



**HAL**  
open science

# Image based species identification of Globodera quarantine nematodes using computer vision and deep learning

Romain Thevenoux, Van-Linh Le, Heloïse Villessèche, Alain Buisson, Marie Beurton-Aimar, Eric Grenier, Laurent Folcher, Nicolas Parisey

## ► To cite this version:

Romain Thevenoux, Van-Linh Le, Heloïse Villessèche, Alain Buisson, Marie Beurton-Aimar, et al.. Image based species identification of Globodera quarantine nematodes using computer vision and deep learning. *Computers and Electronics in Agriculture*, 2021, 186, 10.1016/j.compag.2021.106058 . hal-03319310

**HAL Id: hal-03319310**

**<https://hal.inrae.fr/hal-03319310>**

Submitted on 13 Jun 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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Image based species identification of *Globodera* quarantine nematodes using**  
2 **computer vision and deep learning**

3

4 Romain Thevenoux<sup>1,2</sup>, Linh Le Van<sup>4</sup>, Heloïse Villesseche<sup>1,2,\*</sup>, Alain Buisson<sup>2</sup>, Marie  
5 Beurton-Aimar<sup>3</sup>, Eric Grenier<sup>1</sup>, Laurent Folcher<sup>2</sup>, Nicolas Parisey<sup>1</sup>

6

7 *1 IGEPP, INRAE, Agrocampus Ouest, Université de Rennes 1, 35650, Le Rheu,*  
8 *France*

9 *2 ANSES, Agence nationale de sécurité sanitaire de l'alimentation, de*  
10 *l'environnement et du travail, Laboratoire de la santé des végétaux - Unité de*  
11 *nématologie, 35653 Le Rheu, France*

12 *3 LaBRI, CNRS, Université de Bordeaux, 33405 Talence, France*

13 *4 Inria Bordeaux Sud-Ouest, 200 avenue de la Vieille Tour, 33405 Talence, France*

14

15 *\* Current address: DigitAg, INRAE, MUSE, Montpellier SupAgro, 34196 Montpellier,*  
16 *France*

17

18 Corresponding author: Romain Thevenoux; [romain.thevenoux@inrae.fr](mailto:romain.thevenoux@inrae.fr)

19

20

21

22 **Abstract**

23 Identification of plant parasitic nematode species is usually achieved following  
24 morphobiometric analysis, which requires a certain level of expertise and remains  
25 time consuming. Moreover, molecular and morphological discrimination of a number  
26 of emergent or cryptic species is sometimes difficult. Finding a way to achieve  
27 morphological characterisation quickly and accurately would greatly advance  
28 nematology science. Here, we developed a complete method in order to identify the  
29 two quarantine nematode species *Globodera pallida* and *Globodera rostochiensis*.  
30 First, we chose discriminative metrics on the stylet of nematodes that are able to be  
31 used by algorithms in order to build an automated process. Second, we used a  
32 custom computer vision algorithm (CCVA) and a convolutional neural network (CNN)  
33 to measure our metrics of interest. Third, we compared the CCVA and CNN  
34 predictions and their discriminative power to distinguish closely related species.  
35 Results show accurate identification of *G. pallida* and *G. rostochiensis* with the two  
36 methods, despite small-scale divergence (one to five  $\mu\text{m}$  depending on the metric  
37 used). However, the error rate is higher for *Globodera mexicana*, suggesting that the  
38 algorithms are too specific. Nonetheless, these methods represent a promising novel  
39 approach to automated morphological identification of nematodes and *Globodera*  
40 species in particular.

41

42 **Keywords:** automation; landmarks; machine learning; morphometrics; potato cyst  
43 nematode; nematode taxonomy

44

## 45 **1. Introduction**

46

47 Plant parasitic nematodes (PPNs) are important pests because of their feeding  
48 habits and the role they play in spreading viruses and disease. Over 4 100 species of  
49 plant-parasitic nematodes have been described to date (Decraemer and Hunt, 2006),  
50 and several represent a serious constraint for delivery of global food security (Nicol et  
51 al., 2011). The worldwide economic burden of PPNs has increased annually over the  
52 last period and reached 358 billion USD in 2013 (Abd-Elgawad and Askary, 2015;  
53 Bernard et al., 2017). Consequently, some PPN species are regulated at the  
54 European (Council directive 2016/2031/EC, 2016) and/or international levels.  
55 Implementing these specific regulations requires epidemiological monitoring in order  
56 to apply the possibly associated mandatory management measures. This  
57 epidemiosurveillance requires identification of the species involved.

58 Cyst nematodes (*Heterodera* and *Globodera* spp.) ranked second in the 'top  
59 10' list of plant parasitic nematodes (Jones et al., 2013), and potato cyst nematodes  
60 (PCNs) were clearly highlighted as one of the most damaging groups of species in  
61 these genera in the same study. PCNs are believed to have originated in the Andean  
62 region of Peru and Bolivia (Grenier et al., 2010). Today, PCNs occur on all  
63 continents, in the temperate, tropical or southern tropical zones, both at sea level and  
64 at higher altitudes corresponding to the Andes mountains. All PCN species belong to  
65 the *Globodera* genus. At this time, at least five *Globodera* species parasitizing  
66 Solanaceae have been identified (Subbotin et al., 2020). All these species are able to  
67 develop on tomato, but potato is not a host for two of them (*G. tabacum* and *G.*  
68 *mexicana*). Among the remaining three are the well-known *G. pallida* and *G.*  
69 *rostochiensis* quarantine species. *Globodera rostochiensis* is currently reported in 75

70 countries (EPPO 2020-06-15). *Globodera rostochiensis* seems to have originated in  
71 Bolivia, where the highest genetic diversity was observed (Boucher et al., 2013;  
72 Subbotin et al., 2020). *Globodera pallida*, also known as the pale potato cyst  
73 nematode, is found in 53 countries (EPPO 2020-06-15), mostly in temperate regions.  
74 Following a phylogeographic study carried out on *G. pallida* populations sampled  
75 along the Andean Cordillera in Peru, it was shown that the South of Peru seems to  
76 be the origin of this species, and that all the populations found in Europe originated  
77 from a very limited geographical area in South Peru (Plantard et al., 2008).

78         Following isolation of the nematodes after an extraction process, their  
79 identification to the genus and species level is usually achieved by nematologists  
80 using morphobiometric techniques. Species identification by genome sequencing or  
81 molecular testing is possible but not always available, especially for new quarantine  
82 species or emergent nematodes. In the case of *G. pallida* and *G. rostochiensis*,  
83 several molecular tests exist for their identification but morphobiometric techniques  
84 are often associated or can also be implemented alone. Furthermore, as skills in  
85 nematology are less common than those used in more widely known disciplines,  
86 such as bacteriology, mycology, virology and entomology, morphobiometric  
87 characterisation in nematology can become critical. In fact, discrimination between  
88 species by measurement of morphological criteria through quantitative and qualitative  
89 aspects requires a particular level of expertise and remains time consuming and  
90 repetitive for nematologists. The morphological identification of PCNs is based on  
91 observations and measurements taken on the cyst form stage and/or on second-  
92 stage juveniles. Some of the most useful criteria concerning cysts are: Granek's ratio  
93 (ratio between the distance from the nearest edge of vulval basin to anus and the  
94 diameter of vulval basin), the distance between the vulva and anus, or the number of

95 ridges between the vulva and anus. On second-stage juveniles, corresponding to the  
96 infective stage that will hatch from the cysts, the most useful criteria are: the body  
97 length, the distance from the tail to the excretory pore, the length of the hyaline part  
98 of the tail, the length of the stylet, and the shape of the stylet basal knobs (Perry et  
99 al., 2018).

