Fasciola gigantica, F. hepatica and Fasciola intermediate forms: geometric morphometrics and an artificial neural network to help morphological identification

Suchada Sumruayphol [Equal first author, 1], **Praphaiphat Siribat** [Equal first author, 2], **Jean-Pierre Dujardin** [3], **Sébastien Dujardin** [3], **Chalit Komalamisra** [4], **Urusa Thaenkham** [Corresp. 5]

1 Department of Medical Entomology, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

2 Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok, Thailand

3 IRD, UMR, INTERTRYP IRD, CIRAD, University of Montpellier, Montpellier, France

4 Mahidol-Bangkok School of Tropical Medicine, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

5 Department of Helminthology, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

Corresponding Author: Urusa Thaenkham
Email address: urusa.tha@mahidol.ac.th

Background. *Fasciola hepatica* and *F. gigantica* cause fascioliasis in both humans and livestock. Some adult specimens of *Fasciola* sp. referred to as "intermediate forms" based on their genetic traits, are also frequently reported. Simple morphological criteria are unreliable for their specific identification. In previous studies, promising phenotypic identification scores were obtained using morphometrics based on linear measurements (distances, angles, curves) between anatomical features. Such an approach is commonly termed "traditional" morphometrics, as opposed to "modern" morphometrics, which is based on the coordinates of anatomical points. **Methods.** Here, we explored the possible improvements that modern methods of morphometrics, including landmark-based and outline-based approaches, could bring to solving the problem of the non-molecular identification of these parasites. *F. gigantica* and *Fasciola* intermediate forms suitable for morphometric characterization were selected from Thai strains following their molecular identification. Specimens of *F. hepatica* were obtained from the Liverpool School of Tropical Medicine (UK). Using these three taxa, we tested the taxonomic signal embedded in traditional linear measurements versus the coordinates of anatomical points (landmark- and outline-based approaches). Various statistical techniques of validated reclassification were used, based on either the shortest Mahalanobis distance, the maximum likelihood, or the artificial neural network method. **Results.** Our results revealed that both traditional and modern morphometric approaches can help in the morphological identification of *Fasciola* sp. We showed that the accuracy of the traditional approach could be improved by selecting a subset of characters among the most contributive ones. The influence of size on discrimination by shape was much more important in traditional than in modern

analyses. In our study, the modern approach provided different results according to the type of data: satisfactory when using pseudolandmarks (outlines), less satisfactory when using landmarks. The different reclassification methods provided approximately similar scores, with a special mention to the neural network, which allowed improvements in accuracy by combining data from both morphometric approaches. **Conclusion.** We conclude that morphometrics, whether traditional or modern, represent a valuable tool to assist in *Fasciola* species recognition. The general level of accuracy is comparable among the various methods, but their demands on skills and time differ. Based on the outline method, our study could provide the first description of the shape differences between species, highlighting the more globular contours of the intermediate forms.

1 *Fasciola gigantica*, *F. hepatica* and *Fasciola* intermediate forms: geometric

2 morphometrics and an artificial neural network to help morphological

3 identification

5 Suchada Sumruayphol[1,*], Praphaipat Siribat[2,5,*], Jean-Pierre Dujardin[3], Sébastien Dujardin[3],

6 Chalit Komalamisra[4], Urusa Thaenkham[5]

7 [1] Department of Medical Entomology, Faculty of Tropical Medicine, Mahidol University,

8 Bangkok 10400, Thailand

9 [2] Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok 10400,

10 Thailand

11 [3] IRD, UMR INTERTRYP IRD, CIRAD, University of Montpellier, F-34398 Montpellier,

12 France.

13 [4] Mahidol-Bangkok School of Tropical Medicine, Faculty of Tropical Medicine, Mahidol

14 University, Bangkok, Thailand

15 [5] Department of Helminthology, Faculty of Tropical Medicine, Mahidol University, Bangkok

16 10400, Thailand

17 *These authors contributed equally to this work

18 Corresponding Author:

19 Urusa Thaenkham[5]

20 420/6 Ratchawithi Road, Ratchathewi, Bangkok, Thailand

21 Email address: urusa.tha@mahidol.ac.th

25 **ABSTRACT**

26 **Background.** *Fasciola hepatica* and *F. gigantica* cause fascioliasis in both humans and livestock.

27 Some adult specimens of *Fasciola* sp. referred to as "intermediate forms" based on their genetic

28 traits, are also frequently reported. Simple morphological criteria are unreliable for their specific

29 identification. In previous studies, promising phenotypic identification scores were obtained using

30 morphometrics based on linear measurements (distances, angles, curves) between anatomical

31 features. Such an approach is commonly termed "traditional" morphometrics, as opposed to

32 "modern" morphometrics, which is based on the coordinates of anatomical points.

33 **Methods.** Here, we explored the possible improvements that modern methods of morphometrics,

34 including landmark-based and outline-based approaches, could bring to solving the problem of the

35 non-molecular identification of these parasites. *F. gigantica* and *Fasciola* intermediate forms

36 suitable for morphometric characterization were selected from Thai strains following their

37 molecular identification. Specimens of *F. hepatica* were obtained from the Liverpool School of

38 Tropical Medicine (UK). Using these three taxa, we tested the taxonomic signal embedded in

39 traditional linear measurements versus the coordinates of anatomical points (landmark- and

40 outline-based approaches). Various statistical techniques of validated reclassification were used,

41 based on either the shortest Mahalanobis distance, the maximum likelihood, or the artificial neural

42 network method.

43 **Results.** Our results revealed that both traditional and modern morphometric approaches can help

44 in the morphological identification of *Fasciola* sp. We showed that the accuracy of the traditional

45 approach could be improved by selecting a subset of characters among the most contributive ones.

46   The influence of size on discrimination by shape was much more important in traditional than in

47   modern analyses. In our study, the modern approach provided different results according to the

48   type of data: satisfactory when using pseudolandmarks (outlines), less satisfactory when using

49   landmarks. The different reclassification methods provided approximately similar scores, with a

50   special mention to the neural network, which allowed improvements in accuracy by combining

51   data from both morphometric approaches.

52   **Conclusion.** We conclude that morphometrics, whether traditional or modern, represents a

53   valuable tool to assist in *Fasciola* species recognition. The general level of accuracy is comparable

54   among the various methods, but their demands on skills and time differ. Based on the outline

55   method, our study could provide the first description of the shape differences between species,

56   highlighting the more globular contours of the intermediate forms.

INTRODUCTION

58 *Fasciola hepatica* Linnaeus, 1758 and *F. gigantica* Cobbold, 1855 are well-known as parasites

59 that cause fascioliasis (*Mas-Coma, Bargues & Valero, 2005*). The parasites belong to the family

60 Fasciolidae (Railliet, 1895), subfamily Fasciolinae (Stiles & Hassall, 1898). The adult worm

61 infects the liver of various species of mammals, particularly livestock that consume fresh plant

62 material contaminated with infective-stage metacercaria *(Mas-Coma, Bargues & Valero 2018*).

63 Climatic changes seem to have exacerbated the problem by expanding the geographic distribution

64 of such parasites (*Mas-Coma, Valero & Barynes 2008; 2009a*). Typically, *F. hepatica* is reported

65 worldwide, while *F. gigantica* occurs in a narrower range across Asia and Africa. The geographic

66 distribution of these two parasites can overlap, particularly in the equatorial zone. The hybrid

67 forms that can result between *F. hepatica* and *F. gigantica* are often reported as *Fasciola*

68 "intermediate forms" (*Periago et al., 2008; Mas-Coma, Valero & Barynes 2009a; 2009b*). Like

69 other countries in the tropical and subtropical areas, Thailand has reported *Fasciola* infections in

70 both humans and livestock (*Mas-Coma, Valero & Barynes 2009a; Buranasin & Harinasuta, 1970;*

71 *Srihakim & Pholpark, 1991; Aroonroch et al., 2006; Kanoksil et al., 2006; Tippawangkosol et al.,*

72 *2017; Wannasan et al., 2014*).

73

74 Up until now, traditional morphometrics, which is based on linear measurements between

75 anatomical landmarks, has been used to help in the identification of *Fasciola* species (*Valero,*

76 *Panova & Mas-Coma, 2005*). Many factors have been noted which can modify the accuracy of

77 phenotypical classification done in this way, including: 1) lymnaeid-host specificity, 2) bias arising

78 from different types of definitive hosts, the technical bias for worm fixing, 3) the computerized

79 image analysis system used, 4) the standardized methodology used for measurements, and 5) the

80   effect of allometric growth (*Valero, Panova & Mas-Coma, 2005; Periago et al., 2006*). Size and

81   some other variables or ratios that have been erroneously considered as a proxy for "shape" (body

82   roundness (BR = BP2/4πBA; BP = body perimeter; and BA = body area) could discriminate

83   between adult *F. hepatica*, *F. gigantica*, and *Fasciola* intermediate forms collected from different

84   countries in Asia (*Periago et al., 2006*), as well as being used to examine intraspecific diversity

85   (*Periago et al., 2007*). A molecular identification technique was developed to discriminate

86   between fasciolid adult worms based on ribosomal internal transcribed spacers (ITS) 1 and 2 (*Lin*

87   *et al., 2007; Lotfy et al., 2008*). According to Amer et al. (2011), the identification of fasciolids

88   requires at least ITS2 and mitochondrial cox1 or nad1 sequences. Itagaki et al. (2011) suggested

89   that ITS1 sequences, which do not show variation in Europe or Oceania, are nonetheless useful

90   molecular markers for discriminating *F. hepatica, F. gigantica*, and their hybrids occurring in Asia

91   (*Amer et al., 2011*).

92

93   In Thailand, *Fasciola* intermediate forms were found in the livers of cattle in slaughterhouses and

94   identified as hybrids because of the heterosequences of nuclear ribosomal ITS1 and ITS2

95   (*Chaichanasak et al., 2012; Siribat et al., 2018*). For molecular discrimination between *F.*

96   *gigantica*, *F. hepatica*, and *Fasciola* intermediate forms, ITS1 has been noted as the most effective

97   DNA marker (*Chaichanasak et al., 2012*). *Fasciola* intermediate forms present no qualitative

98   morphological attributes that allow clear identification. Alongside molecular discrimination

99   techniques, traditional morphometrics has been suggested as a relatively simple and effective

100   approach for species identification (*Periago et al., 2006; Valero, Marcos & Mas-Coma, 1996;*

101   *Valero, Panova & Mas-Coma, 2005; Ashrafi et al., 2006; Afshan et al., 2014*). However, the

102   dimensions of *Fasciola* intermediate forms may overlap with those of both *F. gigantica* and *F.*

103   *hepatica* (*Periago et al., 2006; Srimuzipo et al., 2000*).

104

105   Even when using many characters and multivariate statistical tools, the risk of using a

106   morphometrics method is that linear distance measurements mainly provide information relating

107   to size and provide limited information about shape. Previous studies inaccurately labeled as

108   "shape" some variables which are actually another estimation of size, such as body roundness (BR

109    $= BP^2/4\pi BA$), body perimeter (BP), and body area (BA) (*Periago et al., 2006; Mas-Coma, Valero*

110   *& Bargues 2009b; Valero et al., 2018*). These characters are secondary constructions from linear

111   measurements and produce variables redundant with size (*Burnaby, 1966*).

