, pore counts and ratios).

LDA results 4-group scenario

	P_sp_cf_P_cha_&_P_sp_cf_P_goe	P_cha_&P_sp_cf_P_cha	P_cha_&P_sp_cf_P_goe	P_goe_&P_sp_cf_P_cha	P_goe_&P_sp_cf_P_goe	P_goe_&P_cha	total_c orrect
res_							0.7351
mea	0.011162791	0.204651163	0.001627907	0	0.04744186	0	16279
n							0.0612
res_s	0.013819733	0.059589621	0.005963544	0	0.024717954	0	97375
d							

419

420

421 **4.2. Artificial neural networks**

422 **4.2.1. Classification using the four morphogroups**

423 First, we trained neural networks on the dataset of 428 images attributed to the four
 424 morphogroups (**Table 2** 4_classes dataset) using a CNN with a VGG16 architecture. Prior to
 425 the training and testing phases, images were manually grouped into four distinct classes (one

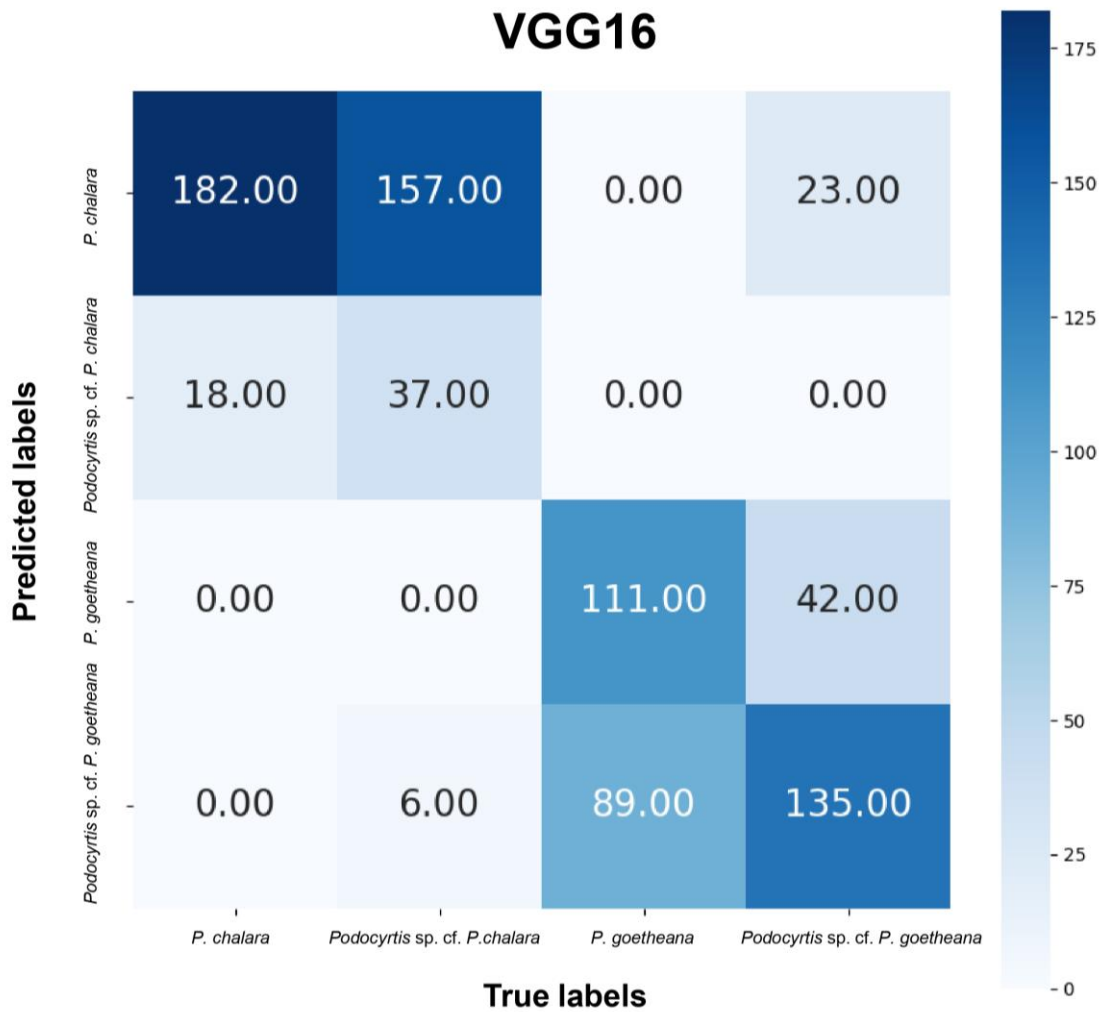
426 class per morphogroup). The analysis was run ten times with the pre-processing parameters
 427 outlined in **Table 5**, which resulted to an identification accuracy of 54.4 ± 1.7 % (**Figure**
 428 **3**). These results indicate that, although the network was able to partially identify the
 429 differences between the general morphologies of *P. chalara* + *Podocyrtis* sp. cf. *P. chalara*
 430 versus those of *P. goetheana* + *Podocyrtis* sp. cf. *P. goetheana*, it was not able to accurately
 431 distinguish all four morphogroups from each other.