100         The so-called 'traditional morphometrics' uses sets of measurements of the  
101 size or length of anatomical parts. It can also concern the proportions and relative  
102 positions of these parts, i.e. the analysis of the shape. In morphometrics, shape is  
103 defined as what remains invariant to rotation, translation and homothety (Kendall,  
104 1989), which are usually represented by sets of landmarks that can be located  
105 precisely on all forms, and establish a clear one-to-one correspondence between all  
106 specimens included in a study. A morphometric study may even combine shapes  
107 with other complex features, like quantitative descriptors of the colour or texture of an  
108 anatomical part. However, independently from measurement complexity, the first step  
109 in any morphometric study is digital imaging of the biological specimens with  
110 controlled illumination and contrasting background. As such, modern morphometrics  
111 can be thought of as a features detection approach using robust digital image  
112 processing (Gonzalez et al., 2004). Historically, there have been numerous image  
113 processing algorithms to measure human morphology that have been used for a  
114 particular purpose such as face detection (Viola and Jones, 2001), or biometrics for  
115 security (Cintas et al., 2016) and medical purposes (Dai et al., 2019). The usual  
116 course of events is that methods are then adapted to other species, leading for  
117 example to a series of applications of advanced image processing for morphometrics  
118 in entomology (Akintayo et al., 2018; Palaniswamy et al., 2010; Porto and Voje,  
119 2020; Vandaele et al., 2018). The use of computer vision has many advantages: (i) it

120 usually comes with a formal quantification of measurement errors, (ii) it enables  
121 better control over repeatability for instance over time, and (iii) it transforms the effort  
122 of training staff into that of deploying specialised software on computing resources.  
123 Yet we will still need experts, at least for building training sets. However, automating  
124 species identification with computer based approaches can ease identification for  
125 non-experts. An interesting fact about automatic morphometric measurement is that it  
126 is a proxy for many different scientific questions. While image-based insect  
127 classification within or among species is quite popular (Martineau et al., 2017),  
128 it serves only one purpose. Automatic generation of landmarks or measurements, on  
129 the other hand, can help in a variety of questions including regressions on  
130 (environmental) covariates, studies of shape-covariations, evaluation of phylogenetic  
131 signals and, of course, classification.

132         Deep learning methods were recently found to achieve good results to  
133 automatically extract features from images. Over the past ten years, in most image-  
134 processing applications, the use of neural networks has been tested and the results  
135 have been compared to those obtained with classical algorithms. Although neural  
136 networks require huge amounts of data and large computing resources to be  
137 efficient, one of their advantages is that case-specific pre-processing operations,  
138 such as segmentation, are not required. They have exhibited good results in image  
139 recognition-classification (Krizhevsky et al., 2012), language recognition (Mikolov et  
140 al., 2011), and face detection (Li et al., 2015). Concerning landmark identification, we  
141 can cite a few examples of applications on human faces: Cintas et al. (2016) have  
142 defined a network to predict landmarks on human ears, Zhang et al. (2018) to  
143 discover landmarks on the face, and Le et al. (2020) for landmarks on beneficial  
144 insects. A review of facial landmark detection can also be found in Wu and Ji (2019).

145 In this context, it would be useful to develop an automated morphometric  
146 analysis tool, based on classical computer vision or deep learning, in order to identify  
147 the two quarantine nematode species, *G. pallida* and *G. rostochiensis*, as a model for  
148 further development to *Globodera* species identification. First, we chose metrics that  
149 can be used by algorithms with the aim of developing an automated process. We  
150 checked whether these metrics are useful to distinguish *G. rostochiensis* from *G.*  
151 *pallida*. In a second step, in order to automatically measure our metrics of interest,  
152 we used and compared a custom computer vision algorithm (CCVA) and a  
153 convolutional neural network (CNN), based on previous work on morphometrics for  
154 carabids (Le et al., 2020).

155

156

## 157 **2. Materials and methods**

158

### 159 2.1. Biological material and image acquisition

160 Fourteen populations from Peru, Europe and Mexico were studied and are  
161 listed in Table 1. All are from the IGEPP/INRAE laboratory collection and were  
162 multiplied on the potato cultivar “Désirée.” For each population, 30 individuals (J2  
163 stage) were randomly chosen and each individual was processed between slide and  
164 slip cover. Slides were exposed using a Firlabo hot plate to a temperature of 45°C for  
165 few seconds to kill the J2 suddenly and have the J2 in a similar elongated shape to  
166 help standardise image acquisition. Individuals were photographed using a Sony  
167 XCD-U100CR camera (6.3 zoom) fitted to a Wild Leitz model DAPLAN microscope.  
168 We took one image for each individual’s head with 40X zoom. For all images, a pixel  
169 corresponds to a square of 0.18 µm per side. The total image set was subsequently



170 divided into a training set (composed of 300 images corresponding to four  
171 populations of *G. rostochiensis* and six populations of *G. pallida*), a test set  
172 (composed of 60 images corresponding to one population of *G. rostochiensis* and  
173 one population of *G. pallida*), and a species test set (composed of two populations of  
174 *G. mexicana*) (Table 1). The test set was used to test the predictive power of the  
175 image processing algorithms (images from these populations were not used during  
176 the algorithm training step), while the species test set was used to test the predictive  
177 power of the algorithms on images of a different but closely related species that  
178 never seen by any parts of our process

## 179 2.2. Manual landmarks and observed metrics

180 Landmarks were placed thanks to Fiji software (Schindelin et al., 2012) and  
181 the Landmarks package (Longair and Jefferis, 2006). Metrics used for this study were  
182 chosen to make image processing easier because they are metrics on objects easy  
183 to spot on the image's scene. In short, two metrics were tested in this study, the  
184 "basal knobs width" (BKW) and the "basal knobs to head length" (BKTH) (Figure 1).  
185 BKTH is a measure similar to the stylet length but which will consider the distance  
186 from the basal knobs to the head instead of to the stylet tip. We hypothesised that  
187 relaxation of the muscle appending the stylet when killing the J2 would be  
188 homogeneous across individuals, thus allowing us to standardise J2 measurements.  
189 Landmarks to extract BKW and BKTH metrics were set by hand in all the images  
190 corresponding to the training set. BKW is a rarely used metric. It was referenced in  
191 the European and Mediterranean Plant Protection Organisation (EPPO) bulletin  
192 (2017) and by Subbotin et al. (2011), but this metric is absent in most J2 identification  
193 keys. This is also true for BKTH which was referenced by Ponce (1977) but is absent  
194 in all nowadays J2 identification keys.

195

### 196 2.3. Automatic extraction of morphological metrics

197 All processes used in this study are summarised in Figure 2. In order to  
198 measure our metric of interest, we used two different types of algorithms: one was a  
199 CCVA based on well-known steps (Gonzalez et al., 2004), and the other was a deep  
200 neural network based on previous work on morphometrics for carabids (Le et al.,  
201 2020) using a CNN.

202 For the CCVA, we first isolated each individual's head from the image  
203 background using a simple combination of contour detection and mathematical  
204 morphology (Gonzalez et al., 2004). We then detected the best ellipsoidal  
205 approximation to the basal knobs by using an elliptical Hough transform on the  
206 head's inner contour (Xie and Ji, 2002). This enabled us to compute BKW as the  
207 radius of the major axis of this ellipse. So, for BKW we expected a prediction twice  
208 lower than the observed metrics. To compute BKTH, we searched for the extremum  
209 of the head's outer contour relative to its orientation (i.e. the tip of the curved head).  
210 We then computed BKTH as the Euclidean distance between the centre of the BKW  
211 detected ellipse and the detected head tip. The most important hyperparameters for  
212 BKTH are the sigma of Canny in the first detection of contour, the morphology square  
213 used for dilated and eroded images with the aim of cleaning the background, and the  
214 minimum and maximum size of the desired ellipse. We fixed the ellipse size thanks to  
215 knowledge of morphometry on the two-target species and all other hyperparameters  
216 using a sparse grid search.

217 For the CNN computation, we used an EB-Net architecture based on the  
218 concept of elementary block (EB) (Le et al., 2020). An EB (see Annex 1) contains a  
219 convolution layer to extract features from the input images, a max pooling operation

220 to reduce the number of parameters, and a drop-out layer to prevent overfitting in the  
221 training process (Srivastava et al., 2014). The EB-Net configuration used for our  
222 purposes contained three EBs. The sequences of the EBs ended with three full  
223 connected layers to output eight values, corresponding to the coordinates of the four  
224 landmarks used on the head. To define a CNN, several hyperparameters must be set  
225 as the number of epochs or the learning rate; for this application, we set the number  
226 of epochs to 5 000 and the learning rate 0.03 to 0.0001. hyperparameters has been  
227 set after several experiments on another biological model (Lé *et al.*, 2020). The root  
228 mean square error (RMSE) was used as the loss function because the output values  
229 are quantitative. To process a CNN, the size of the “training set” must be enlarged.  
230 This was obtained by extraction of colour channels to create new images as  
231 described in Le et al. (2020). The final number of images considered was 14 490 for  
232 training and validation (with a ratio of 80% for training and 20% for validation). A total  
233 of 30 initial images were kept for the testing step; these images were not yet seen by  
234 the network before this step. Implementation of the network used the Lasagne  
235 framework (Dieleman et al., 2015) and can be found at  
236 [https://github.com/linhlevandlu/EBNet\\_nematodes](https://github.com/linhlevandlu/EBNet_nematodes). We trained the network from  
237 scratch using K-fold cross-validation, i.e. training on 9 out of 10 of the populations  
238 and predictions on the individuals of the remaining population. Hence, each  
239 population was predicted at only one timepoint (not trained upon) and we used these  
240 predictions to extract BKW and BKTH as Euclidean distances between the  
241 corresponding predicted landmarks. For populations without manual landmarks  
242 (Ecosse, Chavornay, Tlaxcala and Santa Anna), we trained a network on all the  
243 populations of the training set, and then predicted based on these four landmarks.