112

113   A more convenient approach has recently been applied which enables variations in shape to be

114   estimated from linear measurements (*Valero et al., 2018*). However, there are various techniques

115   available that can be used for this. A comprehensive review was recently published distinguishing

116   two "schools" via which it is possible to approach the allometric concept: the Huxley–Jolicoeur

117   and the Gould–Mosimann schools (*Klingenberg, 2016*). The former, initially developed for

118   bivariate analyses, considers morphological form as a single unified feature (form as both size and

119   shape), emphasizing the covariation among traits as a consequence of variation in size. The latter,

120   initially developed for multivariate morphometric measurements, separates two estimates: an

121   estimate of global size and a multivariate estimate of shape (*Mosimann, 1970; Darroch and*

122   *Mosimann, 1985*). According to Klingenberg (2016), landmark-based geometric morphometrics

123   represents a straightforward extension to this second method. Its concept may be naturally

124   extended to the outline-based approach, which also separates a global estimator of size and a

125 multivariate estimate of shape (*Rohlf & Archie, 1984; Lestrel, 1989*). Instead of using distances

126 between anatomical landmarks, the modern methods use their coordinates in a common system of

127 axes. One of the clear improvements introduced by the modern approach is that it enables a visual

128 representation of shape (*Rohlf & Marcus, 1993*). Visualization of changes in shape are frequently

129 presented using D'Arcy Thompson-like plots, or deformation grids (*Bookstein, 1991*).

130 Deformation grids help to visualize the shifts in aligned coordinates from one shape to another.

131 The grid is a visual aid and can even be amplified to improve the visibility of subtle changes. It

132 can be omitted when these changes are clearly visible, as is frequently the case between species,

133 without the need for any amplification artefacts.

134

135 The use of coordinates usually applies to animal species presenting a solid structure on their body.

136 However, this method has also been applied to soft structures. For example, when combined with

137 photogrammetric techniques, it can reveal variations in the shape of rodent sperm-cell heads and

138 the diversity of these shape variations among species (*Varea Sánchez, Bastir & Roldan, 2013*).

139

140 In our study, we used recommended traditional morphometric techniques and, to preserve the

141 parallelism of shape extraction from the data, we applied the Darroch & Mosimann technique

142 (1985) to the linear measurements, which produces "log-shape ratios" as estimates of shape.

143 We compared the results obtained by traditional techniques with those obtained from two methods

144 based on the coordinates of anatomical points: the landmark-based (*Rohlf, 1990*) and the outline-

145 based (*Kuhl & Giardina, 1982; Lestrel, 1997*) approaches. These two approaches are conceptually

146 different. The landmark-based approach uses anatomical points called "landmarks". These points

147 are small biological structures, as small as a point at the requisite scale. They are considered to be

148   homologous in the sense that they are relocatable points from one individual to another. The

149   outline-based approach also uses an ordered set of discrete point coordinates; these are

150   superficially similar to landmark data, but they are conceptually quite different. They are of a

151   different nature to true landmarks because comparability between them is not expected separately,

152   but from the structure they describe. The outline-based method is generally restricted to closed

153   contours (outlines) where landmarks are lacking. In this approach, the contour is described by

154   points that we call "pseudolandmarks" (*Dujardin, 2017*). There is no need for these

155   pseudolandmarks to be equally spaced, and there is no need to have the same number of them from

156   one individual to another (*Rohlf, 1990a; Dujardin et al. 2014*).

157

158   Pseudolandmarks should not be confused with "semilandmarks". Semilandmarks are landmarks

159   that lack one degree of freedom because they depend on other landmarks. They are generally used

160   to compare open curves between other landmarks, and, most importantly, they are processed as

161   true landmarks by Procrustes superposition, combined with a sliding procedure to minimize total

162   bending energy or to minimize the Procrustes distances to the consensus (*Dujardin, 2019*).

163

164   Outline data are processed in a very different way, either by elliptic Fourier analysis (*Kuhl &*

165   *Giardina, 1982*) or, less frequently, by eigenshape analysis (*Lohmann, 1983*). Briefly, the observed

166   contour is decomposed in terms of sine and cosine curves of successive frequencies called

167   harmonics, with each harmonic being described by four coefficients. These coefficients are

168   normalized by the semi-grand axis of the first harmonic ellipse. The latter may be used as an

169   estimate of global size, as well as the root-squared area of this same starting ellipse, the root-

170   squared area of the digitized contour, or its perimeter. We explored the capacity of these different

171   approaches to discriminate between *F. hepatica*, *F. gigantica*, and *Fasciola* intermediate forms.

172

173   To ensure correct evaluation of the methods, the samples collected from Thailand were also

174   identified using molecular analysis. In addition to the Thailand samples, nine *F. hepatica* strains

175   from the UK were provided and used to give reference data: they were subjected to morphometric

176   analyses only.

177

178   As far as we know, morphometrics applied to the *Fasciola* identification always relied on

179   traditional morphometric techniques. In this study, we tried to answer the following question:

180   could modern morphometric techniques bring any improvement over traditional ones in terms of

181   accuracy and/or simplicity?

182

183   **Materials & methods**

184   **Sample collection**

185   Ninety adult worms of *Fasciola* spp. were collected from the livers of cattle slaughtered in local

186   slaughterhouses in Lamphun, Mae Hong Son, Tak, and Pratum Thani Provinces, Thailand. The

187   cattle in the Pratum Thani slaughterhouse were from Kanchanaburi Province, Thailand. All cattle

188   came from local farms in Thailand. Infected livers were stored on ice and transferred to the

189   laboratory at the Department of Helminthology, Faculty of Tropical Medicine, Mahidol

190   University, within 12 hours and within 1 hour for the samples from Pratum Thani province. Adult

191   *Fasciola* worms were gently separated from the liver tissue. Each individual worm was fixed

192   between two transparent glass slides containing two pieces of paper (0.4 mm) and the two slides

193   were gently held in place using a rubber band. The specimens were placed in normal saline for 1

194   hour. The lateral part of the body of each adult worm was then cut away and stored individually in

195   absolute alcohol at -20 °C until being used for the molecular study. The remaining section of each

196   worm's body was washed in distilled water for 5 min to relax the structure before being placed

197   individually between glass slides. The fixed adult worms were tentatively identified under a

198   stereomicroscope using the taxonomic criteria of primary morphological characters, including

199   cephalic cone, body length, and body width (*Periago et al., 2008; Valero, Marcos & Mas-Coma,*

200   *1996*). The specimens were fixed in 70% ethanol and stored at 4 °C for staining.

201

202   In addition to our field specimens, nine *F. hepatica* specimens were obtained from the Liverpool

203   School of Tropical Medicine (UK). These specimens were used for comparison and subjected to

204   both traditional and geometric morphometric examination. All specimens of *F. hepatica*, which

205   had sheep as their definitive host, were received as permanently mounted slides and deposited in

206   our laboratory in 1993. They were stained with Semichon's acid-carmine and mounted with

207   Canada balsam.

208

209   To make sure that our relatively small samples did not generate outlier values, we

210   compared graphically the observed means obtained using our material to values already published

211   for the same hosts in different countries (*Valero, Panova & Mas-Coma, 2005; Halakou et al.,*

212   *2017; Shafiei et al., 2014; Iyiola et al., 2018*).

213

214   **Molecular identification by PCR-RFLP**

215   A small piece of adult *Fasciola* worm was cut from the lateral part of the body. Genomic DNA

216   was extracted from each individual using a Tissue Genomic DNA Mini Kit (Geneaid, Taipei,

217   Taiwan), according to the manufacturer's protocol. The PCR amplicons of the ribosomal ITS1 and

218   ITS2 regions were amplified using the primers ITS1-F and ITS1-R for ITS1 and 3S and BD2 for

219   ITS2 (*Siribat et al., 2018; Ichikawa & Itagaki, 2010; Huang et al., 2004*). We conducted PCR

220   amplification of the ITS1 and ITS2 regions three times for each sample. Each PCR reaction was

221   performed in 50 mL final volume composed of $1 \times$ Top*Taq* Master Mix Kit (Qiagen, Hilden,

222   Germany) (1U Top*Taq* polymerase, 1.5 mM $MgCl_2$, and Top*Taq* polymerase buffer), 20 pmol of

223   each primer, and 10 ng/µL genomic DNA template. The PCR cycles consisted of initial

224   denaturation at 95 °C for 3 min followed by 29 cycles of denaturation at 95 °C for 45 s, annealing

225   at 55 °C for 30 s, and extension at 72 °C for 60 s. The final extension of each PCR reaction was

226   conducted at 72 °C for 8 min.

227

228   After purifying the PCR products using the ethanol precipitation method, 500 ng of purified PCR

229   products from each sample was digested with restriction enzymes *Rsa*I and *Nla*III for ITS1 and

230   ITS2, respectively (New England BioLab Inc., Massachusetts, USA) at 37 °C for 150 min. The

231   enzymatic digestion was inactivated by heating at 80 °C for 20 min. The final digested PCR

232   amplicons were run on a 2% agarose gel at 50 V for 180 min. The band patterns were visualized

233   using a UV transilluminator. The PCR-RFLP figure was photographed with Gel Documentation

234   (G-Box (HR), Syngene, UK).

235

236   **Parasite staining**

237   The unfolded worms were stained overnight with Semichon's acid-carmine, then de-stained with

238   1% hydrochloric acid for 5 min. The stained specimens were dehydrated in an ethanol dilution

239   series of 70%, 80%, 90% (twice), 95% (twice), and finally in absolute ethanol (twice). Each

240 dehydration step took 30 min. The stained specimens were performed using 1:1 xylene and

241 absolute ethanol for 30 min before being transferred to pure xylene for 30 min; by doing it twice.

242 Finally, the specimens were mounted with Canada balsam (*Waikagul & Thaenkham, 2014*).

243

244 After staining, all specimens were photographed using a Nikon DS-Ri1 SIGHT digital camera

245 connected to a Nikon AZ 100 M stereomicroscope (Nikon Corp., Tokyo, Japan), and the scale was

246 indicated on each photograph. An initial classification was performed based on a visual

247 examination of the specimen's external morphology. The results of these identifications were

248 compared with the outcomes of the PCR-RFLP molecular identification using ITS1 and ITS2 as

249 the molecular markers (Table 1). After that, ImageJ software (freely available at

250 https://imagej.nih.gov/ij/) was used to measure the 20 morphological characters used in traditional

251 morphometric analyses (Table 2).

252

253 **Data processing**

254 Among the 74 and 16 genetically identified *F. gigantica* and *Fasciola* intermediates, respectively,

255 we selected the less damaged specimens that constituted the best samples for the morphometric

256 studies: 17 *F. gigantica* and 10 *Fasciola* intermediate form specimens (from this study). In

257 addition, nine *F. hepatica* (from Liverpool School of Tropical Medicine, UK) were included in the

258 material.

259

260 Four types of morphometric data were separately analyzed: (i) log-transformed linear

261 measurements (LN), (ii) log-shape ratios (LSR), (iii) outlines data (OD), and (iv) landmarks data

262 (LD). For the artificial neural network-based classification (ANN, see below), we also tested the

263 combination of LN with OD.

264

265 **Software**

266 ImageJ image analysis software was used to measure linear dimensions. Landmarks and

267 pseudolandmarks (outlines) were digitized using the CLIC package (https://www.xyom-clic.eu).

268 Subsequent analyses and graphical outputs were performed using both the CLIC package and the

269 recently available online XYOM software (https://xyom.io) (*Dujardin & Dujardin, 2019*). The

270 latter allowed some specific analyses to be performed, including the maximum likelihood (*MxL*)

271 and the ANN methods of classification (see below).

