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

Romain Thevenoux[1,2], Linh Le Van[4], Heloïse Villesseche[1,2,*], Alain Buisson[2], Marie Beurton-Aimar[3], Eric Grenier[1], Laurent Folcher[2], Nicolas Parisey[1]

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

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

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

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

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

Corresponding author: Romain Thevenoux; romain.thevenoux@inrae.fr

Abstract

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

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

1. Introduction

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

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

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

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

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

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

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

2. Materials and methods

2.1. Biological material and image acquisition

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

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

2.2. Manual landmarks and observed metrics

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

2.3. Automatic extraction of morphological metrics

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

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

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

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

2.4. Statistical classification of individuals and populations into species

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

3. Results

3.1. Predicted versus observed metrics

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

and manual landmarks were compared by simple mean comparison (T. test, Figure 3 A, B). For both metrics, there were no significant differences between measurements predicted by the CNN and the observed measurements based on the manual landmarks (Figure 3 A, B). However, the situation appeared to be different regarding the CCVA predicted metrics. For BKW, measurements predicted by the CCVA were significantly smaller than the manual landmarks, but this is clearly linked to the fact that BKW metrics for the CCVA correspond to only the radius of the ellipse. Nonetheless, the two species are still distinguishable (Figure 3A). For BKTH, measurements predicted by the CCVA were also found to be smaller and we hypothesised that this was due to the approximation of the basal knobs position using the centroid of the predicted ellipse, and not the bottom of the ellipse. Nonetheless, a difference between *G. pallida* and *G. rostochiensis* was still observed, and was consistent with the observed metrics. Correlations between predicted and observed measurements were high for all metrics and algorithms, except the prediction of BKTH by CCVA (R2 = 0.28). In most cases, segregation between *G. pallida* and *G. rostochiensis* was clear, although for some individuals, *G. pallida* always had higher measurements and prediction than *G. rostochiensis* (Figure 3 C, D, E, F). Nonetheless, there is still an overlap between the two species for the metrics studied: extreme cases are clear, but mean cases are more confusing. Therefore, there is a need for classifier tools in order to identificate species.

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

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

3.2. Classification using the training set

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

3.3. Classification using the test set

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

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

3.4. Classification using the species test set

There were few differences between the three species that we worked on for the two metrics studied when measurements were done with manual landmarks (Figure 4 A). *Globodera mexicana* showed significantly larger BKW than *G. pallida*, but we observed no significant difference between these two species for the BKTH metric. Importantly, the observed mean difference between *G. mexicana* and *G. pallida* was only 2.5 pixel (0.45 µm) for BKW. BKW was statistically different between the three species studied (Figure 4 A). For the two metrics studied, *G. rostochiensis* had the lower measurements.

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

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

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

For BKW, the observed range of variation (*G. mexicana*: mean = 5.00 µm ± 0.38; *G. rostochiensis*: mean = 3.94 µm ± 0.31; *G. pallida*: mean = 4.74 µm ± 0.31) was similar to the range of variation predicted by CNN (*G. mexicana*: mean = 4.42 µm ± 0.25; *G. rostochiensis*: mean = 4.10 µm ± 0.27; *G. pallida*: mean = 4.51 µm ± 0.25) and by CCVA (*G. mexicana*: mean = 2.44 µm ± 0.25; *G. rostochiensis*: mean = 2.21 µm ± 0.27; *G. pallida*: mean = 2.89 µm ± 0.33). BKW means for CCVA predictions were always lower due to our prediction of the ellipse radius (and not the ellipse diameter) for the basal knob width. However, it must be highlighted that BKW

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

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

4. Discussion and conclusions

4.1. Usefulness of the metrics for *Globodera* taxonomy

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

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

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

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

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

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

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

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

4.3. Interest for systematics in the *Globodera* genus

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

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

Authors contributions

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

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

Acknowledgments

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

Data Archiving Statement

The algorithms used in this study are available at the Zenodo repository for the CCVA (DOI: 10.5281/zenodo.4068341) and at GitHub for the CNN (https://github.com/linhlevandlu/CNN_Beetles_Landmarks).