432

433 **Table 5.** Summary of the tests performed on the 4_classes and 3_classes datasets using VGG16, STDP-based
 434 SNN and SuperSpike-based SNN. Accuracy values are averaged from the results of the respective 10 test runs
 435 and indicated as the mean \pm one standard deviation.

Dataset	Architecture	Test runs	Number of Epochs	Data pre-processing	Simulation time h:m:s	Accuracy (%) \pm std mean \pm std
4_classes	VGG16	10	20	Image resize to (224, 224) px	0:31:46	54.40 ± 1.74
	VGG16	10	20	Image resize to (224, 224) px	0:17:37	92.60 ± 0.77
3_classes	STDP-Network	10	100	Image resize to (128, 128) px & On-Off filter	0:05:43.43	90.40 ± 0.4949
	SuperSpike-Network	10	20	Image resize to (128, 128) px	0:23:26	84.42 ± 1.36

436



437

438 **Figure 3.** Example of a confusion matrix obtained from random single run – Analysis of 4_classes dataset with
 439 VGG16. Correct assignments were reached in 54.40 % of the cases. The color scale indicates the number of
 440 specimens.

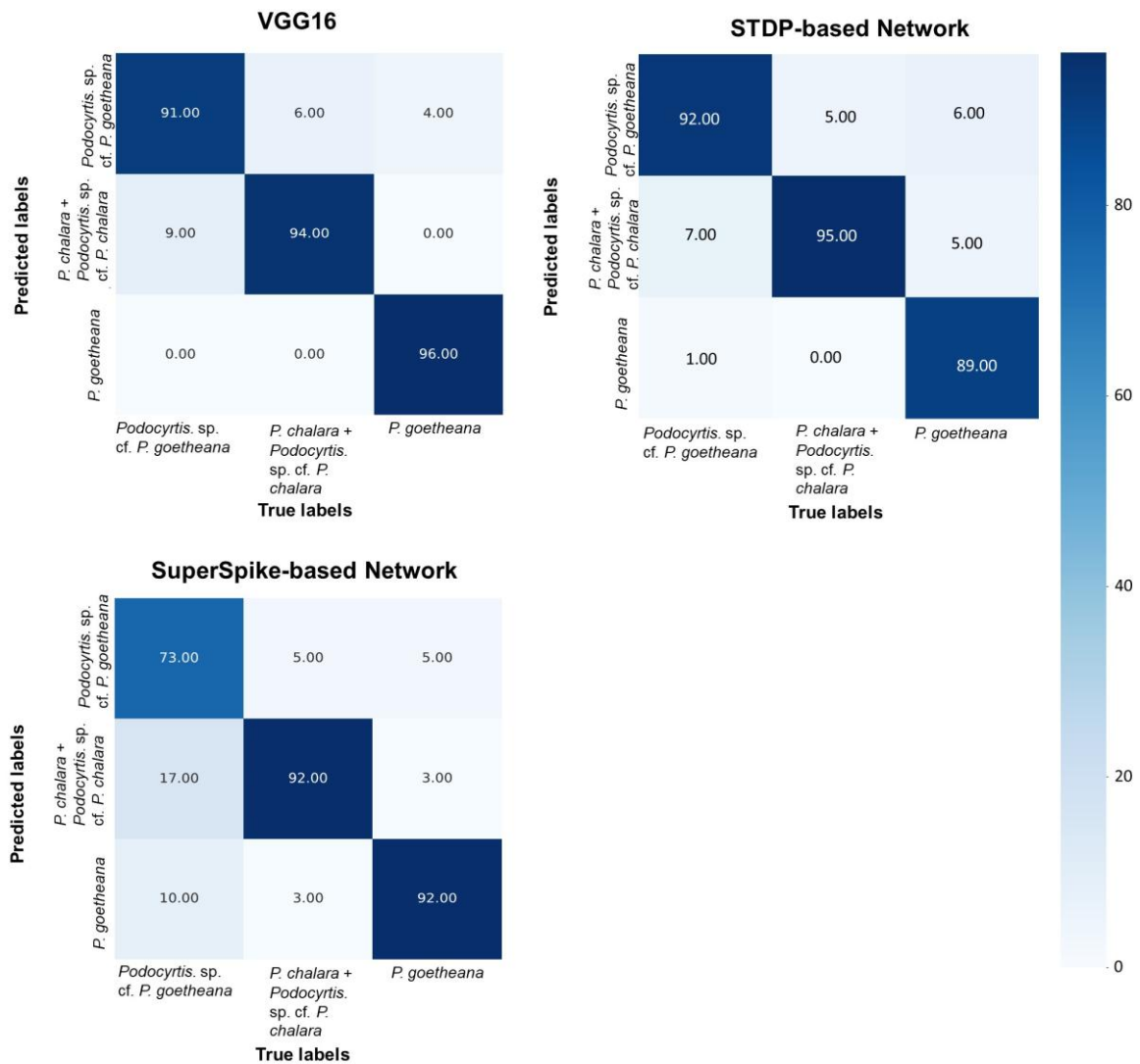
441

442 **4.2.2. Classification using the three morphogroups supported by LDA**

443 Subsequently, we trained neural networks using a dataset with *a priori* assignment to
 444 the three morphogroups that were recognized by LDA, i.e., *P. chalara* + *Podocyrtis sp. cf. P.*
 445 *chalara* *Podocyrtis sp. cf. P. goetheana* and *P. goetheana*. For these analyses we compared
 446 the performance of a VGG16 CNN, a STDP-based SNN and a SuperSpike-based SNN using
 447 the 3_classes dataset (**Table 3**). The resulting assignment results and average network

448 accuracies (**Figure 4**) indicate that, under our specified conditions, all neural networks are
449 able to accurately assign specimens to their correct class when three predefined classes are
450 used in combination with a large dataset of images. However, we observed substantial
451 differences in the speed to conclude analyses; thus the STDP-based SNN was the fastest
452 (around 6 minutes) due to the use of the local learning rule for training and one spike per
453 image per neuron principale. Moreover, VGG16 came second with a time of around 17
454 minutes due to the size of the network and the use of transfer learning and training of only the
455 last layer. Last came the SuperSpike-based SNN with approximately 23 minutes per run
456 because all layers of the network were trained from scratch during the training phase at the
457 start of each run using a global learning rule.

458



459

460 **Figure 4.** Examples of confusion matrices obtained from random single runs – Analyses of 3_classes dataset
 461 with VGG16, STDP-based SNN and SuperSpike-based SNN. Correct assignments were reached in 92.60 %,
 462 90.40 % and 84.42 % of the cases in VGG16, STDP-based SNN and SuperSpike-based SNN, respectively. The
 463 color scale indicates the number of specimens.