272

273 **Traditional morphometrics analyses**

274 Twenty characters were measured using a computerized image analysis system, as described by

275 Valero and colleagues (*Valero, Marcos & Mas-Coma, 1996; Valero, Panova & Mas-Coma, 2005*).

276 We compared graphically our data with previously published ones, showing the departures of

277 observed means (between the published means and ours) from the total means (Fig. S1).

278

279 Univariate morphometric comparisons used non-parametric techniques based on 1000 random

280 permutations of the data between groups. The linear measurements (LN) were log-transformed for

281 the multivariate analyses and validated classification. In addition, the eigenvectors of PC1 and PC2

282 were used to extract the variables that made the largest contribution (>25%), resulting in a total of

283 12 variables instead of 20. These were BL, BW, BP, CL, Vsmin, A-VS, OS-VS, VS-Vit, Vit-P,

284 VS-P, TW, and TP (see Table 2). Log-linear measurements were then subjected to the Darroch &

285 Mosimann (1985) procedure to extract an estimate of global size (log-size) and an estimate of

286   shape variables (log-shape ratios, LSR). The influence of size on interspecific discrimination by

287   LSR was estimated by regressing on log-size the discriminant factors.

288

289   **Outline data (OD)**

290   The contour of each specimen was manually digitized (see Fig. 1A), and the resulting coordinates

291   were subjected to an elliptic Fourier analysis. The "outline data" (OD) used as the input for

292   validated classification were the normalized elliptic Fourier coefficients (NEF). Their numbers

293   were restricted to an average set giving a power of at least 99% (n = 30). Their first eight principal

294   components were used directly as input for the validated classification tests. An estimate of the

295   allometric residue contained in shape variables was performed by regressing the principal

296   components of the NEF on the square root of the area delimited by the contours. The non-

297   parametric estimation of the statistical significance of this multivariate regression followed the

298   procedure of Good (2000). On the other hand, the influence of the global size on interspecific

299   discrimination itself was estimated after regressing the first discriminant factor on size. For the

300   artificial neural network technique (ANN, see below), the OD were also combined with different

301   numbers (4, 6, 8, 10, and 12) of LN.

302

303   **Landmarks data (LD)**

304   The following anatomical landmarks (LM) were digitized: LM1 = left-side of the cephalic cone,

305   LM2 = left-side of the oral sucker, LM3 = right-side of the oral sucker, LM4 = right-side of the

306   cephalic cone, and LM5 = end of the testis (see Fig. 1B). The five landmarks were subjected to a

307   generalized Procrustes analysis, and the resulting aligned individuals were orthogonally projected

308   onto the Euclidean space tangent to the consensus of forms (Rohlf & Slice, 1990). The principal

309   components (PC) of these data (called "Procrustes residuals", "tangent space variables", or

310   "orthogonal projections") were referred to here as "landmarks data" (LD) and used as the input for

311   validated classification. The influence of size on interspecific discrimination by LD was estimated

312   by regressing the discriminant factors on the centroid size.

313

314   **Measurement error due to imaging and digitizing**

315   The precision of our digitization was evaluated for both landmarks and outlines. For size we used

316   the method recommended by Arnqvist & Mårtensson (1998), and for shape we used the Procrustes

317   analysis of variance (Procrustes ANOVA) (*Klingenberg and McIntyre, 1998*).

318

319   **Validated classification**

320   To assess how our results could be generalized to an independent data set, each individual was

321   iteratively removed and its assignation computed following the analysis of the remaining data (the

322   "leave-one-out" method, also known as "validated reclassification") (see Supplementary data 1).

323   The analyses of the remaining data used three different statistical tests: a distance-based (*Dist*)

324   method, a maximum likelihood (*MxL*) method, and an ANN method.

325

326   In the first method (*Dist*), the Mahalanobis distance was computed between the testing set (one

327   individual) and the average of each training group (the remaining data), and the tested individual

328   was assigned to the closest group. In the second method (*MxL*), the testing set (one individual) was

329   assigned to the group it was thought most likely to belong to assuming a normal distribution and

330   independence between characters (*Polly & Head, 2004; Dujardin et al., 2017*). For the *Dist* and

331   *MxL* methods we tried to avoid using a number of input variables that would exceed the number

332   of individuals in the smallest sample. There are various methods to deal with multidimensionality

333   and we selected the simplest one: the set of first PCs lower than the number of individuals in the

334   smallest sample. Note that for the analyses related to the use of five anatomical landmarks, the

335   total number of final shape variables was actually six (two-times the number of LM minus 4 in the

336   case of two-dimensional analyses).

337

338   The ANN method made use of a simple multilayer perceptron with a back propagation algorithm

339   (*Rumelhart, Hinton & Wiliams, 1988*), a variant of the original perceptron model (Rosenblatt,

340   1958) that is not, however, restricted to linear models. The method has been recently applied to

341   morphometric data, including both OD (*Mancuso, 1999*) and LD (*Dobigny, Baylac & Denys,*

342   *2002; Baylac, Villemant & Simbolotti, 2003; Van Bocxlaer & Schultheiß, 2010; Lorenz, Ferraudo*

343   *& Suesdek, 2015; Soda, Slice & Naylor, 2017*). We used the multilayer perceptron program written

344   in JavaScript, which is available at https://www.npmjs.com/package/mlp.

345

346   Following a process of trial-and-error, a single (hidden) layer composed of three neurons provided

347   the best results, irrespective of the type of input data. We used as the input the total number of

348   variables instead of a subset of their PCs. Inputs comprised the morphometric variables (LN, LSR,

349   OD, or LD, see above), while the final outputs were the probabilities of inclusion in each group

350   (in a Bayesian sense). Forms were then classified to the species for which their probability of

351   inclusion was the greatest (*Soda et al., 2017*). The "weights", i.e., the coefficients allowing the

352   outputs to be computed, were applied to the individuals being tested to compute categories and

353   suggest a classification. Thus, each individual was tentatively assigned to its putative species after

354   computing the weights from the remaining data. For a single individual to identify, the remaining

355   data were subjected to a large number of iterations to reduce the classification error to a minimum

356   (less than 25%). All specimens were separately classified 30 times, then an average classification

357   and its standard error were computed (see Supplementary data 1).

358

359   To make the most of the less stringent statistical assumptions of the ANN method with regard to

360   the statistical distribution of the input data (*Ripley, 1996*), we explored its capacity to deal with

361   different metrics combined into a single matrix. Thus, we built a matrix with four LN and made

362   another (normalized) matrix combining them with the OD. We performed a validated

363   reclassification on each matrix and then compared the scores with the correct identification. We

364   repeated this comparison successively using 6, 8, 10, and finally 12 LN, combined with the same

365   set of OD.

366

367   **RESULTS**

368   **External morphological classification**

369   The rough identification of our 90 field specimens did not detect the presence of *Fasciola*

370   intermediate forms despite them being present, as determined by the molecular analyses (see next

371   paragraph). Moreover, the rough identification wrongly assigned 30% of the specimens as *F.*

372   *hepatica* species, despite the complete absence of this species, as revealed by the molecular data

373   (Table 1).

374

375   **Molecular identification**

376   The PCR amplicons of the ITS1 and ITS2 regions (680 and 550 bp) from 90 adult worms of

377   *Fasciola* spp. were digested by *Rsa*I and *Nla*III, respectively, and provided species-specific band

378   patterns that enable the three *Fasciola* species to be distinguished (Fig. S2, S3). The results showed

379   that 74 worms were classified as *F. gigantica* and 16 worms were *Fasciola* intermediate forms. *F.*

380   *hepatica* was not present in the samples collected in the field in Thailand (Table 1).

381

382   **Traditional morphometrics (univariate analyses)**

383   Six out of 20 characters showed significant differences between the three species: CL, CW,

384   VSmax, VSmin, A-VS, and OS-VS (Table 2). Between *Fasciola* intermediate forms and the other

385   two species, BL, BP, OSmax, VS-Vit, VS-P, PhL, PhW, and TL were statistically significantly

386   different. The remaining variables appeared to be similar among the three species (Table 2).

387

388   **Traditional morphometrics (multivariate analyses)**

389   Based on the complete set of 20 linear measurements, the validated classification produced

390   excellent scores of correct species attribution, ranging from 89% to 92% according to the

391   reclassification method (Table 3, column "20LN"). In order to tentatively reduce the number of

392   characters, we selected a subset of contributing variables. We did not select them according to

393   univariate statistical differences between groups but according to their contribution to the first and

394   second principal components. This selection allowed us to restrict the set of LN from 20 to 12

395   variables (BL, BW, BP, CL, Vsmin, A-VS, OS-VS, VS-Vit, Vit-P, VS-P, TL, and TW). The

396   exclusion of poorly contributing variables was beneficial and substantially improved the validated

397   reclassification score from 88% to 97% (Table 3, column "12LN"). The shape variables derived

398   from LN (either LSR from 20LN or LSR from 12LN) produced very similar scores as log-

399   transformed raw variables (Table 3).

400

401 **Allometry**

402 The influence of size on LSR-based discrimination was high: 61% on the first discriminant factor

403 (DF1) and 63% on the second (DF2).

404

405 **Repeatability**

406 The precision of the global size estimation was almost perfect (99%) when using both the landmark

407 and outline techniques. For shape, the precision was slightly lower when using the landmarks

408 technique (93%) compared with the outlines technique (98%).

409

410 **Global size**

411 The global size of the parasites, either the body perimeter as computed from OD, the square-root

412 area within the contour (Fig. 2A; Table 4a) or the centroid size as computed from LD (Fig. 2B,

413 Table 4b) differed significantly between the intermediate forms and the other two species. These

414 other two species, *F. hepatica* and *F. gigantica*, did not show statistically significant differences

415 in global sizes (Table 5). *Fasciola* intermediate forms had the smallest global size, both in terms

416 of centroid size and the perimeter or square-root area of the contour.

417

418 **Outline-based analyses**

419 **Shape variation among *Fasciola* spp. according to the outline-based analysis**

420 Direct visualization, without amplification, of contour differences suggested *Fasciola* intermediate

421 forms possess a more globular contour (Fig. 3A). The factor maps of the discriminant analyses

422 (Fig. 3B) showed non-overlapping distributions. Accordingly, satisfactory validated classification

423 scores were obtained, ranging from 89% to 94% (Table 3, column "OD").

424 Allometry

425 The contribution of global size variation to the first shape-derived discriminant factor derived from

426 OD was 23%, while it was 6% for the second one. According to the non-parametric method used

427 to evaluate the multivariate regression significance between size and shape, the allometric residue

428 contained in OD shape variables was significant ($p < 0.001$).

429

430 Landmark-based analyses

431 The mean specimens aligned according to the five anatomical landmarks used here showed some

432 separation between the three forms, with a seemingly more critical difference for *Fasciola*

433 intermediate forms (Fig. 4A), as for the OD (Fig. 3). The factor map of the discriminant analysis

434 showed that *Fasciola* intermediate forms could diverge from *F. gigantica* and *F. hepatica* (Fig.

435 4B). However, some overlapping was observed, and the validated classification scores based on

436 landmark data were not satisfactory, ranging from 58% to a maximum of 69% (Table 3, column

437 "LD").