244

## 245 2.4. Statistical classification of individuals and populations into species

246 We used a binomial regression, as a generalised linear model (Nelder and  
247 Wedderburn, 1972), to predict the species of each individual based on the BKW and  
248 BKTH measurements. We defined the population with manual landmarks as a  
249 training set, and the populations without as a test set (de Vienne et al., 2013). In  
250 order to assess the predictive power of our regressions, we used leave-one-out  
251 cross-validation on the training set (de Vienne et al., 2013), and prediction on the test  
252 set. We predicted the assignment of each individual to one species or the other.  
253 Measuring the predictive power of our regression using accuracy (the ratio between  
254 the number of correct prediction and the number of individuals in the population) and  
255 kappa index. The kappa index coefficients used to measure the degree of agreement  
256 among raters (in our case human versus algorithms), taking into count the possibility  
257 of said agreement occurring by chance (Cohen, 1960). For the assignment of a  
258 population to a species, we used a majority vote, i.e. the population belongs to the  
259 species of most of its individuals.

260

## 261 **3. Results**

262

### 263 3.1. Predicted versus observed metrics

264 First, we observed that *G. pallida* and *G. rostockiensis* were significantly  
265 differentiated using the observed measurements based on the manual landmarks  
266 (Figure 3 A, B): *G. pallida* had longer BKW and BKTH than *G. rostockiensis*, as  
267 expected. From the 300 images constituting the training set, the CNN extracted  
268 measurements from all images and the CCVA from 81% of them. Most of the  
269 remaining 19% is due to slightly blurry images. Measurements from the CCVA, CNN

270 and manual landmarks were compared by simple mean comparison (T. test, Figure 3  
271 A, B). For both metrics, there were no significant differences between measurements  
272 predicted by the CNN and the observed measurements based on the manual  
273 landmarks (Figure 3 A, B). However, the situation appeared to be different regarding  
274 the CCVA predicted metrics. For BKW, measurements predicted by the CCVA were  
275 significantly smaller than the manual landmarks, but this is clearly linked to the fact  
276 that BKW metrics for the CCVA correspond to only the radius of the ellipse.  
277 Nonetheless, the two species are still distinguishable (Figure 3A). For BKTH,  
278 measurements predicted by the CCVA were also found to be smaller and we  
279 hypothesised that this was due to the approximation of the basal knobs position  
280 using the centroid of the predicted ellipse, and not the bottom of the ellipse.  
281 Nonetheless, a difference between *G. pallida* and *G. rostochiensis* was still observed,  
282 and was consistent with the observed metrics. Correlations between predicted and  
283 observed measurements were high for all metrics and algorithms, except the  
284 prediction of BKTH by CCVA ( $R^2 = 0.28$ ). In most cases, segregation between *G.*  
285 *pallida* and *G. rostochiensis* was clear, although for some individuals, *G. pallida*  
286 always had higher measurements and prediction than *G. rostochiensis* (Figure 3 C,  
287 D, E, F). Nonetheless, there is still an overlap between the two species for the  
288 metrics studied: extreme cases are clear, but mean cases are more confusing.  
289 Therefore, there is a need for classifier tools in order to identify species.

290 Time needed to process the CNN was about seven hours for the learning step  
291 and 30 minutes to extract measurements on the whole training set. Time needed by  
292 CCVA was about 35 minutes to extract measurements on the whole training set, but  
293 building the CCVA took several weeks. Compared to the respective 30 and 35

294 minutes needed by the CCVA and the CNN, landmarking on the same set of images  
295 required at least four hours.

296

### 297 3.2. Classification using the training set

298           At the individual scale, the classifier running with observed metrics showed  
299 an accuracy of 0.96 and a Kappa index of 0.91 (Table 2). Comparatively, the  
300 classifier running with the predicted CNN and CCVA metrics showed an accuracy of  
301 0.83 and 0.85, respectively and a kappa index of 0.65 and 0.74, respectively. All  
302 populations were not predicted with the same accuracy. For example, population  
303 P273/2016 is the most poorly predicted population with a 17% error rate using the  
304 observed landmark, 32% using the CCVA predicted metrics, and 24% using the CNN  
305 predicted metrics. Nonetheless, the rank among population predictions is conserved  
306 throughout the entire set, as populations that showed a high error rate using the  
307 observed landmarks also showed a high error rate using the CNN and CCVA  
308 predicted metrics. The sole exception was population Dunkerque which showed a  
309 perfect prediction score using the observed landmarks, but lower accuracy using the  
310 CNN predicted metrics (23% error rate) or the CCVA predicted metrics (24% error  
311 rate). Nonetheless, at the population scale, with a decision rule of the majority of  
312 individuals predicted in a class, we have perfect accuracy with no errors, regardless  
313 of the data set or the algorithms (CNN or CCVA) used (Table 2).

314

### 315 3.3. Classification using the test set

316           Prediction on the test set showed contrasting results (Table 2). Using metrics  
317 predicted by CCVA, prediction scores (accuracy of 0.83 and kappa index of 0.71)  
318 were similar to the prediction scores observed using the training set. Prediction

319 scores from CNN metric extraction (accuracy of 0.71 and kappa index of 0.43) were  
320 significantly lower than those obtained using the training set. This is mainly due to the  
321 incorrect prediction observed for population Ecosse, which was predicted as *G.*  
322 *pallida*, while it was in fact *G. rostochiensis*. This is the only false prediction between  
323 *G. pallida* and *G. rostochiensis* observed on the whole image set.

324

### 325 3.4. Classification using the species test set

326         There were few differences between the three species that we worked on for  
327 the two metrics studied when measurements were done with manual landmarks  
328 (Figure 4 A). *Globodera mexicana* showed significantly larger BKW than *G. pallida*,  
329 but we observed no significant difference between these two species for the BKTH  
330 metric. Importantly, the observed mean difference between *G. mexicana* and *G.*  
331 *pallida* was only 2.5 pixel (0.45  $\mu\text{m}$ ) for BKW. BKW was statistically different between  
332 the three species studied (Figure 4 A). For the two metrics studied, *G. rostochiensis*  
333 had the lower measurements.

334         The prediction by CNN (Figure 4 B) gave the same result for BKTH  
335 (measurements for *G. pallida* and *G. mexicana* were not significantly different and  
336 measurements for *G. rostochiensis* were significantly different from the other two  
337 species), but the significant difference previously observed between *G. mexicana* and  
338 *G. pallida* for BKW was not observed using CNN predictions. For both metrics, CNN  
339 predicted measurements that were statistically similar for *G. mexicana* and *G. pallida*,  
340 and significantly lower measurements for *G. rostochiensis*. With CCVA,  
341 measurements were extracted for 75% of the individuals. Surprisingly, predictions by  
342 CCVA (Figure 4 C) yielded a significantly smaller BKW for *G. mexicana* than for *G.*  
343 *pallida*, and a similar BKW for *G. mexicana* and *G. rostochiensis*. For BKTH,

344 prediction by CCVA was similar to observed metrics and CNN predictions, meaning  
345 similar prediction for *G. pallida* and *G. mexicana*, and lower prediction for *G.*  
346 *rostochiensis*. Including *G. mexicana* in our dataset showed that the predictions could  
347 yield a different ranking depending on the metric and method used. In fact, even  
348 though BKTH was always measured or predicted with the same ranking between  
349 species (Figure 4, second line), BKW was measured differently, predicted by CNN  
350 and predicted by CCVA (Figure 4, first line).

351 The following results indicate the mean  $\pm$  SD (standard deviation) for both  
352 metrics and species studied. For BKTH, the observed range of variation (*G.*  
353 *mexicana*: mean = 27.86  $\mu\text{m} \pm 0.94$ ; *G. rostochiensis*: mean = 24.49  $\mu\text{m} \pm 1.02$ ; *G.*  
354 *pallida*: mean = 26.49  $\mu\text{m} \pm 0.83$ ) was similar to the range of variation predicted by  
355 CNN (*G. mexicana*: mean = 25.77  $\mu\text{m} \pm 0.85$ ; *G. rostochiensis*: mean = 24.84  $\mu\text{m} \pm$   
356 0.75; *G. pallida*: mean = 26.08  $\mu\text{m} \pm 0.73$ ). These values are smaller than the range  
357 of variation predicted by CCVA (*G. mexicana*: mean = 24.71  $\mu\text{m} \pm 2.15$ ; *G.*  
358 *rostochiensis*: mean = 23.04  $\mu\text{m} \pm 2.13$ ; *G. pallida*: mean = 25.07  $\pm 1.59$ ). The means  
359 were always lower in the predictions by CCVA due to the approximation of the ellipse  
360 centre, used as a proxy in the CCVA for the bottom of the basal knobs.