464

465

466 5. Discussion

467 In this study, we examined the morphological variability in Eocene *Podocyrtilis*
468 belonging to the anagenetic sequence that starts with *P. chalara* and ends with *P. goetheana*.
469 Specifically, we examined and compared the performance of two analytical approaches in
470 assigning individuals to *a priori* defined morphogroups that were constructed from qualitative
471 observations, i.e., *P. chalara*, *Podocyrtilis* sp. cf. *P. chalara*, *Podocyrtilis* cf. *P. goetheana* and
472 *P. goetheana*. The first approach involved LDA based on morphometric data, whereas the
473 second was a neural network approach based on automatic image recognition. Both methods
474 gave very similar results, which indicates that both morphometrics and image analysis
475 evaluated shape differences in a highly similar way, suggesting that the results obtained with
476 these two different methods are robust.

477 Comparing LDA and neural network approaches, assignment performances were
478 comparatively low when four morphogroups were considered in the morphological transition
479 from *P. chalara* to *P. goetheana*. The scatterplot of the LDA (**Figure 2**), indicated a large
480 morphospace overlap between *P. chalara* and *Podocyrtilis* sp. cf. *P. chalara*, revealing that the
481 qualitatively observed ‘differences’ were either not sampled in our datasets or that these
482 differences are part of a larger spectrum of morphological variation and not informative to
483 distinguish morphogroups. Given the similarity of LDA and neural networks, based on
484 different datasets, the second hypothesis is more likely; it appears that *P. chalara* and
485 *Podocyrtilis* sp. cf. *P. chalara* represent a single highly variable morphogroup. Morphospace
486 overlap is also observed in the LDA between *Podocyrtilis* sp. cf. *P. goetheana* and *P.*
487 *goetheana*, but it is much more limited than between *P. chalara* and *Podocyrtilis* sp. cf. *P.*
488 *chalara*. Finally, *Podocyrtilis* sp. cf. *P. goetheana* also overlaps with *P. chalara* and *Podocyrtilis*
489 sp. cf. *P. chalara* in morphospace occupation, however this overlap is small. The analyses
490 conducted with the machine learning approach on four morphogroups confirm the results of

491 LDA, as VGG16 had significant difficulties in differentiating specimens of *P. chalara* from
492 *Podocyrtis* sp. cf. *P. chalara* and vice-versa using stacked and segmented images, when all
493 four morphogroup classes were pre-defined. This was also the case for some specimens of
494 *Podocyrtis* sp. cf. *P. goetheana* and *P. goetheana*, which all resulted in an inferior
495 performance of VGG16 compared to that of the LDA for the scenario with four
496 morphogroups, i.e., 54.4 ± 1.7 % versus 73.5 ± 6.1 %, respectively.