438

439 Statistical significance of shape variation

440 The pairwise Mahalanobis distances based on shape, for both landmark- and outline-based shapes,

441 were highly significant ($p < 0.001$) between *Fasciola* intermediate forms and *F. gigantica* or *F.*

442 *hepatica*. Between *F. gigantica* and *F. hepatica*, the distance was smaller, albeit almost significant

443 ($p \sim 0.06$) (Table 6).

444

445 Allometry

446   The first and second discriminant factors derived from the Procrustes residuals were still under the

447   influence of size (42% and 4%, respectively) after regression on centroid size, which was more

448   than for the outline-based analysis (23% and 6%, respectively) but much less than for the

449   traditional method (61% and 63%, respectively).

450

451   **Combination of log-transformed linear measurements and OD**

452   The ANN performance was significantly improved when traditional linear measurements (LN),

453   even a small number of them, were combined with outline shape descriptors (OD). This

454   performance could be appreciated by comparing the ANN accuracy when applied to these data

455   either separately or combined (Table 7). For instance, with ten traditional measurements as the

456   input, the ANN performance achieved 88% of correct identification, with OD only it reached 89%

457   (Table 3), whereas by combining both types of data the total score jumped to 94% (Table 7).

458

459   **DISCUSSION**

460   The specimens in this study were processed using the same treatment method. If some deformation

461   occurred as a result of this treatment, it is reasonable to expect that this would be the same for each

462   species. The average values of linear measurements of 12 selected morphological characters of

463   each *Fasciola* sp. were graphically compared with previously published values from the same

464   parasites and the same hosts (see Fig. S1). This comparison showed that our samples were in the

465   range of average values and suggested that the small number of specimens did not result in any

466   sample bias.

467

468 To reduce the effects of growth on the phenotype, only fully mature stages were selected from

469 each molecularly defined group. However, in addition to possible growth heterogeneity, other

470 divergent influences could have affected phenotypic variations in our total sample, such as

471 different geographic origins, microhabitats, host, and seasonality. We considered that, within each

472 species or group, these environmental effects would probably have had much more of an impact

473 on size than on shape. We used the geometric morphometrics approach to tentatively set apart

474 these size effects from the morphometric data, extracting the remaining shape variables and using

475 them exclusively to describe the groups distinguished by molecular data. The allometric residue

476 that possibly remained in the shape variables was not removed because our comparisons concerned

477 species rather than intraspecific groups. Species may differ by their allometric growth, which

478 becomes part of their evolutionary divergence (*Klingenberg, 2016*).

479

480 Our tentative removal of the size effect from the linear measurements (either 20 LN or 12 LN) did

481 not change the global classification scores as obtained by raw measurements, probably because it

482 could not perform a consistent removal of size, as indicated by our allometric analysis. Size

483 remained dominant in the LSR (~61%) relative to the PR (~41%) and OD (~24%). Another

484 inconvenience of LSR is that they do not allow a visual representation of shape changes between

485 taxa (*Rohlf & Marcus, 1993*). Methods based on coordinates data allow such visual representation

486 (see Fig. 2D).

487

488 Our samples represented limited geographic isolates and, because of the low sample sizes, could

489 suffer from low representativity. Moreover, the number of variables could be greater than the

490 number of individuals in our smallest sample (nine *F. hepatica*). We followed the "rule" for a

491   multivariate analysis, especially discriminant analysis, which is to use a number of variables that

492   does not exceed the number of individuals in the smallest sample. There are various methods

493   available to deal with multidimensionality. This species number of PC scores was selected so as

494   to use a number of variables that does not exceed the number of individuals in the smallest sample.

495   This procedure allowed multidimensionality to be reduced where necessary; this did not make our

496   samples more representative, but at least the information they provided was not over-interpreted.

497

498   Since linear measurements can provide information about both size and shape, and because of the

499   relatively small size of *Fasciola* intermediate forms (Fig. 2), it was expected that this sample could

500   be identified using traditional techniques. Nevertheless, linear measurements did not just reflect

501   global size variation since they could also allow an excellent score of recognition for the pair *F.*

502   *hepatica* and *F. gigantica*, which are parasites of similar global size (Fig. 2, Table 4). The

503   extraction of shape (LSR) did not produce different results, confirming the relatively poor

504   efficiency of shape removal from our linear measurements. With or without shape extraction,

505   traditional morphometrics appeared to be a powerful means of species determination. It does,

506   however, require an excellent knowledge of the parasite's internal anatomy.

507

508   The previously recommended number of linear measurements necessary to distinguish *F. hepatica*

509   from *F. gigantica* and *Fasciola* intermediate forms was as high as 20 (*Valero, Panova & Mas-*

510   *Coma, 2005; Ashrafi et al., 2006*). By considering the contribution of each variable to their first

511   two PC, we were able to reduce this number from 20 to 12, with similar or better results (see Table

512   3). This improvement could be explained by the elimination of redundancy among measurements,

513   and a possible lack of precision for some of them. Either way, our examination of repeatability

514   produced satisfactory results for both shape (93%) and size (99%), although these results were

515   lower than the results of the outline-based technique (98% and 99%, respectively).

516

517   The outline-based morphometrics analysis was based on manual contour digitization. Although

518   this type of manual data collection is relatively fast, an automated capture technique might be

519   preferable. Automated curve tracing is already included in many types of imaging software, and it

520   may be more appropriate for generalizability (*Yang et al., 2015*), but its useful application is

521   dependent on the complexity of the contours (*Sheets at al., 2006; Firmat et al., 2010*). We would

522   recommend that a study be performed that is specifically aimed at comparing automatic and

523   manual capture techniques. The fact remains, however, that whether manual or automatic

524   techniques are used, the outline digitization process is a relatively easy task and does not require

525   highly specialized skills from the user.

526

527   The landmark-based approach required a little more knowledge about the internal anatomy of the

528   parasites and did not provide satisfactory results. The lower reclassification scores of the landmark

529   data relative to the OD should not be regarded as a rule. In our case, very few reliable LM could

530   be suggested, reducing the amplitude of shape capture, which could explain the unsatisfactory

531   scores obtained. It is possible that better results might be obtained with, for instance, a different

532   landmark configuration, or a different number of landmarks.

533

534   In our validated method of reclassification each individual was sequentially removed from the total

535   sample and its identification performed using the model computed without it, a method also called

536   a "leave-one-out" or "jack-knife" procedure (*Manly, 1986*). Thus, individual identification was

537   repeated as many times as there were individuals in the total sample (see Supplementary data 1).

538   Our "accuracy" for each method (Tables 3 and 7) was the percentage of individuals that were

539   correctly identified at the end of each such procedure. For the ANN classification, each

540   classification session was repeated 30 times, and the accuracies were presented as the average and

541   standard deviation of the 30 sessions (Tables 3 and 7).

542

543   The validated classification applied to different kinds of data used different methods (Table 3).

544   The one based on the Mahalanobis distance (*Dist*) generally produced higher scores with more

545   correct identifications than the other classification techniques, which conforms with previous

546   studies and highlights the frequent relative superiority of this technique (*Dujardin et al., 2017;*

547   *MacLoad et al., 2007; Sonnenschein et al., 2015*). However, reliable classification and assessment

548   of group separation using this analysis is based on several assumptions, some of which frequently

549   do not coincide with morphometric data (*Mitteroecker & Bookstein, 2011*). For this reason, we

550   added additional classification techniques, such as the simple maximum likelihood (*MxL*)

551   approach (*Polly & Head, 2004; Dujardin et al., 2017*) and the ANN technique.

552

553   Since the *MxL* method assumes normally distributed data and independence among the characters

554   used, our best results were obtained when applied to (a subset of) the PC of the log-transformed

555   input variables, rather than when applied to the raw variables (detailed comparisons not shown).

556   In agreement with previous comparisons (*Dujardin et al., 2017*), the *MxL* technique provided a

557   slightly lower accuracy than the *Dist* method.

558

559 In our study, the ANN produced slightly lower scores than the *Dist* or the *MxL* approaches. Our

560 idea was to compare the taxonomic signal embedded in different types of data rather than to deeply

561 explore the power of each classification method. So, for each classification method, we tried to

562 subject each type of data to the same process. The neural network was thus configured in the same

563 way, irrespective of which data were used as the input (see Supplementary data 2). Certainly, some

564 improvements could be obtained by refining the configuration according to the kind of data being

565 used. Moreover, a possible explanation for the relative weakness of the ANN method used here is

566 the low sample sizes of the training data. The ANN method may require many data to achieve an

567 accurate level of performance (*Soda, Slice & Naylor, 2017*). However, the general scores were

568 satisfactory, and we showed that this technique could greatly benefit from using combined data as

569 the input, for example using a few traditional measurements together with outline shape data

570 (Table 7).

571

572 **CONCLUSION**

573 Both traditional and modern morphometric approaches can be helpful in the morphological

574 identification of *Fasciola* sp. We showed that by reducing the number of linear measurements

575 from the 20 commonly recommended ones to a selected subset of 12 measurements, the accuracy

576 of identification could be improved.

577

578 In our study, the outline-based approach produced satisfactory classification scores, more so than

579 the landmark-based approach. Moreover, we showed for the first time the apparent differences in

580 shape between "intermediate forms" of *Fasciola* sp. and *F. gigantica* and *F. hepatica*. The relative

581 simplicity of data collection (even manual collection) makes the outline-based approach an

582 attractive alternative.

583   By comparing various data collection and classification techniques, our study suggests promising

584   new strategies to follow in helping with the identification of species in the genus *Fasciola* based

585   on their morphology.

586

587   **ACKNOWLEDGEMENTS**

588   The authors would like to thank the Center Equipment Unit, Faculty of Tropical Medicine,

589   Mahidol University for supporting the excellent imaging system. We would like to thank Mr. Ulf

590   Biallas, Software Engineer, Axel Springer Ideas Engineering GmbH for his kind assistance.

591

592   **ADDITIONAL INFORMATION AND DECLARATION**

593   **Funding**

594   The work was financially supported for sample collection from the Department of Helminthology,

595   Faculty of Tropical Medicine, Mahidol University.

596   **Competing Interest**

597   The authors declare that they have no competing interests.

598   **Author Contributions**

599   ● Suchada Sumruayphol analyzed the data using GM analyses and prepared the manuscript.

600   ● Praphaipat Siribat photographed *Fasciola* spp., performed molecular works and prepared

601   the manuscript.

602   ● Jean-Pierre Dujardin and Sebastien Dujardin analyzed the data using TM, GM and ANN

603   analyses and reviewed the final version of the manuscript.

604   ● Chalit Komalamisra and Urusa Thaenkham supervised this study and reviewed the final

605   version of the manuscript.

606   All authors read and approved the final manuscript.

607 **Data Availability**

608 The following information was supplied regarding data availability:

609   The raw data are provided as a Supplemental File

610 **Supplemental Information**

611 Supplemental information for this article can be found online at xxxxxxx

612

613 **REFERENCES**

614 **Afshan K, Valero M, Qayyum M, Peixoto R, Magraner A, Mas-Coma S. 2014.** Phenotypes of

615   intermediate forms of *Fasciola hepatica* and *F. gigantica* in buffaloes from Central Punjab,

616   Pakistan. *Journal of Helminthology* **88(4)**:417–426. DOI 10.1007/s11692-011-9109-8.

617 **Amer S, Dar Y, Ichikawa M, Fukuda Y, Tada, C, Itagaki T, Nakai Y. 2011.** Identification of

618   *Fasciola* species isolated from Egypt based on sequence analysis of genomic (ITS1 and

619   ITS2) and mitochondrial (NDI and COI) gene markers. *Parasitology International* **60(1)**:5-

620   12. DOI 10.1016/j.parint.2010.09.003.