361 For BKW, the observed range of variation (*G. mexicana*: mean = 5.00  $\mu\text{m} \pm$   
362 0.38; *G. rostochiensis*: mean = 3.94  $\mu\text{m} \pm 0.31$ ; *G. pallida*: mean = 4.74  $\mu\text{m} \pm 0.31$ )  
363 was similar to the range of variation predicted by CNN (*G. mexicana*: mean = 4.42  
364  $\mu\text{m} \pm 0.25$ ; *G. rostochiensis*: mean = 4.10  $\mu\text{m} \pm 0.27$ ; *G. pallida*: mean = 4.51  $\mu\text{m} \pm$   
365 0.25) and by CCVA (*G. mexicana*: mean = 2.44  $\mu\text{m} \pm 0.25$ ; *G. rostochiensis*: mean =  
366 2.21  $\mu\text{m} \pm 0.27$ ; *G. pallida*: mean = 2.89  $\mu\text{m} \pm 0.33$ ). BKW means for CCVA  
367 predictions were always lower due to our prediction of the ellipse radius (and not the  
368 ellipse diameter) for the basal knob width. However, it must be highlighted that BKW



369 estimated by CCVA was overestimated for *G. pallida*, as it was more than half of the  
370 measurements done with manual landmarks. It is probable that this overestimation  
371 observed solely for *G. pallida* was the reason why BKW estimated by CCVA allowed  
372 the distinction between *G. pallida* and *G. mexicana*.

373 The species test set contains 60 images and CCVA extract measurement for  
374 45 images. the classifier using CCVA prediction showed an accuracy of 91% for the  
375 species test set. Using the measurements based on landmarking and the CNN  
376 predicted metrics, the classifier showed accuracy of 0.85 and 0.5, respectively.

377

## 378 **4. Discussion and conclusions**

379

### 380 4.1. Usefulness of the metrics for *Globodera* taxonomy

381 We described here a complete framework, from image acquisition to species  
382 identification. There have been very few attempts to develop automated  
383 measurement in nematology. Stylet detection has already been explored (de la  
384 Blanca et al., 1992), but results were not sufficiently convincing, mostly due to image  
385 quality and difficulties related to background extraction. Another project attempted to  
386 approach species identification using a similarity coefficient (Fortuner and Ahmadi,  
387 1986; Fortuner and Wong, 1983). To our knowledge, a complete process, such as  
388 the one described here, has never been explored in nematology. Moreover, several  
389 improvements to the process proposed in this study could be put forward, especially  
390 the possibility of guidelines in image acquisition and pre-treatment. Of note, all  
391 images that we processed were oriented in the same direction due to manual re-  
392 orientation before image processing. To decrease the time needed for image  
393 acquisition, the option of taking more than one larva on images should be explored.

394 Conventionally, distinction between *G. pallida* and *G. rostochiensis* needs at  
395 least three criteria, Granek's ratio (calculated on the cyst), the stylet length, and a  
396 criterion on basal knob shape (EPPO, 2017). One of the two metrics used in this  
397 study, BKW, is a poorly used metric, and BKTH uses in its construction the stylet  
398 length metric, commonly used in *Globodera* taxonomy. Stylet length variation is  
399 known to range between 19  $\mu\text{m}$  and 23  $\mu\text{m}$  for *G. rostochiensis*, and 22.5  $\mu\text{m}$  and 25  
400  $\mu\text{m}$  for *G. pallida* (Perry et al., 2018). Even though the mean of our measurements  
401 was always higher, the range observed for BKTH was at the same scale, meaning  
402 that BKTH variation is due to length stylet variation. Also, this similar range of  
403 variation compared to stylet length metrics published in the literature (Perry et al.,  
404 2018) allowed us to validate our standardisation process of image acquisition  
405 assuming that the stylet "rest" position is similar across all individuals. The variation  
406 observed for BKW between *G. pallida* and *G. rostochiensis* was about 0.67  $\mu\text{m}$ , while  
407 within each species the variation range remained  $< 0.5 \mu\text{m}$ . Both the BKW and BKTH  
408 metrics appear to be relevant in terms of species identification and could, in time, be  
409 incorporated into a revised species identification key. Moreover, the automatisation of  
410 process delete the "observer effect" that could bias measure. Thus, the  
411 automatization gave more reproducibility in measurements.

412 *Globodera pallida* has more rounded knobs and *G. rostochiensis* more  
413 flattened ones. The basal knob shape criterion requires some expertise in  
414 nematology to be accurately defined and could be subjective. The BKW metric could  
415 make it possible to avoid this subjective and difficult criterion. Moreover, it shows very  
416 good discriminative power between *G. pallida* and *G. rostochiensis*. Surprisingly,  
417 CCVA overestimates this metric for *G. pallida* and we could hypothesise that this is  
418 linked to the knob form. Importantly, basal knobs are a 3D structure and when

419 images were taken, we reduced them to a 2D image. Thus, the shapes of basal  
420 knobs, on 2D images are more or less rounded depending on the species observed.  
421 As a consequence, the ellipse estimated by CCVA will fit differently on the real shape  
422 of basal knobs and induce a bias compared to reality. Nonetheless, it was surprising  
423 that this bias was observed only for *G. pallida* and not for *G. mexicana*, which is also  
424 described as a species displaying prominent rounded knobs.

425         Results from the species test set show us that the process could be extended  
426 to other species, with an adaptation of the metrics studied. Only the BKTH metric is  
427 well predicted by CNN and CCVA. The poor prediction for the BKW metric supports  
428 the fact that knowledge of taxonomy, systematics and morphology is still essential for  
429 species discrimination. Automation of the process can only be done with the help of  
430 an expert. However, we could also consider that the prediction of BKW by CCVA may  
431 reflect knob form differentiation. It is important to test this metric on an extended set  
432 of populations and individuals for *G. mexicana* to also take into account the genetic  
433 and phenotypic diversity in this species. Since there are few descriptions of *G.*  
434 *mexicana* in the literature, only described in a thesis (Campos-Vela, 1967), it could  
435 be useful to continue to describe this species more precisely with an automated  
436 process, and by developing other metrics adapted to image computing. We are  
437 already able to detect the body length metric (data not shown) currently used for  
438 species identification in nematology. Increasing the number of useful metrics and  
439 further exploring computing imagery methods are essential to develop automation of  
440 species identification in nematology. This would partially relieve the lack of specific  
441 skills for nematode identification, but more importantly, it should allow us to save time  
442 and process far more individuals or samples for accurate and robust identification.

443 This kind of automated tool could also decrease the cost of analyses compared to  
444 molecular tests, where some of the required consumables are expensive.

445 We have shown that we are able, mostly with data predicted by CCVA, to  
446 predict the species class of an individual and more accurately of a population.  
447 Nonetheless, the procedure used to classify individuals in this work will necessarily  
448 yield an answer between proposed choices (i.e in the species test set, between *G.*  
449 *pallida*, *G. rostochiensis* and *G. mexicana*). The classifier builds edges between each  
450 class with given data, with the aim of choosing between them, but does not build a  
451 high boundary of class. This means that we are not able to detect other species. With  
452 the aim of species detection, it would be essential to build the boundaries of each  
453 species to allow us to determine whether an unknown specimen belongs to one of  
454 the studied classes or to none of them.

455

#### 456 4.2. Comparison between the two methods tested

457 In this study, we compared two types of algorithms to predict metrics for the  
458 purpose of classification. Firstly, the two methods used showed good classification  
459 results at the individual scale. Using the training set, all populations were correctly  
460 affiliated with their species; nonetheless, species affiliation using CNN measurements  
461 seems to be less accurate than using CCVA measurements. Species affiliation using  
462 observed data was always more accurate than using the other two methods,  
463 meaning that there is always a loss of accuracy of the measurements when  
464 automation is performed. The CCVA propensity to filter poor quality images seems to  
465 improve its accuracy. This is supported by the fact that the case where the classifier  
466 using CNN measurements led to incorrect prediction at the population scale also  
467 corresponds to the case where the CCVA removed 15 images on the species test

468 set, mostly in the *G. rostochiensis* population. The choice of population was done to  
469 incorporate broad genetic diversity, and results from the classifier showed that  
470 despite this diversity, we were able to discriminate *G. pallida* from *G. rostochiensis*,  
471 even for CCVA on the test set. This gave us confidence about the robustness of the  
472 metrics chosen and the methods.

473 The point of this study was not to choose between CNN and CCVA, because  
474 these two methods could be complementary. One of the most important points  
475 regarding CNN is its short building time and flexibility. The CNN learns alone and  
476 provides results independently of image quality, while the CCVA took more time to be  
477 built and appears to be more species-specific.