497 When *P. chalara* and *Podocyrtis* sp. cf. *P. chalara* are lumped in the same morphogroup,
498 resulting in a three-group configuration, assignment probabilities improved strongly both for
499 LDA and neural networks. On average 94.5 ± 3.1 % of assignments were correct with LDA,
500 whereas 92.6 ± 0.8 %, 90.4 ± 0.5 % and 84.4 ± 1.4 % of the assignments were correct for
501 VGG16, STDP-based SNN and SuperSpike-based SNN, respectively. Neural networks,
502 mainly our CNN can accurately and quickly, assign specimens to the three pre-defined classes
503 using a large dataset. Whereas the CNN performed highly similarly to LDA, both SNNs we
504 used here performed less well, as had been documented for other tasks before (Tavanaej et al.,
505 2019). Runs with the STDP-based SNN ran to completion fastest and given that the accuracy
506 was only slightly reduced compared to LDA and VGG16, this approach may be preferred for
507 datasets that require a very long runtime with similar CNNs. The accuracy of the SuperSpike-
508 based SNN was significantly reduced compared to all the other classification methods that we
509 used. Further work is required to determine the cause of this underperformance, but the lower
510 accuracy for the SuperSpikeSuperSpike-based SNN is possibly due to the network size, as we
511 used only using eight layers in our case compared to 16 layers in VGG-16.

512 As for the performance evaluation of the neural network approach, it is noteworthy
513 that images were obtained manually for the purpose of our study, but advances in image
514 technology now allow that much of the image acquisition and preparation procedures
515 (photographing, stacking and segmentation) to be automated by the use of automatic

516 microscopes and modification of the AutoRadio_Segmenter plugin's code (Marchant et al.,
517 2020; Tetard et al., 2020). Using these automated procedures would facilitate the scalability
518 of the entire analysis with larger datasets. If such automated procedures were to be used,
519 constructing image datasets may potentially become more time-efficient than the various
520 procedures that are required to develop a morphometric dataset. Another advantage of using
521 neural network is their quick run time (**Table 5.**), although both accuracy and run time would
522 increase upon using larger image datasets. Furthermore, standardizing the rotation and
523 orientation of specimens is essential in morphometric studies; however, this requirement can
524 be relaxed for neural networks, as in our case re-rotating and re-orienting were used in the
525 data augmentation process to enlarge input datasets.

526 In our study, we evaluated assignment accuracy of LDA and machine learning with
527 neural networks based on *a priori* group assignments; however, in the future, it would be
528 useful to examine morphological variation without considering such assignments, e.g., by
529 using other ordination techniques and/or by using unsupervised machine learning techniques.
530 These methods could help in attempts to evaluate whether the three retained morphogroups
531 represent natural entities, although based on fossil evidence only such assessments are very
532 difficult. We refrain from such analyses here, as we believe these would be best conducted
533 with a larger set of material that ideally would cover the total geographic range and the total
534 stratigraphic interval covered by *P. chalara*, *Podocyrtis* sp. cf. *P. goetheana* and *P.*
535 *goetheana*.

536 Future work could also be focused on trying to evaluate the capacity of VGG16 to
537 accurately differentiate between the *P. chalara* and *Podocyrtis* sp. cf. *P. chalara* specimens,
538 with an altered set of images. Indeed, the weighted activation heatmap (Grad-CAM)
539 (Servaraju et al., 2017) generated from the runs with the 3-class dataset showed that the
540 analytical focus of the network was centered around the thoracic and abdominal walls of the

541 specimens (**Figure S22** in **Supplementary material**). One could try to develop a dataset
542 including several unstacked and non-segmented images per specimen, each one with a focus
543 on specific morphological features. This would allow some images to contain as much detail
544 as possible on abdominal features, whereas others would focus on other regions (e.g. the
545 thorax) and include blurred features of the outer walls and backside of the abdomen. We
546 hypothesize that these manipulations could force the network to focus its recognition
547 capabilities on a larger set of morphological features. If this hypothesis would be correct, it
548 would also facilitate the data acquisition process by eliminating the need to stack and segment
549 the images themselves. Alternatively, it is possible that providing more fragmented
550 information to neural networks would hamper an efficient learning process, with negative
551 consequences on the accuracy of the following predictions, somewhat similar to what we
552 observed in the VGG16 evaluation based on the 4-class dataset.