References

Abd-Elgawad, M.M., Askary, T.H., 2015. Impact of phytonematodes on agriculture economy, In: Askary, T.H., Martinelli, P.R.P. (Eds.), Biocontrol Agents of Phytonematodes. CABI Publishing, Wallingford, UK, pp. 3-49.

Akintayo, A., Tylka, G.L., Singh, A.K., Ganapathysubramanian, B., Singh, A., Sarkar, S., 2018. A deep learning framework to discern and count microscopic nematode eggs. Sci. Rep. 8, 1-11.

Bernard, G.C., Egnin, M., Bonsi, C., 2017. The impact of plant-parasitic nematodes on agriculture and methods of control, Nematology-Concepts, Diagnosis and Control. InTech.

Boucher, A.C., Mimee, B., Montarry, J., Bardou-Valette, S., Belair, G., Moffett, P., Grenier, E., 2013. Genetic diversity of the golden potato cyst nematode *Globodera rostochiensis* and determination of the origin of populations in Quebec, Canada. Mol. Phylogenet. Evol. 69, 75-82.

Campos-Vela, A., 1967. Taxonomy, life cycle and host range of *Heterodera mexicana* n. sp. (Nematoda: Heteroderidae). University of Wisconsin, Madison, WI, USA.

Cintas, C., Quinto-Sánchez, M., Acuña, V., Paschetta, C., De Azevedo, S., de Cerqueira, C.C.S., Ramallo, V., Gallo, C., Poletti, G., Bortolini, M.C., 2016. Automatic ear detection and feature extraction using geometric morphometrics and convolutional neural networks. IET Biometrics 6, 211-223.

Council directive 2016/2031/EC, 2016. Regulation 2016/2031 of the European parliament of the Council of 26 October 2016 on protective measures against pests of plants, amending Regulations (EU) No 228/2013, (EU) No 652/2014 and (EU) No 1143/2014 of the European Parliament and of the Council and repealing Council Directives 69/464/EEC, 74/647/EEC, 93/85/EEC, 98/57/EC, 2000/29/EC, 2006/91/EC and 2007/33/EC. Official Journal of the European Union, L317: 4-104.

Dai, M., Sun, T., Chen, X., Yu, L., Chen, M., Hao, P., Zeng, X., Yan, J., Chen, S., 2019. A B-scan imaging method of conductivity variation detection for magneto–acousto–electrical tomography. IEEE Access 7, 26881-26891.

de la Blanca, N.P., Fdez-Valdivia, J., Castillo, P., Gomez-Barcina, A., 1992. Detecting nematode features from digital images. J. Nematol. 24, 289.

de Vienne, D.M., Refregier, G., Lopez-Villavicencio, M., Tellier, A., Hood, M.E., Giraud, T., 2013. Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. New Phytologist 198, 347-385.

Decraemer, W., Hunt, D.J., 2006. Structure and Classification, In: Perry, R.N., Moens, M. (Eds.), Plant Nematology. CABI Publishing, Wallingford, UK, pp. 153-184.

Dieleman, S., Schlüter, J., Raffel, C., Olson, E., Sønderby, S.K., Nouri, D., Maturana, D., Thoma, M., 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.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676-682.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929-1958.

Subbotin, S.A., Franco, J., Knoetze, R., Roubtsova, T.V., Bostock, R.M., Vera, I.C.d.P., 2020. DNA barcoding, phylogeny and phylogeography of the cyst nematode species from the genus *Globodera* (Tylenchida: Heteroderidae). Nematology 22, 269-297.

Subbotin, S.A., Vera, I.C.D.P., Mundo-Ocampo, M., Baldwin, J.G., 2011. Identification, phylogeny and phylogeography of circumfenestrate cyst nematodes (Nematoda: Heteroderidae) as inferred from analysis of ITS-rDNA. Nematology 13, 805-824.