621 **Arnqvist G., Mårtensson T. 1998.** Measurement error in geometric morphometrics: empirical

622   strategies to assess and reduce its impact on measure of shape. *Acta Zoologica Academiae*

623   *Scientiarum Hungaricae* **44 (1–2)**:73–96.

624 **Aroonroch R, Worawichawong S, Nitiyanant P, Kanchanapitak A, Bunyaratvej S. 2006.**

625   Hepatic fascioliasis due to *Fasciola hepatica*: a two-case report. *Journal of the Medical*

626   *Association of Thailand* **89(10)**:1770–1774.

627 **Ashrafi K, Valero MA, Panova M, Periago MV, Massoud J, Mas-Coma S. 2006.** Phenotypic

628   analysis of adults of *Fasciola hepatica*, *Fasciola gigantica* and intermediate forms from the

629   endemic region of Gilan, Iran. *Parasitology International* **55(4)**:249–260. DOI

630   10.1016/j.parint.2006.06.003.

631   **Baylac M, Villemant C, Simbolotti G. 2003.** Combining geometric morphometrics with pattern

632   recognition for the investigation of species complexes. *Biological Journal of the Linnean*

633   *Society* **80(1)**:89–98. DOI 10.1046/j.1095-8312.2003.00221.x.

634   **Buranasin P, Harinasuta T. 1970.** A case of fascioliasis in Thailand. *The Southeast Asian*

635   *Journal of Tropical Medicine and Public Health* **1**:146–147.

636   **Burnaby TP. 1966.** Growth invariant discriminant functions and generalized distances.

637   *Biometrics* **22**:96-110.

638   **Chaichanasak P, Ichikawa M, Sobhon P, Itagaki T. 2012.** Identification of *Fasciola* flukes in

639   Thailand based on their spermatogenesis and nuclear ribosomal DNA, and their intraspecific

640   relationships based on mitochondrial DNA. *Parasitology International* **61**:545–549. DOI

641   10.1016/j.parint.2012.03.009.

642   **Darroch JN, Mosimann JE. 1985.** Canonical and principal components of shape. *Biometrika*

643   **72(2)**:241–252. DOI 10.1093/biomet/72.2.241.

644   **Dobigny G, Baylac M, Denys C. 2002.** Geometric morphometrics, neural networks and diagnosis

645   of sibling *Taterillus* species (Rodentia, Gerbillinae). *Biological Journal of the Linnean*

646   *Society* **77(3)**: 319–327. DOI 10.1046/j.1095-8312.2002.00074.x.

647   **Dujardin JP, Kaba D, Solano P, Dupraz M, McCoy KD, Jaramillo-o N. 2014.** Outline-based

648   morphometrics, an overlooked method in arthropod studies?. *Infection, Genetics and*

649   *Evolution* **28**: 704-714. DOI 10.1016/j.meegid.2014.07.035

650   **Dujardin JP, Dujardin S, Kaba D, Santillán-Guayasamín S, Villacís AG, Piyaselakul S,**

651   **Sumruayphol S, Samung Y, Vargas RM. 2017.** The maximum likelihood identification

652 method applied to insect morphometric data. *Zoological Systematics* **42(1)**:46–58. DOI

653 10.11865/zs.201704.

654 **Dujardin JP. 2019.** A template-dependent semilandmarks treatment and its use in medical

655 entomology. *Infection, Genetics and Evolution* **70**:197–207. DOI 10.1016/j.meegid.2019.

656 03.002

657 **Dujardin S, Dujardin JP. 2019.** Geometric morphometrics in the cloud. *Infection, Genetics and*

658 *Evolution* **70**:189-196. DOI 10.1016/j.meegid.2019.02.018. Available at *https://xyom.io*.

659 **Firmat C, Gomes-Rodrigues H, Renaud S, Claude J, Hutterer R, Garcia-Talavera F,**

660 **Michaux J. 2010.** Mandible morphology, dental microwear, and diet of the extinct giant rats

661 Canariomys (Rodentia: Murinae) of the Canary Islands (Spain). *Biological Journal of the*

662 *Linnean Society* **101(1)**:28–40. DOI 10.1111/j.1095-8312.2010.01488.x.

663 **Good P. 2000**. Permutation tests: a practical guide to resampling methods for testing hypotheses.

664 New York, Springer

665 **Halakou A, Khazan H, Bendehpour M, Kazemi B. 2017**. Morphological study of *Fasciola*

666 parasites isolated from cattle and sheep in Golestan Province (Iran). *Novelty in Biomedicine*

667 **5(4)**:166–171.

668 **Huang WY, He B, Wang CR, Zhu XQ. 2004**. Characterisation of *Fasciola* species from

669 Mainland China by ITS-2 ribosomal DNA sequence. *Veterinary Parasitology* **120(1-2)**: 75-

670 83. DOI 10.1016/j.vetpar.2003.12.006.

671 **Ichikawa M, Itagaki T. 2010**. Discrimination of the ITS1 types of *Fasciola* spp. based on a PCR-

672 RFLP method. *Parasitology Research* **106(3)**:757–761. DOI 10.1007/s00436-010-1724-2.

673 **Itagaki T, Ichinomiya M, Fukuda K, Fusyuku S, Carmona C. 2011**. Hybridization experiments

674 indice incomplete reproductive isolating mechanism between *Fasciola hepatica* and

675 *Fasciola gigantica*. *Parasitology* **138(10)**:1278-1284. DOI 10.1017/S0031182011000965.

676 **Iyiola OA, Shittu O, Owolodun OA, Animasaun DA, Udeze AO. 2018**. Morphometric

677 phenotypes and molecular identification of *Fasciola* species isolated from cattle in Ilorin,

678 North-Central Nigeria. *Sri Lankan Journal of Biology* **3(2)**:9–23. DOI 10.4038/sljb.v3i2.23.

679 **Kanoksil W, Wattanatranon D, Wilasrusmee C, Mingphruedh S, Bunyaratvej S. 2006**.

680 Endoscopic removal of one live biliary *Fasciola gigantica*. *Journal of the Medical*

681 *Association of Thailand* **89(12)**:2150–2154.

682 **Klingenberg CP, McIntyre GS. 1998.** Geometric morphometrics of developmental instability:

683 analyzing patterns of fluctuating asymmetry with Procrustes methods. *Evolution* **52**: 1363–

684 1375.

685 **Klingenberg CP. 2016**. Size, shape, and form: concepts of allometry in geometric morphometrics.

686 *Development genes and evolution* **226(3)**:113-137. DOI 10.1007/s00427-016-0539-2.

687 **Kuhl FP, Giardina CR. 1982.** Elliptic Fourier features of a closed contour. *Computer graphics*

688 *and image processing* **18(3**):236-258. DOI 10.1016/0146-664X (82)90034-X.

689 **Lin RQ, Dong SJ, Nie K, Wang CR, Song HQ, Li AX, Huang WY, Zhu XQ. 2007**. Sequence

690 analysis of the first internal transcribed spacer of rDNA supports the existence of the

691 intermediate *Fasciola* between *F. hepatica* and *F. gigantica* in mainland China. *Parasitology*

692 *Research* **101(3)**:813-817. DOI 10.1007/s00436-007-0512-0

693 **Lorenz C, Ferraudo AS, Suesdek L. 2015**. Artificial Neural Network applied as a methodology

694 of mosquito species identification. *Acta Tropica* **152**:165–169. DOI

695 10.1016/j.actatropica.2015.09.011.

696 **Lotfy WM, Brant SV, DeJong RJ, Le TH, Demiaszkiewicz A, Rajapakse RJ, Perera VB,**

697   **Laursen JR, Loker ES. 2008**. Evolutionary origins, diversification, and biogeography of

698   liver flukes (Digenea, Fasciolidae). *The American Journal of Tropical Medicine and*

699   *Hygiene* **79(2)**:248-255. DOI 10.4269/ajtmh.2008.79.248.

700 **Mancuso S. 1999**. Elliptic Fourier Analysis (EFA) and Artificial Neural Networks (*ANN*s) for the

701   identification of grapevine (*Vitis vinifera* L.) genotypes. *Vitis* **38(2)**:73-78.

702 **Mas-Coma S, Bargues MD, Valero MA. 2005**. Fascioliasis and other plant-borne trematode

703   zoonoses. *International Journal Parasitology* **35(11-12)**:1255–1278. DOI

704   10.1016/j.ijpara.2005.07.010

705 **Mas-Coma S, Valero MA, Bargues MD. 2008**. Effects of climate change on animal and zoonotic

706   helminthiases. *Revue Scientifique et Technique* **27(2)**:443–457.

707 **Mas-Coma S, Valero MA, Bargues MD. 2009a**. Climate change effects on Trematodiases, with

708   emphasis on zoonotic fascioliasis and schistosomiasis. *Veterinary Parasitology* **163(4)**:264–

709   280. DOI 10.1016/j.vetpar.2009.03.024.

710 **Mas-Coma S, Valero MA, Bargues MD. 2009b**. *Fasciola*, lymnaeids and human fascioliasis,

711   with a global overview on disease transmission, epidemiology, evolutionary genetics,

712   molecular epidemiology and control. *Advances in Parasitology* **69**:41-146. DOI

713   10.1016/S0065-308X(09)69002-3.

714 **Mas-Coma S, Bargues MD, Valero MA. 2018**. Human fascioliasis infection sources, their

715   diversity, incidence factors, analytical methods and prevention measures. *Parasitology*

716   145(13):1665-1699. DOI 10.1017/S0031182018000914

717   **Mitteroecker P, Bookstein F. 2011**. Linear Discrimination, Ordination, and the Visualization of

718   Selection Gradients in Modern Morphometrics. *Evolutionary Biology* **38(1)**:100–114. DOI

719   10.1007/s11692-011-9109-8

720   **Periago MV, Valero MA, Panova M, Mas-Coma S. 2006**. Phenotypic comparison of allopatric

721   populations of *Fasciola hepatica* and *Fasciola gigantica* from European and African bovines

722   using a computer image analysis system (CIAS). *Parasitology Research* **99(4)**:368-378. DOI

723   10.1007/s00436-006-0174-3

724   **Periago MV, Valero MA, Sayed ME, Ashrafi K, Wakeel AE, Mohamed MY, Desquesnes M,**

725   **Curtale F, Mas-Coma S. 2008**. First phenotypic description of *Fasciola hepatica*/ *Fasciola*

726   *gigantica* intermediate forms from the human endemic area of the Nile Delta, Egypt.

727   *Infection, Genetics and Evolution* **8(1)**:51–58. DOI 10.1016/j.meegid.2007.10.001

728   **Polly PD, Head JJ. 2004**. Maximum-likelihood identification of fossils: taxonomic identification

729   of Quaternary marmots (Rodentia, Mammalia) and identification of vertebral position in the

730   pipesnake *Cylindrophis* (Serpentes, Reptilia). In: Elewa AMT, editor. *Morphometrics,*

731   *Applications in Biology and Paleontology*. Springer Science & Business Media, 197–221.

732   **Ripley BD. 1996**. Pattern recognition and neural networks. Cambridge: Cambridge University

733   Press.

734   **Rohlf FJ, Marcus LF. 1993**. A revolution in morphometrics. *Trends in Ecology & Evolution*

735   **8(4)**:129–132. DOI 10.1016/0169-5347(93)90024-J.