553

554 **6. Conclusion**

555 The aim of our work was to study the morphological variability in the anagenetic sequence of
556 *P. chalara* to *P. goetheana* and to evaluate the performance of recognizing and classifying
557 four *a priori* identified morphogroups with various machine learning algorithms based on
558 neural networks that use image data as direct input in comparison to linear discriminant
559 analysis using morphometric data. Our results demonstrate that LDA and neural networks
560 provide very similar outcomes, indicating robust performances. With both approaches we
561 encountered difficulties distinguishing *P. chalara* and *Podocyrtilis* sp. cf. *P. chalara*,
562 suggesting that the qualitative basis on which these morphogroups were recognized is to be
563 revised. For both approaches, assignment probabilities drastically increased for the scenario
564 where three morphogroups were recognized, lumping *P. chalara* and *Podocyrtilis* sp. cf. *P.*
565 *chalara*, and thus retaining *P. chalara* + *Podocyrtilis* sp. cf. *P. chalara*, *Podocyrtilis* sp. cf. *P.*

566 *goetheana* and *P. goetheana*. Further studies with more comprehensive sampling are required
567 to document the likelihood of *Podocyrtilis* sp. cf. *P. goetheana* representing a separate natural
568 entity, which additionally has implications for the position of bioevent RP16. However, our
569 results indicate that the morphometric data that we used for LDA samples the morphological
570 variability in the anagenetic sequence of *P. chalara* to *P. goetheana* in a comprehensive way.
571 Secondly, neural network approaches were able to correctly assign most specimens, and
572 therewith to accurately distinguish the three morphogroups directly from specimen images.
573 These results indicate that VGG16, STDP-based SNNs, and even SuperSpike-based SNNs are
574 capable of recognizing morphological variation in images and thus of reliably distinguishing
575 radiolarian morphogroups, which could facilitate identification and help with reaching more
576 objective taxonomic decisions. Furthermore, neural network approaches can be combined
577 with automated image acquisition and preparation procedures (photographing, stacking and
578 segmentation) that enable the creation of much larger image databases in a time-efficient
579 manner. Analyses based on neural network architecture could thus take a fraction of the time
580 that would be required for a trained taxonomist/(paleo)biologist to create and analyze
581 quantitative morphometric datasets. In conclusion, neural network approaches based on
582 images of (paleo)biological specimens may provide promising opportunities to guide more
583 objective taxonomic decisions.

584

585 **Data availability**

586 Microscopic slides are prepared and stored at UMR 8198 – Evo-Eco-Paleo of the University
587 of Lille, France. The datasets (<https://doi.org/10.57745/8KBOFP>, Pinto et al. 2023) have been
588 archived in the repository of the University of Lille at Recherche Data Gouv. The codes are
589 available at

590 <https://archive.softwareheritage.org/browse/directory/cc7d8ef1505299a208adcde597a98d90b>

591 [0ca47d6/https://archive.softwareheritage.org/browse/directory/cc7d8ef1505299a208adcde597](https://archive.softwareheritage.org/browse/directory/cc7d8ef1505299a208adcde597a98d90b0ca47d6/)
592 [a98d90b0ca47d6/](https://archive.softwareheritage.org/browse/directory/cc7d8ef1505299a208adcde597a98d90b0ca47d6/) (Elbez, 2023).

593

594 **Acknowledgments**

595 This study was partly funded by the French government through the program
596 “Investissements d’avenir” (I-ULNE SITE/ANR-16-IDEX-0004 ULNE) managed by the
597 National Research Agency. It also received funding from the European Union’s Horizon 2020
598 research and innovation program under the Marie Skłodowska-Curie grant agreement no.
599 847568. It was also supported by UMR 8198 Evo-Eco-Paléo and IRCICA (CNRS and Univ.
600 Lille USR-3380).

601 A special thanks to Mazdak Fatahi for very helpful discussions regarding machine learning
602 techniques and to Sylvie Regnier for help and advice with the sample preparation.

603

604

605

606

607

608

609

610

611

612 **References**

- 613 Balouek, D., Carpen Amarie, A., Charrier, G., Desprez, F., Jeannot, E., Jeanvoine, E., Lèbre,
614 A., Margery, D., Niclausse, N., Nussbaum, L., Richard, O., Perez, C., Quesnel, F.,
615 Rohr, C., & Sarzyniec, L., 2013, Adding virtualization capabilities to the Grid'5000
616 testbed: In *Cloud Computing and Services Science, Communications in Computer and
617 Information Science*). Springer International Publishing, v. 367, p. 3-20.
- 618 Beaufort, L., and Dollfus, D., 2004, Automatic recognition of coccoliths by dynamical neural
619 networks: *Marine Micropaleontology*, v. 51, p. 57-73.
620 doi:10.1016/j.marmicro.2003.09.003.
- 621 Bourel, B., Marchant, R., de Garidel-Thoron, T., Tetard, M., Barboni, D., Gally, Y., and
622 Beaufort, L., 2020, Automated recognition by multiple convolutional neural networks
623 of modern, fossil, intact and damaged pollen grains: *Computers and Geosciences*, v.
624 140, 104498. doi:10.1016/j.cageo.2020.104498.
- 625 Brocher, J., 2022, biovoxxel/BioVoxxel-Toolbox: BioVoxxel Toolbox v2.5.3. Zenodo.
626 doi:10.5281/zenodo.5986129.
- 627 Carlsson, V., Danelian, T., Boulet, P., Devienne, P., Laforge, A., and Renaudie, J., 2022,
628 Artificial intelligence applied to the classification of eight middle Eocene species of
629 the genus *Podocyrthis* (polycystine radiolaria): *Journal of Micropalaeontology*, v. 41, p.
630 165–182.
- 631 Cortes, C., and Vapnik, V., 1995, Support-vector networks: *Machine Learning*, v. 20, p. 273–
632 297.
- 633 Danelian, T., and Johnson, K.G., 2001, Patterns of biotic change in Middle Jurassic to Early
634 Cretaceous Tethyan radiolarian: *Marine Micropaleontology*, v. 43, p. 239-260.
635 doi:10.1016/S0377-8398(01)00029-9.