Thevenoux, R., Folcher, L., Esquibet, M., Fouville, D., Montarry, J., Grenier, E., 2020. The hidden diversity of the potato cyst nematode *Globodera pallida* in the south of Peru. Evol. Appl. 13, 727-737.

Thewlis, J., Albanie, S., Bilen, H., Vedaldi, A., 2019. Unsupervised learning of landmarks by descriptor vector exchange. Proceedings of the IEEE International Conference on Computer Vision, pp. 6361-6371.

Vandaele, R., Aceto, J., Muller, M., Peronnet, F., Debat, V., Wang, C.-W., Huang, C.-T., Jodogne, S., Martinive, P., Geurts, P., 2018. Landmark detection in 2D bioimages for geometric morphometrics: a multi-resolution tree-based approach. Sci. Rep. 8, 1-13.

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition, CVPR 2001, Kauai, HI, USA, pp. I-I, doi: 10.1109/CVPR.2001.990517.

Xie, Y., Ji, Q., 2002. A new efficient ellipse detection method. Object recognition supported by user interaction for service robots, Quebec City, Quebec, Canada, pp. 957-960.

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

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

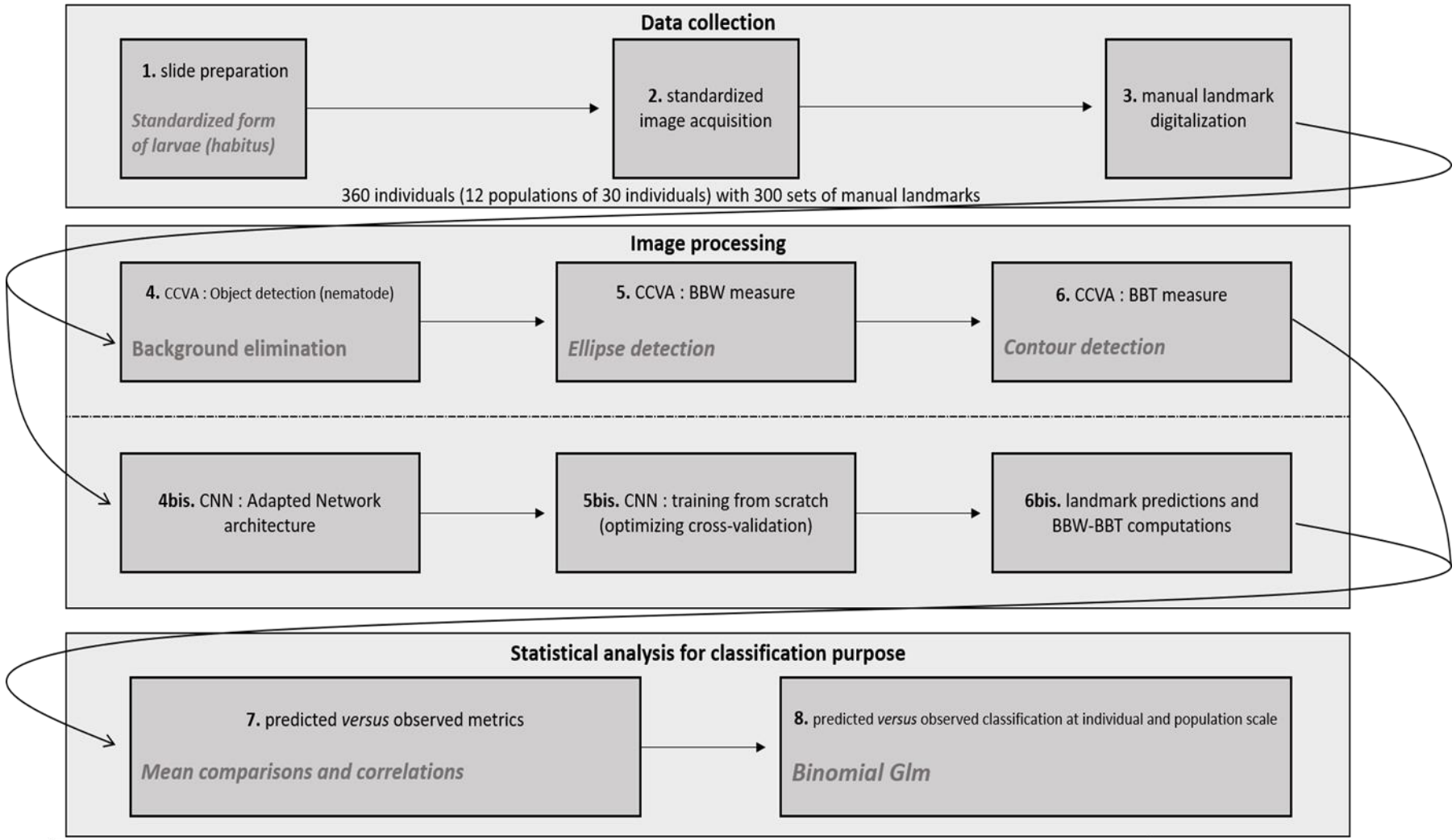


Figure 2: Processing chain used to extract and analyse measurements. The top and bottom boxes represent processes common to the CCVA and CNN analysis. The main steps of the CCVA and CNN algorithms are indicated in the middle box.