478 As both CCVA and CNN are supervised learning algorithms, one could think  
479 about using unsupervised learning algorithms instead. There is a recent trend  
480 towards unsupervised methods for automated landmarking, where landmarks are  
481 learnt from the actual data (Jakab et al., 2018; Li et al., 2020; Thewlis et al., 2019).  
482 This type of approach could fit our needs well, for example by learning landmarks  
483 based on inter-intra population consistencies. The main drawback is that the  
484 algorithms require large sample sizes to generate landmarks. To go further, it would  
485 be interesting to search automatically for novel morphometric descriptors, not only  
486 landmarks. To a certain extent, at least semi-automatically, one could rely on  
487 algorithms to interpret machine learning predictions (Lundberg and Lee, 2017),  
488 analysing the interpreted predictions in order to derive new morphological  
489 descriptors.

490

#### 491 4.3. Interest for systematics in the *Globodera* genus

492 More than simply testing algorithms and classifying organisms, extracting  
493 morphobiometric data on certain organisms will provide information for systematics,  
494 in genera suffering from unresolved taxonomy, i.e. species complexes. Effectively,  
495 the systematics and taxonomy of *Globodera* are still under discussion (Subbotin et  
496 al., 2020; Subbotin et al., 2011). For example, at this time in the species complex  
497 known as the tobacco cyst nematode complex, there is little evidence for  
498 morphobiometric differentiation (Mota and Eisenback, 1993), even though some  
499 genetic differentiation has been reported (Madani et al., 2010; Subbotin et al., 2020).  
500 Providing evidence for morphobiometric differentiation will help to understand  
501 speciation in this complex. Recently, the synonymisation of *G. bravoae* (Franco et al.,  
502 2000) and *G. mexicana* (Campos-Vela, 1967) was proposed based on molecular  
503 data (Subbotin et al., 2020). It would be interesting to determine whether  
504 morphological characterisation of these two species with the algorithms proposed in  
505 this study supports this proposal. Automated measurement extraction could also be  
506 of great value in the study of newly discovered species. A potential new species of  
507 PCN has been identified genetically in the south of Peru and in Chile (Thevenoux et  
508 al., 2020), but for now there is no evidence that it is morphologically different from *G.*  
509 *pallida* and *G. mexicana*. In these cases, the contribution of automated extraction of  
510 measurements could be dual. First, the time spent to study a population is lower than  
511 when using manual measurements, which makes it possible to work with more  
512 individuals to strengthen the statistical power and the conclusions. Second, it could  
513 allow us to create new metrics, like in this study, potentially more useful and easier  
514 for morphological differentiation. As a result, automation of the identification process

515 appears to be a real interface between taxonomy, systematics, and image  
516 computing.

517

518

## 519 **Authors contributions**

520 R.T. performed the biological experiments, part of the image processing, and  
521 most statistical analyses, and drafted and managed the manuscript. L.LeV.  
522 developed relevant deep-learning tools. H.V. helped with biological experiments and  
523 early image processing. A.B. provided expertise on nematode systematics. M.B-A.  
524 provided overall guidance on image processing and helped in manuscript maturation.  
525 E.G. and L.F. provided overall guidance on ecology and evolution, and helped in  
526 manuscript maturation. N.P. provided overall guidance on image processing and  
527 statistical analysis, and helped in manuscript maturation.

528 All authors critically reviewed the manuscript. Authorship order reflects roles:  
529 from R.T. to A.B. “doers” and from M.B.A. to N.P. “enablers”; author rank in each role  
530 type is to be read left-to-right for “doers” and right-to-left for “enablers”.

531

532

## 533 **Acknowledgments**

534 We wish to thank all the colleagues who helped in material multiplication, and would  
535 like to thank the INRAE Plant Health Department and ANSES Plant Health  
536 Laboratory for their support of the NemAlliance partnership agreement which helped  
537 to make this work possible. This work is part of the PhD project of Romain  
538 Thevenoux, who is funded by ANSES and the Conseil Régional Bretagne (Brittany  
539 Regional Council).

540

541

## 542 **Data Archiving Statement**

543 The algorithms used in this study are available at the Zenodo repository for the  
544 CCVA (DOI: 10.5281/zenodo.4068341) and at GitHub for the CNN  
545 ([https://github.com/linhlevandlu/CNN\\_Beetles\\_Landmarks](https://github.com/linhlevandlu/CNN_Beetles_Landmarks)).

546

547

## 548 **References**

- 549 Abd-Elgawad, M.M., Askary, T.H., 2015. Impact of phytonematodes on agriculture economy, In:  
550 Askary, T.H., Martinelli, P.R.P. (Eds.), *Biocontrol Agents of Phytonematodes*. CABI Publishing,  
551 Wallingford, UK, pp. 3-49.
- 552 Akintayo, A., Tylka, G.L., Singh, A.K., Ganapathysubramanian, B., Singh, A., Sarkar, S., 2018. A deep  
553 learning framework to discern and count microscopic nematode eggs. *Sci. Rep.* 8, 1-11.
- 554 Bernard, G.C., Egnin, M., Bonsi, C., 2017. The impact of plant-parasitic nematodes on agriculture and  
555 methods of control, *Nematology-Concepts, Diagnosis and Control*. InTech.
- 556 Boucher, A.C., Mimee, B., Montarry, J., Bardou-Valette, S., Belair, G., Moffett, P., Grenier, E., 2013.  
557 Genetic diversity of the golden potato cyst nematode *Globodera rostochiensis* and  
558 determination of the origin of populations in Quebec, Canada. *Mol. Phylogenet. Evol.* 69, 75-  
559 82.
- 560 Campos-Vela, A., 1967. Taxonomy, life cycle and host range of *Heterodera mexicana* n. sp.  
561 (Nematoda: Heteroderidae). University of Wisconsin, Madison, WI, USA.
- 562 Cintas, C., Quinto-Sánchez, M., Acuña, V., Paschetta, C., De Azevedo, S., de Cerqueira, C.C.S., Ramallo,  
563 V., Gallo, C., Poletti, G., Bortolini, M.C., 2016. Automatic ear detection and feature extraction  
564 using geometric morphometrics and convolutional neural networks. *IET Biometrics* 6, 211-223.
- 565 Council directive 2016/2031/EC, 2016. Regulation 2016/2031 of the European parliament of the  
566 Council of 26 October 2016 on protective measures against pests of plants, amending  
567 Regulations (EU) No 228/2013, (EU) No 652/2014 and (EU) No 1143/2014 of the European  
568 Parliament and of the Council and repealing Council Directives 69/464/EEC, 74/647/EEC,  
569 93/85/EEC, 98/57/EC, 2000/29/EC, 2006/91/EC and 2007/33/EC. *Official Journal of the*  
570 *European Union*, L317: 4-104.
- 571 Dai, M., Sun, T., Chen, X., Yu, L., Chen, M., Hao, P., Zeng, X., Yan, J., Chen, S., 2019. A B-scan imaging  
572 method of conductivity variation detection for magneto–acousto–electrical tomography. *IEEE*  
573 *Access* 7, 26881-26891.
- 574 de la Blanca, N.P., Fdez-Valdivia, J., Castillo, P., Gomez-Barcina, A., 1992. Detecting nematode  
575 features from digital images. *J. Nematol.* 24, 289.
- 576 de Vienne, D.M., Refregier, G., Lopez-Villavicencio, M., Tellier, A., Hood, M.E., Giraud, T., 2013.  
577 Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations  
578 and relation to coevolution. *New Phytologist* 198, 347-385.
- 579 Decraemer, W., Hunt, D.J., 2006. Structure and Classification, In: Perry, R.N., Moens, M. (Eds.), *Plant*  
580 *Nematology*. CABI Publishing, Wallingford, UK, pp. 153-184.
- 581 Dieleman, S., Schlüter, J., Raffel, C., Olson, E., Sønderby, S.K., Nouri, D., Maturana, D., Thoma, M.,  
582 Battenberg, E., Kelly, J., 2015. Lasagne: first release, Version v0.1 ed. Zenodo.



583 Fortuner, R., Ahmadi, A., 1986. NEMAID 2.0. Computer program for identification of Nematodes.  
584 User's manual. California Department of Food and Agriculture, Sacramento, CA, USA.

585 Fortuner, R., Wong, Y., 1983. NEMAID. Computer program for identification of nematodes. User's  
586 manual. California Department of Food and Agriculture, Publication, Sacramento, CA, USA.

587 Franco, N.F., Prado Vera, I.C.d., Lamothe-Argumedo, R., 2000. *Globodera bravoae* sp. n. (Tylenchida:  
588 Heteroderidae) from Mexico. Int. J. Nematol. 10, 169-176.