636 Danelian, T., and Macleod, N., 2019, Morphometric Analysis of Two Eocene Related
637 Radiolarian Species of the Podocyrtis (Lampterium) Lineage: Paleontological
638 Research, v. 23, p. 314–330.

639 Danelian, T., Zambetakis-Lekkas, A., Galoyan, G., Sosson, M., Asatryan, G., Hubert, B., and
640 Grigoryan, A., 2014, Reconstructing Upper Cretaceous (Cenomanian)
641 paleoenvironments in Armenia based on Radiolaria and benthic Foraminifera;
642 implications for the geodynamic evolution of the Tethyan realm in the Lesser
643 Caucasus: Palaeogeography, Palaeoclimatology, Palaeoecology, v. 413, p. 123–132.

644 Danelian, T., Baudin, F., Gardin, S., Masure, E., Ricordel, C., Fili, I., Meçaj, T., and Muska,
645 K., 2007, The record of mid Cretaceous oceanic anoxic events from the Ionian zone of
646 southern Albania: Revue de Micropaléontologie, v. 50, p. 225–237.

647 Danelian, T., Le Callonec, L., Erbacher, J., Mosher, D.C., Malone, M.J., Berti, D., Bice, K.L.,
648 Bostock, H., Brumsack, H.-J., Forster, A., Heidersdorf, F., Henderiks, J., Janecek, T.J.,
649 Junium, C., Macleod, K., Meyers, P.A., Mutterlose, J.H., Nishi, H., Norris, R.D., Ogg,
650 J.G., O'Regan, M.A., Rea, B., Sexton, P., Sturt-Fredricks, H., Suganuma, Y., Thurow,
651 J.W., Wilson, P.A., Wise, S.W., and Glatz, C., 2005, Preliminary results on
652 Cretaceous-Tertiary tropical Atlantic pelagic sedimentation (Demerara Rise, ODP Leg
653 207): Comptes Rendus Geoscience, v. 337, p. 609–616.

654 Dollfus, D., and Beaufort, L., 1999, Fat neural network for recognition of position-normalized
655 objects. Neural Networks, v.12, p. 553-560. doi:10.1016/S0893-6080(99)00011-8.

656 [code] Elbez, H., 2023, Code for: Morphometrics and machine learning discrimination of the
657 middle Eocene radiolarian species *Podocyrtis chalara*, *P. goetheana* and their
658 morphological intermediates, deposited at
659 [https://archive.softwareheritage.org/browse/directory/cc7d8ef1505299a208adcde597a](https://archive.softwareheritage.org/browse/directory/cc7d8ef1505299a208adcde597a98d90b0ca47d6/)
660 [98d90b0ca47d6/](https://archive.softwareheritage.org/browse/directory/cc7d8ef1505299a208adcde597a98d90b0ca47d6/) 2023-08-24.

661 Haeckel, E., 1887, Report on the Radiolaria collected by H.M.S. Challenger during the years
662 1873-1876. Report on the Scientific Results of the Voyage of H.M.S. Challenger
663 during the years 1873-1876. 18: 1-1803.

664 Falez, P., 2019, Improving Spiking Neural Networks Trained with Spike Timing Dependent
665 Plasticity for Image Recognition [phdthesis]: Université de Lille.

666 Hijazi, S., Kumar, R., and Rowen, C., 2015, Using Convolutional Neural Networks for Image
667 Recognition: Cadence, Cadence Design Systems Inc., p. 1–12.

668 Hsiang, A.Y., Brombacher, A., Rillo, M.C., Mleneck-Vautravers, M.J., Conn, S., Lordsmith,
669 S., Jentzen, A., Henehan, M.J., Metcalfe, B., Fenton, I.S., Wade, B.S., Fox, L.,
670 Meilland, J., Davis, C. V., Baranowski, U., Groeneveld, J., Edgar, K.M., Movellan,
671 A., Aze, T., Dowsett, H. J., Giles Miller, C., Rios, N., and Hull, P. M., 2019, Endless
672 Forams: >34,000 modern planktonic foraminiferal images for taxonomic training and
673 automated species recognition using convolutional neural networks: *Paleoceanography*
674 and *Paleoclimatology*, 34, 1157-1177. doi:10.1029/2019PA003612.