736   **Rohlf FJ, Slice D. 1990**. Extensions of the procrustes method for the optimal superimposition of

737   landmarks. *Systematics Zoology* **39(1)**:40–59. DOI 10.2307/2992207.

738   **Rosenblatt F. 1958**. The Perceptron: A theory of statistical separability in cognitive systems.

739   Cornell Aeronautical Laboratory, Report No. VG1196-G-1.

740 **Rumelhart DE, Hinton G E, Williams RJ. 1988**. Learning representations by back-propagating

741   errors. *Nature* **323**:533-536.

742 **Siribat P, Dekumyoy P, Komalamisra C, Sumruayphol S, Thaenkham U. 2018**. Molecular

743   identification of *Fasiola* spp. representative samples from Thailand based on PCR-RFLP.

744   *The Journal of Tropical Medicine and Parasitology* **41**:1–7.

745 **Shafiei R, Sarkari B, Sadjjadi SM, Mowlavi GR, Moshfe A. 2014**. Molecular and

746   morphological characterization of *Fasciola* spp. isolated from different host species in a

747   newly emerging focus of human fascioliasis in Iran. *Veterinary Medicine International*

748   **405740**. DOI 10.1155/2014/405740

749 **Sheets HD, Covino KM, Panasiewicz JM, Morris SR. 2006**. Comparison of geometric

750   morphometric outline methods in the discrimination of age-related differences in feather

751   shape. *Frontiers in Zoology* **3(1)**:2–15. DOI 10.1186/1742-9994-3-15.

752 **Soda KJ, Slice DE, Naylor GJP. 2017**. Artificial neural networks and geometric morphometric

753   methods as a means for classification: A case-study using teeth from Carcharhinus sp.

754   (Carcharhinidae). *Journal of Morphology* **278(1)**:131–141. DOI 10.1002/jmor.20626).

755 **Sonnenschein A, VanderZee D, Pitchers WR, Chari S, Dworkin I. 2015**. An image database

756   of *Drosophila melanogaster* wings for phenomic and biometric analysis. *Gigascience* **4**:25.

757   DOI 10.1186/s13742-015-0065-6.

758 **Srihakim S, Pholpark M. 1991**. Problem of fascioliasis in animal husbandry in Thailand. *The*

759   *Southeast Asian Journal of Tropical Medicine and Public Health* **22(Suppl)**:352–355.

760 **Srimuzipo P, Komalamisra C, Choochote W, Jitpakdi A, Vanichthanakorn, P, Keha P,**

761   **Riyong D, Sukontason K, Komalamisra N, Sukontason K. 2000**. Comparative

762   morphometry, morphology of egg and adult surface topography under light and scanning

763   electron microscopics, and metaphase karyotype among three size-races of *Fasciola*

764   *gigantica* in Thailand. *The Southeast Asian Journal of Tropical Medicine and Public Health*

765   **31(2)**:366–373.

766   **Tippawangkosol P, Saingamsook J, Kijdamrongthum P, Thinrungroj N, Jariyawat K,**

767   **Somboon P. 2017**. Endoscopic management of biliary fascioliasis and identification of

768   *Fasciola* fluke based on mitochondrial DNA: A case report from Thailand. *The Southeast*

769   *Asian Journal of Tropical Medicine and Public Health* **48(3)**:524–531.

770   **Valero MA, Marcos MD, Mas-Coma S. 1996**. Mathematical model for the ontogeny of *Fasciola*

771   *hepatica* in the definitive host. *Research & Reviews in Parasitology* **56**:13–20.

772   **Valero MA, Panova M, Mas-Coma S. 2005**. Phenotypic analysis of adults and eggs of *Fasciola*

773   *hepatica* by computer image analysis system. *Journal of Helminthology* **79(3)**:217–225. DOI

774   10.1079/JOH2005301.

775   **Van Bocxlaer B, Schultheiß R. 2010**. Comparison of morphometric techniques for shapes with

776   few homologous landmarks based on machine-learning approaches to biological

777   discrimination. *Paleobiology* **36(3)**:497-515. DOI 10.1666/08068.1.

778   **Varea Sánchez M, Bastir M, Roldan ER. 2013**. Geometric morphometrics of rodent sperm head

779   shape. *PloS One* **8(11)**: e80607. DOI 10.1371/journal.pone.0080607.

780   **Waikagul J, Thaenkham U. 2014**. Approaches to research on the systematics of fish-borne

781   trematodes. Academic Press.

782   **Wannasan A, Khositharattanakool P, Chaiwong P, Piangjai S, Uparanukraw P, Morakote**

783   **N. 2014.** Identification of *Fasciola* species based on mitochondrial and nuclear DNA reveals

784   the co-existence of intermediate *Fasciola* and *Fasciola gigantica* in Thailand. *Experimental*

785   *Parasitology* **146:** 64–70. DOI 10.1016/j.exppara.2014.09.006.

786   **Yang HP, Ma CS, Wen H, Zhan QB, Wang XL. 2015**. A tool for developing an automatic insect

787   identification system based on wing outlines. *Scientific report* **5**:12786. DOI

788   10.1038/srep12786.

789

Figure 1

Geometric morphometric analysis of *Fasciola* adult worms

(A) Contour digitized on *Fasciola* spp. body for outline-based geometric morphometrics analysis, (B) Position of 5 landmarks digitized on *Fasciola* spp. body for landmark-based geometric morphometric analysis.

A

1 cm

B

1 cm

Figure 2

Global size of *Fasciola* adult worms

(A) Quantile boxes showing body perimeter variation obtained after outline-based geometric morphometric analysis. Each box shows the group median that separates the 25th and 75th quartiles. Dots on the left represent individuals. (B) Variation of body centroid size of *Fasciola* spp. from landmark-based GM analysis. Each box shows the group median that separates the 25th and 75th quartiles. Dots on the left represent individuals.

A

B

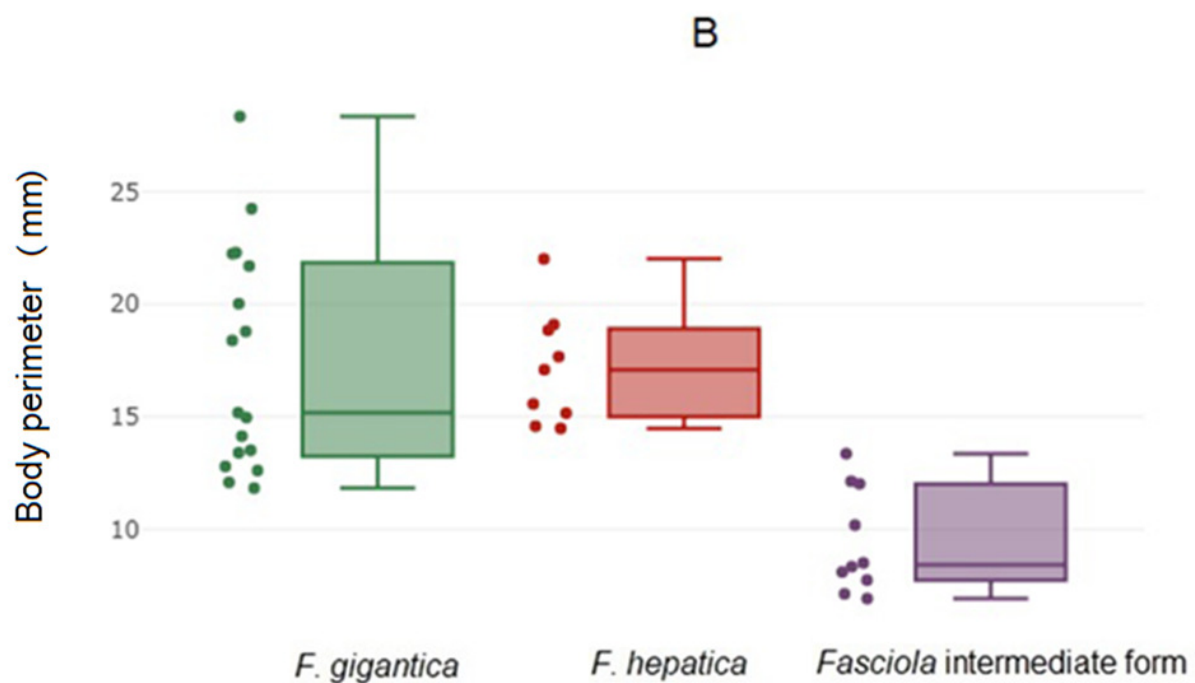
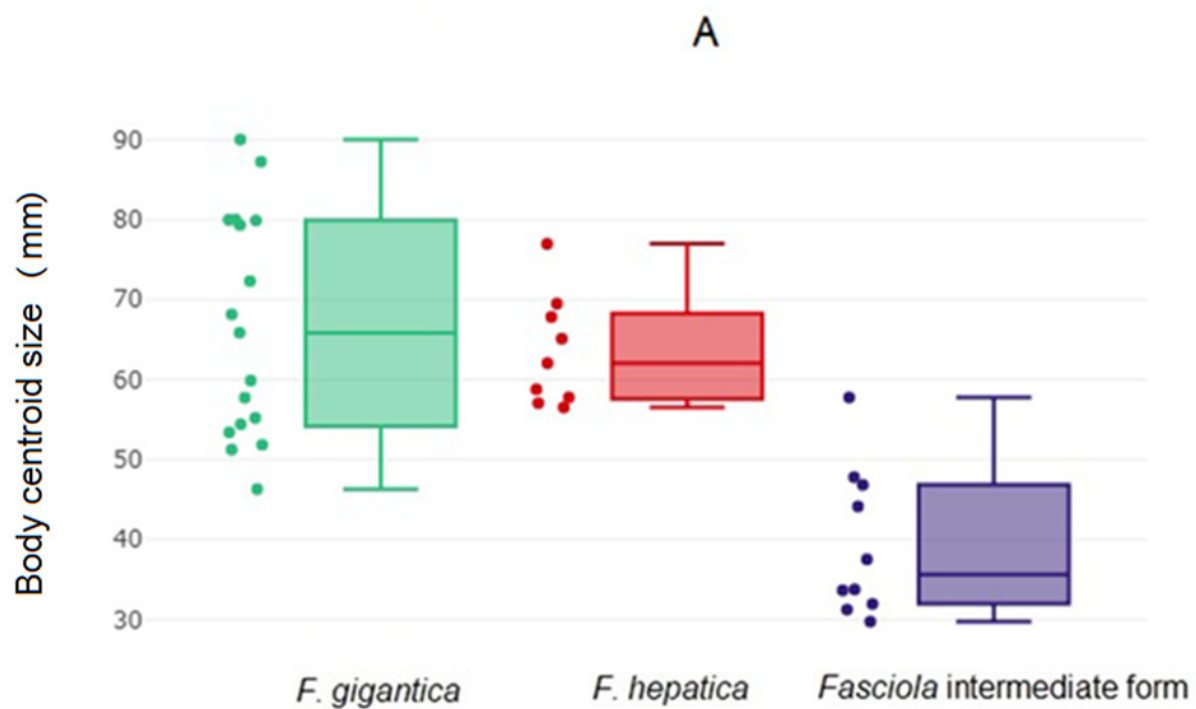



Figure 3

Shape variation of *Fasciola* spp. based on the outline-based analysis

<!--[if !supportLists]-->(A) <!--[endif]-->Mean shape of whole body contour of *F. gigantica*, *F. hepatica*, and *Fasciola* intermediate forms. (B) Outline-based discriminant analysis. Factor map of the two discriminant factor (DFs) derived from shape variables for *Fasciola* species. Each point represents an individual. The horizontal axis is the first DF; the vertical axis is the second DF; their cumulated contributions reach 100% of the total variation.