589 Gonzalez, R.C., Woods, R.E., Eddins, S.L., 2004. Digital image processing using MATLAB. Pearson  
590 Education India.

591 Grenier, E., Fournet, S., Petit, E., Anthoine, G., 2010. A cyst nematode 'species factory' called the  
592 Andes. Nematology 12, 163-169.

593 Jakab, T., Gupta, A., Bilen, H., Vedaldi, A., 2018. Unsupervised learning of object landmarks through  
594 conditional image generation. Advances in neural information processing systems, NIPS  
595 Proceedings, pp. 4016-4027.

596 Jones, J.T., Haegeman, A., Danchin, E.G., Gaur, H.S., Helder, J., Jones, M.G., Kikuchi, T., Manzanilla-  
597 López, R., Palomares-Rius, J.E., Wesemael, W.M., 2013. Top 10 plant-parasitic nematodes in  
598 molecular plant pathology. Mol. Plant Pathol. 14, 946-961.

599 Kendall, D. G., 1989. "A survey of the statistical theory of shape." Statistical Science: pp. 87-99.

600 Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional  
601 neural networks. Advances in neural information processing systems, NIPS Proceedings, pp. 1097-  
602 1105.

603 Le, V.-L., Beurton-Aimar, M., Zemmari, A., Marie, A., Parisey, N., 2020. Automated landmarking for  
604 insects morphometric analysis using deep neural networks. Ecol. Inform. 60, 101175.

605 Li, H., Lin, Z., Shen, X., Brandt, J., Hua, G., 2015. A convolutional neural network cascade for face  
606 detection. Proceedings of the IEEE conference on computer vision and pattern recognition, pp.  
607 5325-5334.

608 Li, W., Liao, H., Miao, S., Lu, L., Luo, J., 2020. Unsupervised Learning of Landmarks based on Inter-  
609 Intra Subject Consistencies. ArXiv preprint arXiv:2004.07936.

610 Longair, M., Jefferis, G.S., 2006. Name landmarks and register. Fiji plugin.

611 Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Advances in  
612 neural information processing systems, pp. 4765-4774.

613 Madani, M., Subbotin, S.A., Ward, L.J., Li, X., De Boer, S.H., 2010. Molecular characterization of  
614 Canadian populations of potato cyst nematodes, *Globodera rostochiensis* and *G. pallida* using  
615 ribosomal nuclear RNA and cytochrome b genes. Can. J. Plant Pathol. 32, 252-263.

616 Martineau, M., Conte, D., Raveaux, R., Arnault, I., Munier, D., Venturini, G., 2017. A survey on image-  
617 based insect classification. Pattern Recognit. 65, 273-284.

618 Mikolov, T., Deoras, A., Povey, D., Burget, L., Černocký, J., 2011. Strategies for training large scale  
619 neural network language models, 2011 IEEE Workshop on Automatic Speech Recognition &  
620 Understanding. IEEE, pp. 196-201.

621 Mota, M.M., Eisenback, J.D., 1993. Morphometrics of *Globodera tabacum tabacum*, *G. t. virginiae*,  
622 and *G. t. solanacearum* (Nemata: Heteroderinae). J. Nematol. 25, 148-160.

623 Nelder, J.A., Wedderburn, R.W., 1972. Generalized linear models. J. R. Stat. Soc. Serie A 135, 370-384.

624 Nicol, J.M., Turner, S.J., Coyne, D.L., den Nijs, L., Hockland, S., Maafi, Z.T., 2011. Current Nematode  
625 Threats to World Agriculture. Springer, New York.

626 Palaniswamy, S., Thacker, N.A., Klingenberg, C.P., 2010. Automatic identification of landmarks in  
627 digital images. IET Comput. Vis. 4, 247-260.

628 Perry, R.N., Moens, M., Jones, T., 2018. Cyst Nematodes. CAB International, Oxfordshire, UK, Boston,  
629 MA, USA.

630 Plantard, O., Picard, D., Valette, S., Scurrah, M., Grenier, E., Mugniery, D., 2008. Origin and genetic  
631 diversity of Western European populations of the potato cyst nematode (*Globodera pallida*)  
632 inferred from mitochondrial sequences and microsatellite loci. Mol. Ecol. 17, 2208-2218.

633 Porto, A., Voje, K.L., 2020. ML-morph: A fast, accurate and general approach for automated detection  
634 and landmarking of biological structures in images. Methods Ecol. Evol. 11, 500-512.

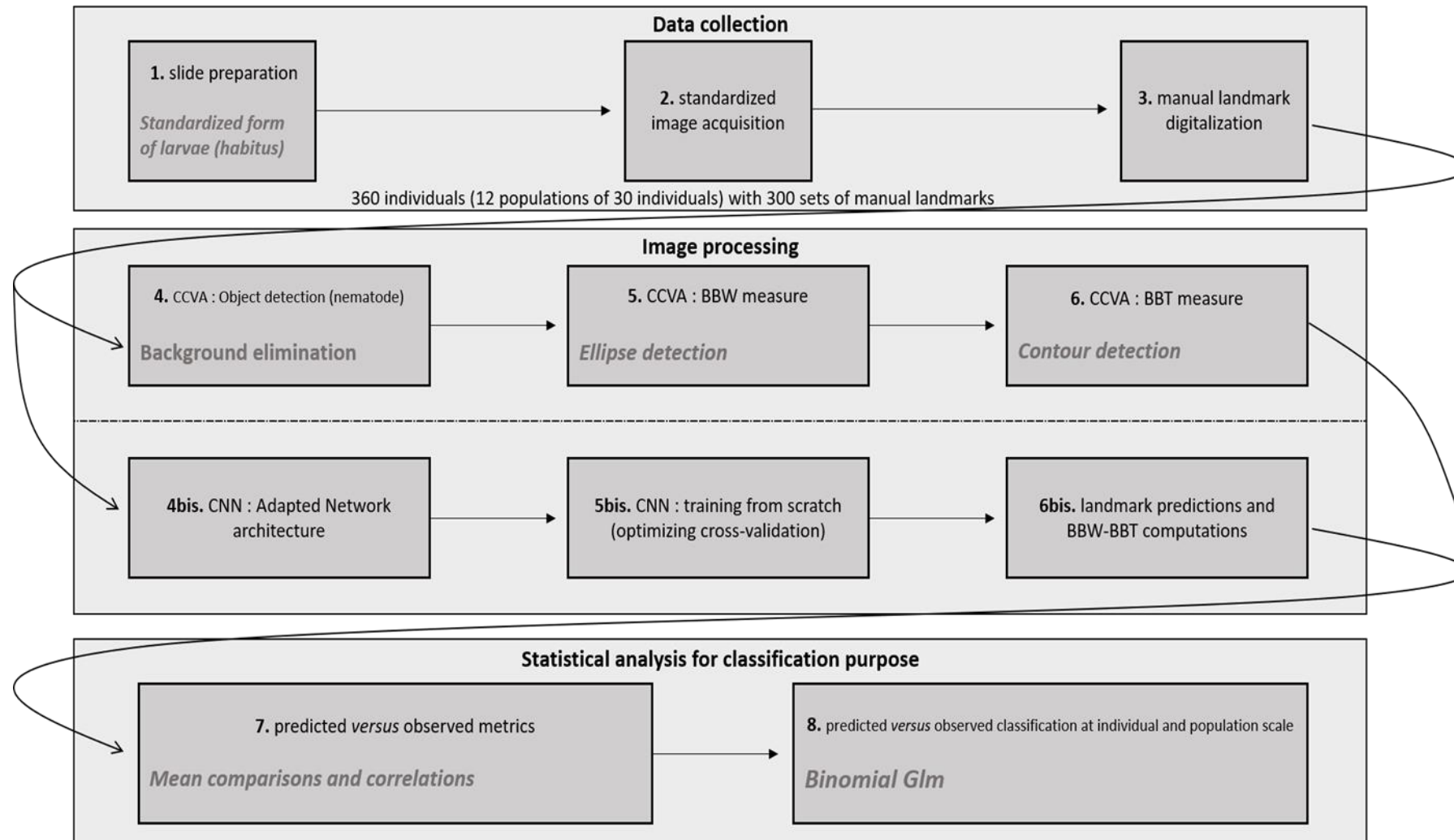
- 635 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden,  
636 C., Saalfeld, S., Schmid, B., 2012. Fiji: an open-source platform for biological-image analysis.  
637 Nat. Methods 9, 676-682.
- 638 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way  
639 to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929-1958.
- 640 Subbotin, S.A., Franco, J., Knoetze, R., Roubtsova, T.V., Bostock, R.M., Vera, I.C.d.P., 2020. DNA  
641 barcoding, phylogeny and phylogeography of the cyst nematode species from the genus  
642 *Globodera* (Tylenchida: Heteroderidae). Nematology 22, 269-297.
- 643 Subbotin, S.A., Vera, I.C.D.P., Mundo-Ocampo, M., Baldwin, J.G., 2011. Identification, phylogeny and  
644 phylogeography of circumfenestrate cyst nematodes (Nematoda: Heteroderidae) as inferred  
645 from analysis of ITS-rDNA. Nematology 13, 805-824.
- 646 Thevenoux, R., Folcher, L., Esquibet, M., Fouville, D., Montarry, J., Grenier, E., 2020. The hidden  
647 diversity of the potato cyst nematode *Globodera pallida* in the south of Peru. Evol. Appl. 13,  
648 727-737.
- 649 Thewlis, J., Albanie, S., Bilen, H., Vedaldi, A., 2019. Unsupervised learning of landmarks by descriptor  
650 vector exchange. Proceedings of the IEEE International Conference on Computer Vision, pp.  
651 6361-6371.
- 652 Vandaele, R., Aceto, J., Muller, M., Peronnet, F., Debat, V., Wang, C.-W., Huang, C.-T., Jodogne, S.,  
653 Martinive, P., Geurts, P., 2018. Landmark detection in 2D bioimages for geometric  
654 morphometrics: a multi-resolution tree-based approach. Sci. Rep. 8, 1-13.
- 655 Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features.  
656 Proceedings of the 2001 IEEE computer society conference on computer vision and pattern  
657 recognition, CVPR 2001, Kauai, HI, USA, pp. I-I, doi: 10.1109/CVPR.2001.990517.
- 658 Xie, Y., Ji, Q., 2002. A new efficient ellipse detection method. Object recognition supported by user  
659 interaction for service robots, Quebec City, Quebec, Canada, pp. 957-960.