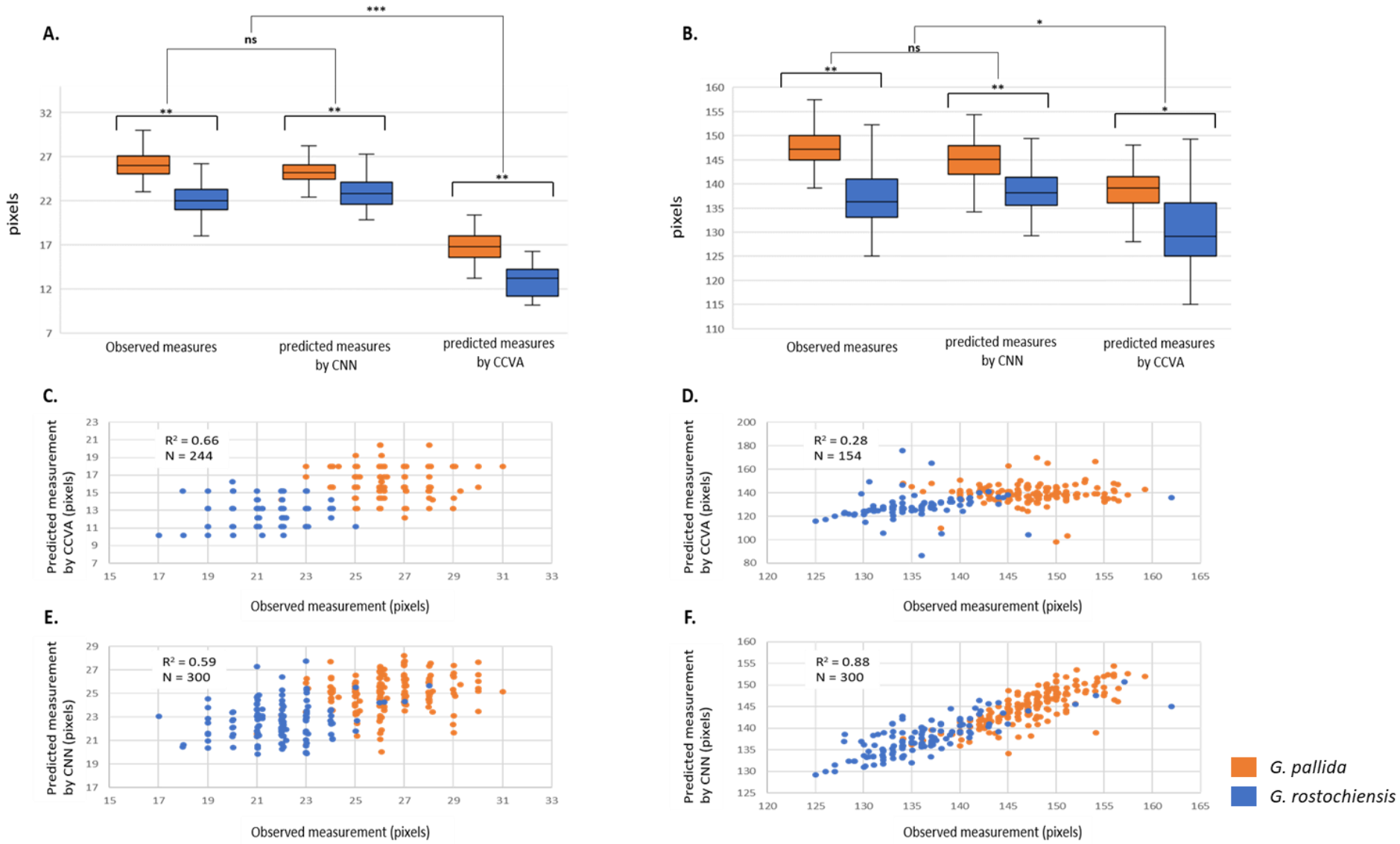


Figure 3. Comparison between predicted and observed measurements for *G. pallida* and *G. rostochiensis*. Measurements observed, predicted by CNN and predicted by CCVA are shown in (A) and (B) for the BKW and BKTH metrics, respectively. Correlations between observed and predicted measurements by CCVA 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 (R2) 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	*G. rostochiensis*	0.00	0.24	0.23
Lind	*G. pallida*	1.00	1.00	0.86
Nimes	*G. rostochiensis*	0.06	0.09	0.20
P260	*G. pallida*	0.96	0.92	0.80
P273/2016	*G. pallida*	0.83	0.68	0.76
P282	*G. rostochiensis*	0.10	0.20	0.36
P285	*G. pallida*	1.00	0.96	1.00
P308	*G. pallida*	1.00	1.00	0.93
P312	*G. rostochiensis*	0.00	0.14	0.10
Rook	*G. pallida*	0.96	0.95	0.90
Total on training set		*Accuracy: 0.96*	*Accuracy: 0.88*	*Accuracy: 0.83*
Chav	*G. pallida*		0.83	0.96
Ecos	*G. rostochiensis*		0.12	0.53
Total on testing set			*Accuracy: 0.85*	*Accuracy: 0.71*