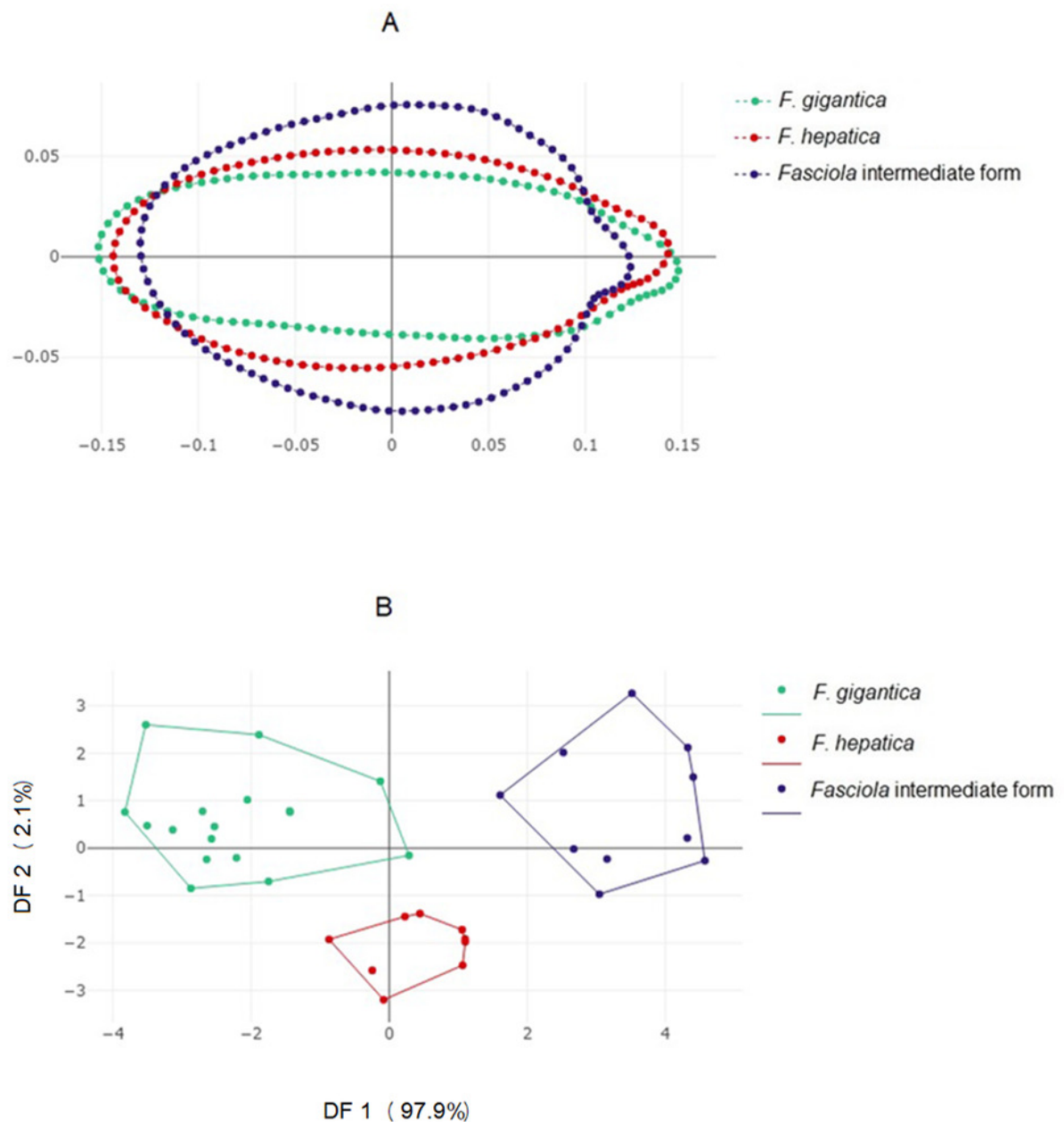
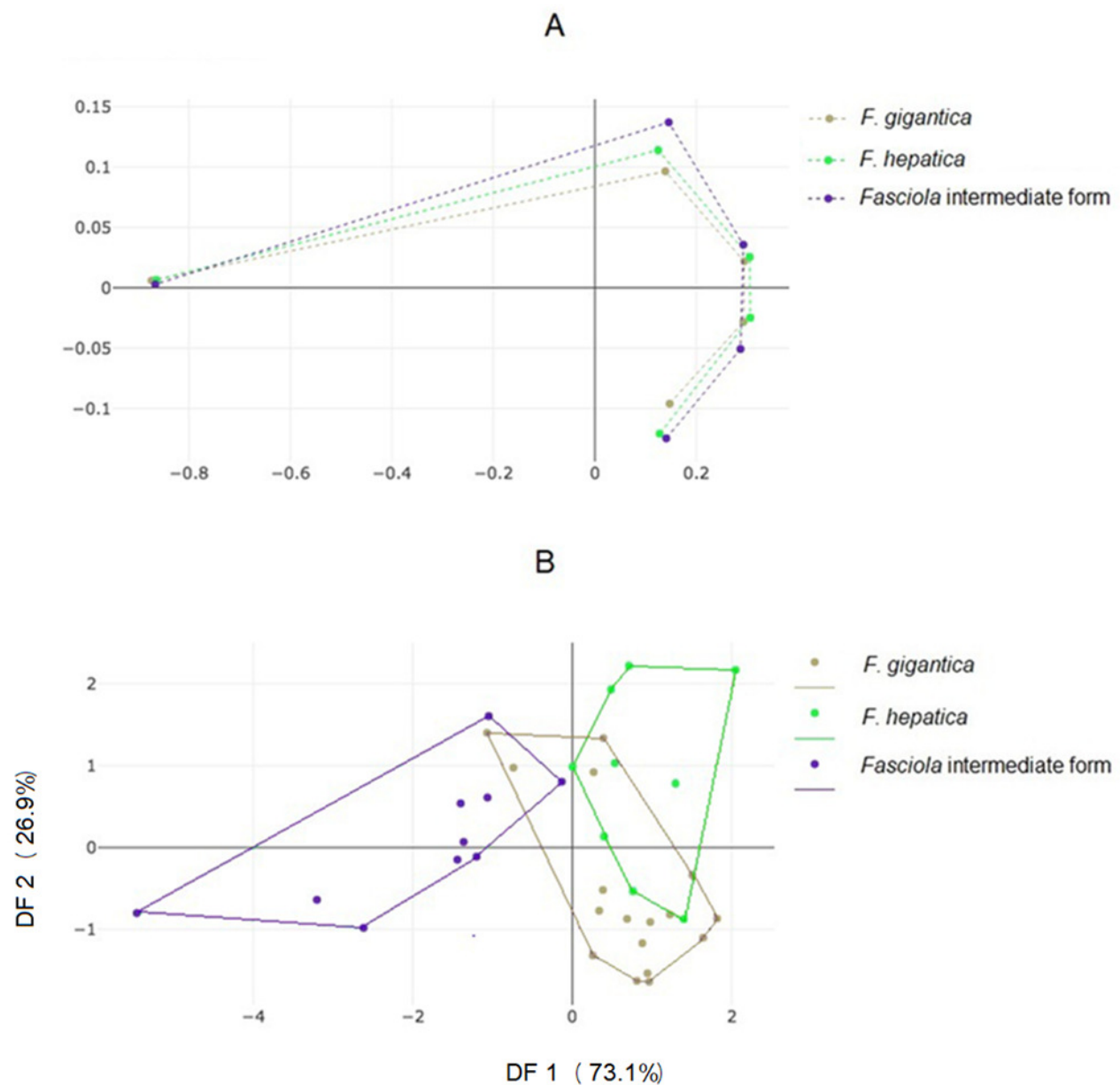



Figure 4

Shape variation of *Fasciola* spp. based on landmark-based analyses

(A) Average positions of anatomical landmarks (dots) after Procrustes superposition, representing the mean shapes of *F. gigantica*, *F. hepatica*, and of *Fasciola* intermediate forms, (B) Landmark-based discriminant analysis. Factor map of the two discriminant factors (DFs) derived from final shape variables of *F. gigantica*, *F. hepatica*, and *Fasciola* intermediate forms. Each point represents an individual. The horizontal axis is the first DF; the vertical axis is the second DF, their cumulated contributions reach 100% of the total variation.

A

B

Visual identification versus molecular typing of field specimen

The visual examination of external morphology could not recognize any of the 16 intermediate forms, most of them (14) wrongly identified as *F. hepatica*. A minor part of *F. gigantica* (13) was misidentified also with *F. hepatica*.

1 **Table 1.** Visual identification versus molecular typing of field specimen

Species	External Morphology	PCR-RFLP
F. gigantica	63	74
Fasciola intermediate forms	0	16
F. hepatica	27	0

4
5 The visual examination of external morphology could not recognize any of the 16 intermediate

6 forms, most of them (14) wrongly identified as *F. hepatica*. A minor part of *F. gigantica* (13)

7 was misidentified also with *F. hepatica*.

Traditional morphometric analyses of *Fasciola* species

Fg, *Fasciola gigantica*; *Fh*, *Fasciola hepatica*; *Fif*, *Fasciola* intermediate forms. M, mean. n, sample size; Min, minimum, Max, maximum; SD, standard deviation; BL, Body length, BL; BW, Body width, BW; BWOv, BW at ovary level; BP, Body perimeter; CL, Cone length; CW, Cone width; Osmax, Oral sucker maximum diameter; Osmin, Oral sucker minimum diameter; Vsmax, Ventral sucker maximum diameter; Vsmin, Ventral sucker minimum diameter; A-VS, Distance between anterior end of body and VS; OS-VS, Distance between suckers; VS-Vit, Distance between VS and union of vitelline glands; Vit-P, Distance between Vit and posterior end of body; VS-P, Distance between VS and posterior end of body; PhL, Pharynx length; PhW, Pharynx width; TL, Testicular space length; TW, Testicular space width; TP, Testicular space perimeter.