675 Itaki, T., Taira, Y., Kuwamori, N., Saito, H., Ikehara, M., and Hoshino, T., 2020, Innovative
676 microfossil (radiolarian) analysis using a system for automated image collection and
677 AI-based classification of species: *Scientific Reports*. doi:10.1038/s41598-020-77812-
678 6.

679 Lever, J., Krzywinski, M., and Altman, N., 2017, Principal component analysis: *Nature*
680 *Methods*, v. 14, p. 641–642.

681 Maass, W., 1997, Networks of spiking neurons: The third generation of neural network
682 models: *Neural Networks*, v. 10, p. 1659–1671.

683 Marchant, R., Tetard, M., Pratiwi, A., Adebayo, M., and de Garidel-Thoron, T., 2020,
684 Automated analysis of foraminifera fossil records by image classification using a
685 convolutional neural network: *Journal of Micropalaeontology*, v. 39, p. 183–202.

686 Masquelier, T., and Thorpe, S.J., 2007, Unsupervised Learning of Visual Features through
687 Spike Timing Dependent Plasticity: *PLOS Computational Biology*, v. 3, p. e31.

688 Meunier, M., and Danelian, T., 2022, Astronomical calibration of late middle Eocene
689 radiolarian bioevents from ODP Site 1260 (equatorial Atlantic, Leg 207) and
690 refinement of the global tropical radiolarian biozonation: *Journal of*
691 *Micropalaeontology*, v. 41, p. 1–27.

692 Meunier, M., and Danelian, T., 2023, Progress in understanding middle Eocene nassellarian
693 (Radiolaria, Polycystinea) diversity; new insights from the western equatorial Atlantic
694 Ocean: *Journal of Paleontology*, v. 97, p. 1–25.

695 Mitra, R., Marchitto, T.M., Ge, Q., Zhong, B., Kanakiya, B., Cook, M.S., Fehrenbacher, J.S.,
696 Ortiz, J.D., Tripathi, A., and Lobaton, E., 2019, Automated species-level identification
697 of planktic foraminifera using convolutional neural networks, with comparison to
698 human performance: *Marine Micropaleontology*, v. 148, p. 1-14.
699 doi:10.1016/j.marmicro.2019.01.005.

700 Moore, T.C., 1972, Mid-Tertiary Evolution of the Radiolarian Genus *Calocycletta*:
701 *Micropaleontology*, v. 18, p. 144–152.

702 Nigrini, C.A., Sanfilippo, A., and Moore, T.J., 2005, Radiolarian biostratigraphy and
703 chronology of radiolarian events of ODP Leg 199 sites and EW9709 sediment cores:
704 Supplement to: Nigrini, CA et al. (2005): Cenozoic Radiolarian Biostratigraphy: A
705 Magnetobiostratigraphic Chronology of Cenozoic Sequences from ODP Sites 1218,
706 1219, and 1220, Equatorial Pacific. In: Wilson, PA; Lyle, M; Firth, JV (Eds.)
707 Proceedings of the Ocean Drilling Program, Scientific Results, College Station, TX
708 (Ocean Drilling Program), v. 199, p. 1-76,
709 <https://doi.org/10.2973/Odp.Proc.Sr.199.225.2005>.

710 Obut, O., and Iwata, K., 2000, Lower Cambrian Radiolaria from the Gorny Altai (southern
711 West Siberia): Journal of Geology and Geophysics, v. 41.

712 Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson,
713 G.L., Solymos, P., Stevenes, M.H.H., Wagner, H., 2013, Package 'vegan': Community
714 ecology package, v. 2, p. 1–295.

715 Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z.,
716 Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison,
717 M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S., 2019,
718 PyTorch: An Imperative Style, High-Performance Deep Learning Library: 33rd
719 NeurIPS, Vancouver, British Columbia, Canada, <https://arxiv.org/abs/1912.01703>

720 Pehle, C.-G., and Pedersen, J.E., 2021, Norse - A deep learning library for spiking neural
721 networks (0.0.5), Zenodo, <https://doi.org/10.5281/zenodo.4422025>.

722 Pinto, F., Carlsson, V., Meunier, M., Van Bocxlaer., B., Elbez, H., Cueille, M., Boulet, P and
723 Danelian, T, 2023, Data for the: Morphometrics and machine learning discrimination
724 of the middle Eocene radiolarian species *Podocyrstis chalara*, *P. goetheana* and their
725 morphological intermediates, Recherche Data Gouv,
726 <https://doi.org/10.57745/8KBOFP>.

727 Pouille, L., Obut, O., Danelian, T., and Sennikov, N., 2011, Lower Cambrian (Botomian)
728 polycystine Radiolaria from the Altai Mountains (southern Siberia, Russia): Comptes
729 Rendus Palevol, v. 10, p. 627–633.