**Table 1:** Name, geographic origin and characteristics of all populations used.

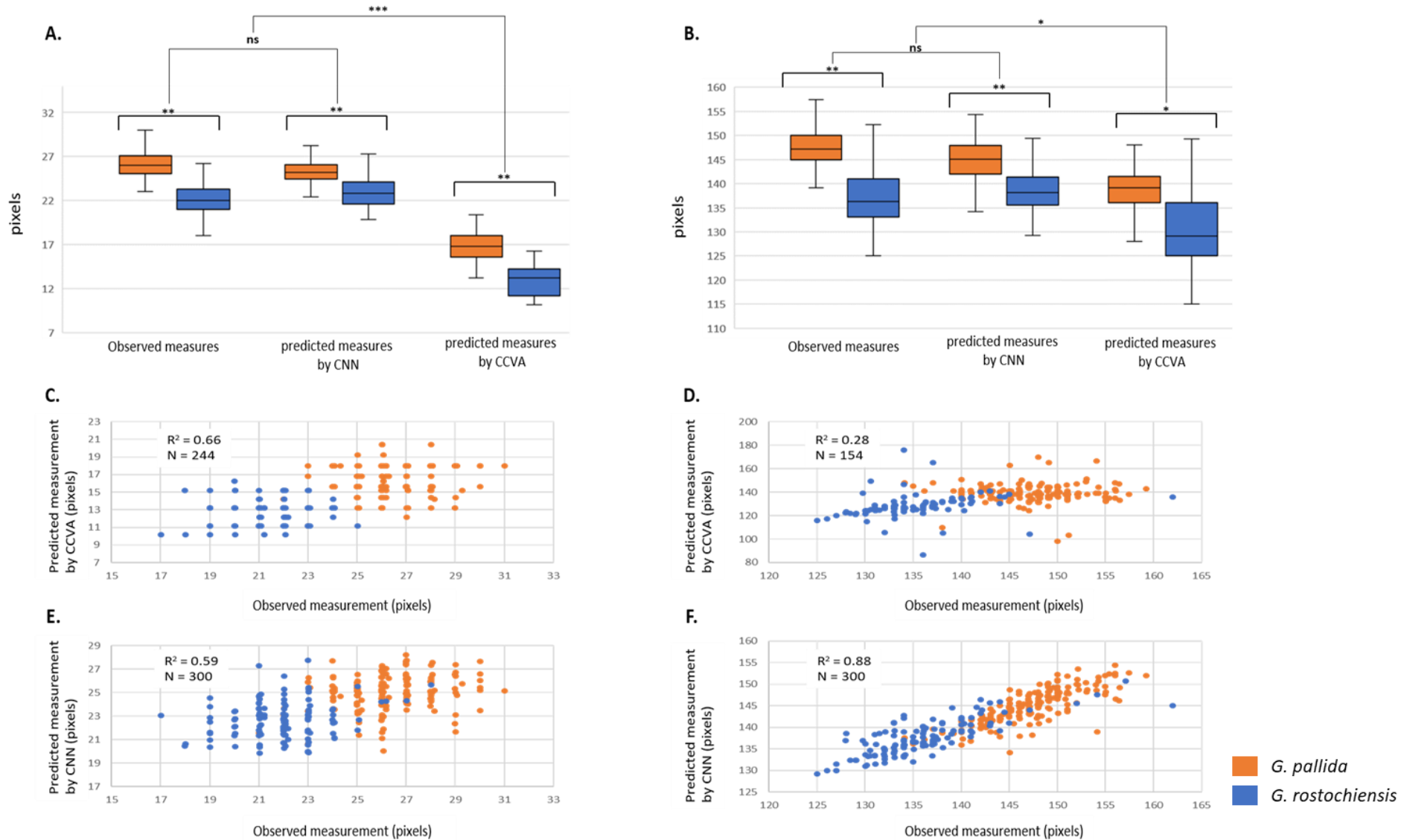
<b>Population</b>	<b>Population code</b>	<b>Species</b>	<b>Origin</b>	<b>Number of individuals</b>	<b>Data set</b>
Dunkerque	Dunk	<i>G. rostochiensis</i>	France	30	Training set
Nimes	Nimes	<i>G. rostochiensis</i>	France	30	Training set
P282	P282	<i>G. rostochiensis</i>	Peru	30	Training set
P312	P312	<i>G. rostochiensis</i>	Peru	30	Training set
Ecosse	Ecos	<i>G. rostochiensis</i>	Scotland	30	Test set
Lindley	Lind	<i>G. pallida</i>	UK	30	Training set
P260	P260	<i>G. pallida</i>	Peru	30	Training set
P273/2016	P273/2016	<i>G. pallida</i>	Peru	30	Training set
P285	P285	<i>G. pallida</i>	Peru	30	Training set
P308	P308	<i>G. pallida</i>	Peru	30	Training set
Rookmaker	Rook	<i>G. pallida</i>	Netherlands	30	Training set
Chavornay	Chav	<i>G. pallida</i>	Switzerland	30	Test set
Tlaxcala	TXL	<i>G. mexicana</i>	Mexico	30	Species test set
Santa Anna	Stan	<i>G. mexicana</i>	Mexico	30	Species test set



**Figure 1.** Illustration of the two new metrics developed for this study. In orange, the length between the base of basal knobs and the head (BKTH). In green, the width of basal knobs (BKW).



**Figure 2:** Processing chain used to extract and analyse measurements. The top and bottom boxes represent processes common to the CCVA and CNN analysis. The main steps of the CCVA and CNN algorithms are indicated in the middle box.