1 **Table 2.** Traditional morphometric analyses of *Fasciola* species

Character	Fg n = 17			Fif n = 10			Fh n = 9			P		
	M	Min-Max	SD	M	Min-Max	SD	M	Min-Max	SD	*Fg-Fif*	*Fg-Fh*	*Fif-Fh*
1. BL	29.9	20.8- 42.1	6.7	16.0	10.9-23.2	3.6	28.2	23.8-33.3	3.1	0.000	0.787	0.000
2. BW	7.4	4.8-10.9	1.9	8.6	5.9-12.8	2.2	10.2	8.2-12.9	1.5	0.141	0.001	0.107
3. BWOv	4.4	3.3-5.9	0.7	5.1	3.3-7.7	1.3	5.1	4.2-5.6	0.5	0.113	0.063	0.746
4. BP	65.4	46.3-88.2	13.6	37.5	29.2-57.5	9.0	63.3	54.6-74.8	6.6	0.000	0.900	0.000
5. CL	2.9	2.4-3.7	0.4	1.6	1.2-2.1	0.3	3.4	2.6-3.9	0.4	0.000	0.241	0.000
6. CW	3.3	2.4-4.1	0.5	2.4	1.80-3.35	0.5	4.1	3.4-4.6	0.4	0.003	0.021	0.000
7. OSmax	0.8	0.5-1.0	0.1	0.6	0.5-0.7	0.1	0.8	0.5-1.1	0.2	0.001	0.861	0.001
8. OSmin	0.5	0.3-0.7	0.1	0.4	0.3-0.6	0.1	0.6	0.3-0.8	0.2	0.399	0.089	0.023
9. VSmax	1.3	0.9-1.6	0.2	0.9	0.7-1.1	0.1	1.5	0.9-1.7	0.3	0.000	0.106	0.000
10. VSmin	1.3	1.0-1.6	0.2	0.8	0.5-0.9	0.2	1.5	1.1-1.7	0.2	0.000	0.239	0.000
11. A-VS	2.2	1.5-2.9	0.1	1.5	1.1-2.3	0.3	2.9	2.6-3.5	0.3	0.000	0.032	0.000
12. OS-VS	1.6	0.9-2.5	0.4	1.0	0.6-1.6	0.3	2.3	1.9-3.1	0.4	0.003	0.031	0.000
13. VS-Vit	14.4	9.8-21.8	4.2	8.1	6.0-12.1	2.0	17.0	12.8-25.9	4.4	0.000	0.264	0.000
14. Vit-P	13.5	5.8-33.4	7.1	4.1	1.3-8.1	1.7	7.7	5.8-10.8	1.4	0.000	0.059	0.014
15. VS-P	24.3	14.1-37.8	7.3	12.3	8.8-20.0	3.3	23.7	19.3-29.0	4.3	0.000	0.946	0.000
16. PhL	0.6	0.4-0.8	0.1	0.4	0.2-0.5	0.1	0.7	0.6-0.9	0.1	0.000	0.265	0.000
17. PhW	0.5	0.3-0.8	0.1	0.3	0.2-0.5	0.8	0.5	0.3- 0.6	0.1	0.000	0.794	0.000
18. TL	12.6	8.3-19.1	3.7	7.4	5.7-11.0	1.7	13.7	10.7-19.7	2.8	0.000	0.499	0.000

19. TW	4.2	2.5-7.5	4.2	5.3	3.7-8.1	1.5	6.2	4.7-7.6	0.9	0.049	0.000	0.239
20. TP	31.6	22.2-43.1	7.6	25.0	17.8-33.2	6.1	35.9	30.6-46.1	4.7	0.019	0.163	0.000

Fg, *Fasciola gigantica*; *Fh*, *Fasciola hepatica*; *Fif*, *Fasciola* intermediate forms. M, mean. n, sample size; Min, minimum, Max, maximum; SD, standard deviation; BL, Body length, BL; BW, Body width, BW; BWOv, BW at ovary level; BP, Body perimeter; CL, Cone length; CW, Cone width; Osmax, Oral sucker maximum diameter; Osmin, Oral sucker minimum diameter; Vsmax, Ventral sucker maximum diameter; Vsmin, Ventral sucker minimum diameter; A-VS, Distance between anterior end of body and VS; OS-VS, Distance between suckers; VS-Vit, Distance between VS and union of vitelline glands; Vit-P, Distance between Vit and posterior end of body; VS-P, Distance between VS and posterior end of body; PhL, Pharynx length; PhW, Pharynx width; TL, Testicular space length; TW, Testicular space width; TP, Testicular space perimeter.

Validated classifications according to three methods and three morphometric approaches (Traditional, Outlines and Landmarks).

Percentages of correct classification according to data sets and classification methods. *Fg*, *Fasciola gigantica*; *Fg*, *Fasciola hepatica*; *Fif*, *Fasciola* intermediate forms; LN, natural logarithms; LSR, Log-Shape-Ratios; OD, outline data, here Normalized Elliptic Fourier coefficients; LD, Procrustes residuals (orthogonal projections). *ANN*, artificial neural network. For ANN, the results are presented as averages over 30 sessions and, between square brackets, the standard deviation; *MxL*, maximum likelihood validated reclassification: between brackets, the number of PC giving the best reclassification score; the *Dist* (PC) the Mahalanobis distance-based validated method used the 8 first principal components of the variables LN and OD, and the 6 PC of ORP.

1   **Table 3.** Validated classifications according to three methods and three morphometric approaches

2   (Traditional, Outlines and Landmarks).

Methods	Samples/ Variables	Traditional				Outlines	Landmarks
		20 LN	LSR (20 LN)	12 LN	LSR (12 LN)	30 OD	6 LD
ANN	17 *Fg*	13.63 [0.85]	13.5 [1.01]	13.47 [1.04]	14.17 [1.05]	14.96 [0.69]	10.81[2]
	9 *Fh*	8.4 [0.67]	7.6 [0.72]	8.43 [0.68]	7.73 [0.69]	7.13 [0.85]	5.65 [1.21]
	10 *Fif*	9.97 [0.18]	9.93 [0.25]	9.83 [0.46]	8.43 [0.73]	9.96 [0.20]	8.33 [0.95]
	36	89% [2.8]	**86% [3.5]**	88% [3.5]	**84% [2.8]**	**89% [3.5]**	**68% [5.6]**
MxL (PC)	17 *Fg*	16	15	17	15	14	9
	9 *Fh*	7	7	7	7	8	6
	10 *Fif*	9	10	8	9	10	6
	36	89% (3)	**89% (4)**	89% (4)	**86% (3)**	**89% (4)**	**58% (3)**
Dist (PC)	17 *Fg*	15	15	17	17	16	13
	9 *Fh*	8	8	8	9	8	5
	10 *Fif*	10	10	10	9	10	7
	36	92%	**92%**	97%	**97%**	**94%**	**69%**

4   Percentages of correct classification according to data sets and classification methods. *Fg*,

5   *Fasciola gigantica*; *Fg*, *Fasciola hepatica*; *Fif*, *Fasciola* intermediate forms; LN, natural

6   logarithms; LSR, Log-Shape-Ratios; OD, outline data, here Normalized Elliptic Fourier

7   coefficients; LD, Procrustes residuals (orthogonal projections). *ANN*, artificial neural network. For

8   ANN, the results are presented as averages over 30 sessions and, between square brackets, the

9   standard deviation; *MxL,* maximum likelihood validated reclassification: between brackets, the

10   number of PC giving the best reclassification score; the *Dist* (PC) the Mahalanobis distance-based

11   validated method used the 8 first principal components of the variables LN and OD, and the 6 PC

12   of ORP.

Global sizes of *Fasciola* spp.

Fg, *Fasciola gigantica*; *Fif*, *Fasciola* intermediate forms; *Fh*, *Fasciola hepatica*; n, number of specimen; Min, minimum; Max, maximum; SD Standard deviation. Perimeter and square-root area as computed by the outline-based technique, and centroid size by the Generalized Procrustes Analysis (landmark-based technique).

1 **Table 4**. Global sizes of *Fasciola* spp.

2 **Table 4a.** Body perimeter and Square-root area of *Fasciola* spp.

Species	n	Mean	Min	Max	SD
			Body Perimeter		
Fg	17	65.44	45.22	88.47	13.96
Fh	9	62.25	55.55	75.40	6.88
Fif	10	38.67	29.05	56.83	9.14
			Square-root area		
Fg	17	13.5	9.04	17.89	2.54
Fh	9	14.3	10.64	14.68	1.76
Fif	10	9.8	5.20	10.40	2.33

5 **Table 4b.** Centroid size of *Fasciola* spp.

Species	n	Mean	Min	Max	SD
Fg	17	17.44	11.83	28.33	4.98
Fh	9	17.16	14.48	22.00	2.52
Fif	10	9.45	6.93	13.36	2.31

7 *Fg, Fasciola gigantica*; *Fif, Fasciola* intermediate forms; *Fh, Fasciola hepatica*; n, number of

8 specimen; Min, minimum; Max, maximum; SD Standard deviation. Perimeter and square-root

9 area as computed by the outline-based technique, and centroid size by the Generalized Procrustes

10 Analysis (landmark-based technique).

Non-parametric comparisons of global size estimations, 1000 permutations (*P*-values)

P, percentage of values equal or higher than the observed differences of mean values between species after 1000 random permutations. It is the empirical risk of error saying that the differences are significant. BP, body perimeter.

1 **Table 5.** Non-parametric comparisons of global size estimations, 1000 permutations (*P*-values)

Species	Centroid size	BP / Square-root area
F. gigantica / *F. hepatica*	*P = 0.909*	*P = 0.650 / P = 0.555*
F. gigantica / *F.* intermediate forms	*P < 0.001*	*P < 0.001 / P = 0.003*
F. intermediate forms / *F. hepatica*	*P < 0.001*	*P < 0.004 / P < 0.001*

3 *P*, percentage of values equal or higher than the observed differences of mean values between

4 species after 1000 random permutations. It is the empirical risk of error saying that the

5 differences are significant. BP, body perimeter.

Mahalanobis distance and *P*-value for outline-based and landmark-based GM analyses

Fg, *Fasciola gigantica*; *Fg*, *Fasciola hepatica*; *Fif*, *Fasciola* intermediate forms

1 **Table 6.** Mahalanobis distance and *P*-value for outline-based and landmark-based GM analyses

Species	Outline-based GM			Landmark-based GM		
	Fg	*Fh*	*Fif*	*Fg*	*Fh*	*Fif*
Fg	0.00			0.00		
Fh	2.28	0.00		1.68	0.00	
	(p = 0.108)			(*p* = 0.063)		
Fif	4.42	4.23	0.00	2.85	2.86	0.00
	(p<0.001)	(p = 0.001)		(*p* < 0.001)	(*p* < 0.001)	

3 *Fg*, *Fasciola gigantica*; *Fg*, *Fasciola hepatica*; *Fif*, *Fasciola* intermediate forms

Scores of validated classification using as input for the multilayer perceptron either log-transformed linear measurements (LN) alone, or a combination of LN with outline data.

Percentages are the average of 30 ANN classification sessions (see Supplementary data 1), with the standard deviation between square brackets. LN, log-transformed linear measurements; OD, outline data, i.e., normalized elliptic Fourier coefficients. The first row may be read as follows: with only four log-transformed linear measurements (4LN, first column), 78% of correct species assignment was obtained after validated classification (second column), 88% when these data were combined with outline data (third column).

1   **Table 7.** Scores of validated classification using as input for the multilayer perceptron either log-

2   transformed linear measurements (LN) alone, or a combination of LN with outline data.

	LN		LN + OD	
	average	stdev	average	stdev
4 LN	73%	[4.8]	88%	[4.0]
	BL, VS-Vit, Vit-P, VS-P			
6 LN	78%	[4.2]	88%	[4.0]
	4LN above + OS-VS, A-VS			
8 LN	81%	[4.5]	88%	[5.3]
	6LN above + OS-VS, CL, TL, BP			
10 LN	88%	[3.5]	94%	[3.7]
	8 LN above + BW, TW			
12 LN	88%	[3.5]	96%	[3.0]
	10 LN above + VSmin, A-VS			

4   Percentages are the average of 30 ANN classification sessions (see Supplementary data 1), with

5   the standard deviation between square brackets. LN, log-transformed linear measurements; OD,

6   outline data, i.e., normalized elliptic Fourier coefficients. The first row may be read as follows:

7   with only four log-transformed linear measurements (4LN, first column), 78% of correct species

8   assignment was obtained after validated classification (second column), 88% when these data were

9   combined with outline data (third column).