730 R Core Team, 2022, R: a language and environment for statistical computing: R Foundation
731 for Statistical Computing. <https://www.R-project.org>.

732 Renaudie, J., Danelian, T., Saint Martin, S., Le Callonnec, L., and Tribovillard, N., 2010,
733 Siliceous phytoplankton response to a Middle Eocene warming event recorded in the

734 tropical Atlantic (Demerara Rise, ODP Site 1260A): *Palaeogeography,*
735 *Palaeoclimatology, Palaeoecology*, v. 286, p. 121–134.

736 Renaudie, J., Lazarus, D., 2013, On the accuracy of paleodiversity reconstructions: a case
737 study in Antarctic Neogene radiolarians: *Paleobiology*, v. 39(3), p. 491-509.

738 Renaudie, J., Gray, R., and Lazarus, D.B., 2018, Accuracy of a neural net classification of
739 closely-related species of microfossils from a sparse dataset of unedited images. *PeerJ*
740 *Preprints*, 6:e27328v1.

741 Riedel, R.W., 1971, Cenozoic Radiolaria from the western tropical Pacific, Leg 7.: *Init. Repts.*
742 *DSDP*, v. 7, p. 1592–1627.

743 Riedel, W.R., and Sanfilippo, A., 1978, Stratigraphy and Evolution of Tropical Cenozoic
744 Radiolarians: *Micropaleontology*, v. 24, p. 61–96.

745 Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D., Ripley, M.B.,
746 2013, Package ‘mass’: *Cran r*, 538, p. 113–120.

747 Rohlf, F., and Bookstein, F., 1991, Size and Shape (Book Reviews: Proceedings of the
748 Michigan Morphometrics Workshop.): *Science*, v. 253, p. 345-362.

749 Sanfilippo, A., and Riedel, W.R., 1970, Post-Eocene “Closed” Theoperid Radiolarians:
750 *Micropaleontology*, v. 16, p. 446–462.

751 Sanfilippo, A., and Riedel, W.R., 1992, The Origin and Evolution of Pterocorythidae
752 (Radiolaria): A Cenozoic Phylogenetic Study: *Micropaleontology*, v. 38, p. 1.

753 Sanfilippo, A., and Nigrini, C., 1998, Code numbers for Cenozoic low latitude radiolarian
754 biostratigraphic zones and GPTS conversion tables: *Marine Micropaleontology*, v. 33,
755 p. 109–156.

756 Schneider, C.A., Rasband, W.S., and Eliceiri, K.W., 2012, NIH Image to ImageJ: 25 years of
757 image analysis: *Nature Methods*, v. 9, p. 671–675.

758 Simonyan, K., and Zisserman, A., 2015, Very Deep Convolutional Networks for Large-Scale
759 Image Recognition: ICLR, San Diego, California, USA,
760 <https://arxiv.org/abs/1409.1556>.

761 Tavanaei, A., Ghodrati, M., Kheradpisheh, S. R., Masquelier, T., & Maida, A., 2019, Deep
762 learning in spiking neural networks, *Neural Networks*, v. 111, p. 47-63,
763 doi:10.1016/j.neunet.2018.12.002.

764 Tetard, M., Monnet, C., Noble, P., and Danelian T., 2017, Biodiversity patterns of Silurian
765 Radiolaria: *Earth-Science Reviews*, v.173, p. 77-83.

766 Tetard, M., Marchant, R., Cortese, G., Gally, Y., de Garidel-Thoron, T., and Beaufort, L.,
767 2020, Technical note: A new automated radiolarian image acquisition, stacking,
768 processing, segmentation and identification workflow: *Climate of the Past*, v. 16, p.
769 2415–2429.

770 Venables, W.N., and Ripley, B.D., 2002, Random and Mixed Effects, *in* Venables, W.N. and
771 Ripley, B.D., eds., *Modern Applied Statistics with S*, Statistics and Computing: New
772 York, NY, Springer, p. 271–300.

773 Watanabe, M., Kawagata, S., Aita, Y., Suzuki, N., and Kamikuri, S., 2022, Changes in
774 morphological parameters of the radiolarian *Lampterium* lineage from the middle
775 Eocene in the tropical Pacific: *Marine Micropaleontology*, v. 173, 102125.

776 Westerhold, T., and Röhl, U., 2013, Orbital pacing of Eocene climate during the Middle
777 Eocene Climate Optimum and the chron C19r event: Missing link found in the tropical
778 western Atlantic: *Orbital Pacing of Eocene Climate: Geochemistry, Geophysics,*
779 *Geosystems*, v. 14, p. 4811–4825.

780 Yang, Z.R., and Yang, Z., 2014, Artificial Neural Networks, *Comprehensive Biomedical*
781 *Physics*: Elsevier, p. 1–17.

782 Zenke, F., and Ganguli, S., 2018, SuperSpike: Supervised Learning in Multilayer Spiking
783 Neural Networks: Neural Computation, v. 30, p. 1514–1541