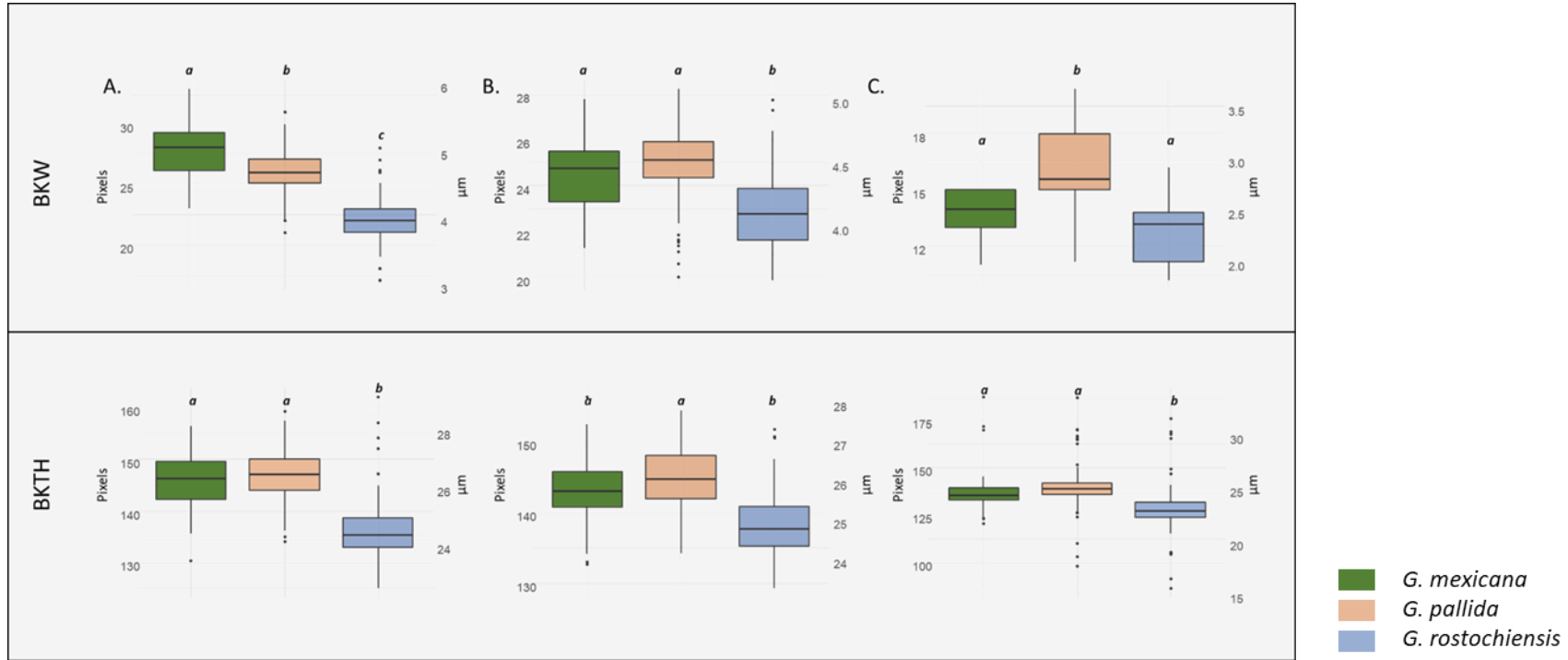
**Figure 3.** Comparison between predicted and observed measurements for *G. pallida* and *G. rostochiensis*. Measurements observed, predicted by CNN and predicted by CCVA are shown in (A) and (B) for the BKW and BKTH metrics, respectively. Correlations between observed and predicted measurements by CCVA are shown in (C) and (E) for the measurements of BKW and BKTH, respectively. Correlations between observed and predicted measurements by CNN are shown in (D) and (F) for the measurements of BKW and BKTH, respectively. The coefficients of correlation ( $R^2$ ) and the number of measurements considered ( $N$ ) are indicated inside each graph.

\*\*\* < 0.001, \*\* < 0.01, \* < 0.05

**Table 2.** Prediction scores obtained at the population scale by using a majority rule assignation. Numbers are the normalised score for the whole population computed on available individuals predicted as *G. pallida*. Therefore, scores less than 0.5 stand for *G. rostochiensis*, conversely, scores greater than 0.5 stand for *G. pallida*. A score of 0.5 means that half of the individuals are predicted as *G. pallida* and half as *G. rostochiensis*

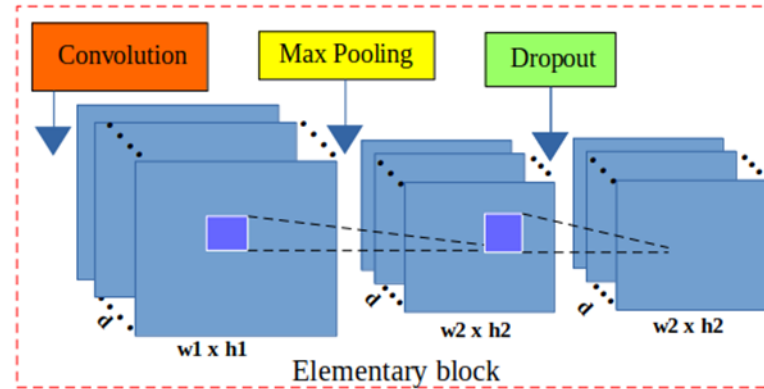
Pop	Reality	Prediction from observed measurements	CCVA Prediction	CNN prediction
Dunk	<i>G. rostochiensis</i>	0.00	0.24	0.23
Lind	<i>G. pallida</i>	1.00	1.00	0.86
Nimes	<i>G. rostochiensis</i>	0.06	0.09	0.20
P260	<i>G. pallida</i>	0.96	0.92	0.80
P273/2016	<i>G. pallida</i>	0.83	0.68	0.76
P282	<i>G. rostochiensis</i>	0.10	0.20	0.36
P285	<i>G. pallida</i>	1.00	0.96	1.00
P308	<i>G. pallida</i>	1.00	1.00	0.93
P312	<i>G. rostochiensis</i>	0.00	0.14	0.10
Rook	<i>G. pallida</i>	0.96	0.95	0.90
<i>Total on training set</i>		<i>Accuracy: 0.96</i>	<i>Accuracy: 0.88</i>	<i>Accuracy: 0.83</i>
Chav	<i>G. pallida</i>		0.83	0.96
Ecos	<i>G. rostochiensis</i>		0.12	0.53
<i>Total on testing set</i>			<i>Accuracy: 0.85</i>	<i>Accuracy: 0.71</i>

Correct prediction at population scale  
 Incorrect prediction at population scale



**Figure 4.** Comparison between three species of the observed and predicted metrics. The top line corresponds to BKW measurements and the bottom line to BKTH measurements. (A) observed metrics, (B) metrics predicted by CNN, and (C) metrics predicted by CCVA. Significant differences were estimated by the ANOVA method and are indicated by the letter on top of each boxplot.







**Annex 1** : Elementary block layers used in the CNN computation

**Table 1:** Name, geographic origin and characteristics of all populations used.

<b>Population</b>	<b>Population code</b>	<b>Species</b>	<b>Origin</b>	<b>Number of individuals</b>	<b>Data set</b>
Dunkerque	Dunk	<i>G. rostochiensis</i>	France	30	Training set
Nimes	Nimes	<i>G. rostochiensis</i>	France	30	Training set
P282	P282	<i>G. rostochiensis</i>	Peru	30	Training set
P312	P312	<i>G. rostochiensis</i>	Peru	30	Training set
Ecosse	Ecos	<i>G. rostochiensis</i>	Scotland	30	Test set
Lindley	Lind	<i>G. pallida</i>	UK	30	Training set
P260	P260	<i>G. pallida</i>	Peru	30	Training set
P273/2016	P273/2016	<i>G. pallida</i>	Peru	30	Training set
P285	P285	<i>G. pallida</i>	Peru	30	Training set
P308	P308	<i>G. pallida</i>	Peru	30	Training set
Rookmaker	Rook	<i>G. pallida</i>	Netherlands	30	Training set
Chavornay	Chav	<i>G. pallida</i>	Switzerland	30	Test set
Tlaxcala	TXL	<i>G. mexicana</i>	Mexico	30	Species test set
Santa Anna	Stan	<i>G. mexicana</i>	Mexico	30	Species test set

**Table 2.** Prediction scores obtained at the population scale by using a majority rule assignation. Numbers are the normalised score for the whole population computed on available individuals predicted as *G. pallida*. Therefore, scores less than 0.5 stand for *G. rostochiensis*, conversely, scores greater than 0.5 stand for *G. pallida*. A score of 0.5 means that half of the individuals are predicted as *G. pallida* and half as *G. rostochiensis*.

Pop	Reality	Prediction from observed measurements	CCVA Prediction	CNN prediction
Dunk	<i>G. rostochiensis</i>	0.00	0.24	0.23
Lind	<i>G. pallida</i>	1.00	1.00	0.86
Nimes	<i>G. rostochiensis</i>	0.06	0.09	0.20
P260	<i>G. pallida</i>	0.96	0.92	0.80
P273/2016	<i>G. pallida</i>	0.83	0.68	0.76
P282	<i>G. rostochiensis</i>	0.10	0.20	0.36
P285	<i>G. pallida</i>	1.00	0.96	1.00
P308	<i>G. pallida</i>	1.00	1.00	0.93
P312	<i>G. rostochiensis</i>	0.00	0.14	0.10
Rook	<i>G. pallida</i>	0.96	0.95	0.90
<i>Total on training set</i>		<i>Accuracy: 0.96</i>	<i>Accuracy: 0.88</i>	<i>Accuracy: 0.83</i>
Chav	<i>G. pallida</i>		0.83	0.96
Ecos	<i>G. rostochiensis</i>		0.12	0.53
<i>Total on testing set</i>			<i>Accuracy: 0.85</i>	<i>Accuracy: 0.71</i>

 Correct prediction at population scale  
 Incorrect prediction at population scale