Correct prediction at population scale
Incorrect prediction at population scale

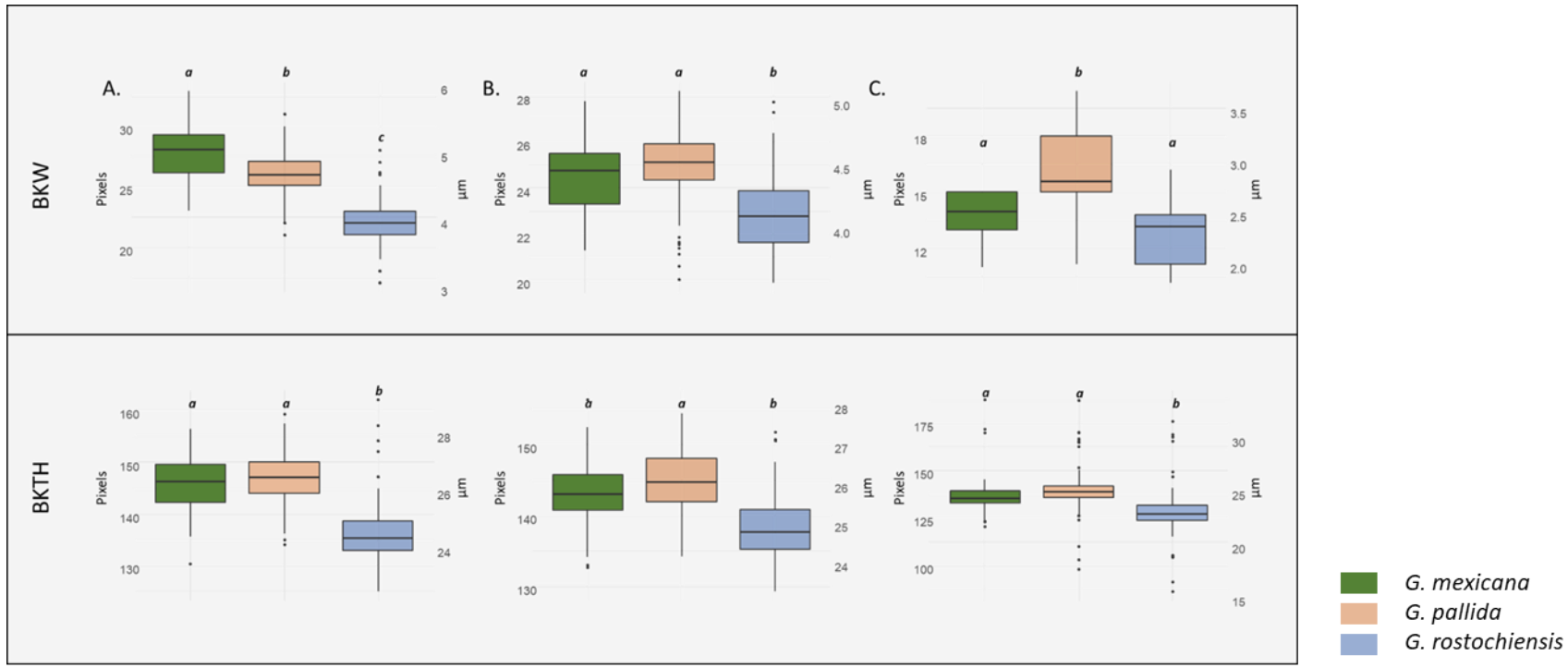
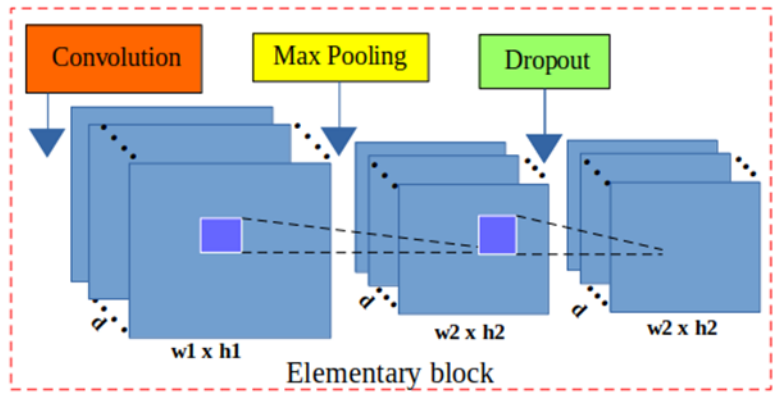


Figure 4. Comparison between three species of the observed and predicted metrics. The top line corresponds to BKW measurements and the bottom line to BKTH measurements. (A) observed metrics, (B) metrics predicted by CNN, and (C) metrics predicted by CCVA. Significant differences were estimated by the 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.

Population	Population code	Species	Origin	Number of individuals	Data set
Dunkerque	Dunk	*G. rostochiensis*	France	30	Training set
Nimes	Nimes	*G. rostochiensis*	France	30	Training set
P282	P282	*G. rostochiensis*	Peru	30	Training set
P312	P312	*G. rostochiensis*	Peru	30	Training set
Ecosse	Ecos	*G. rostochiensis*	Scotland	30	Test set
Lindley	Lind	*G. pallida*	UK	30	Training set
P260	P260	*G. pallida*	Peru	30	Training set
P273/2016	P273/2016	*G. pallida*	Peru	30	Training set
P285	P285	*G. pallida*	Peru	30	Training set
P308	P308	*G. pallida*	Peru	30	Training set
Rookmaker	Rook	*G. pallida*	Netherlands	30	Training set
Chavornay	Chav	*G. pallida*	Switzerland	30	Test set
Tlaxcala	TXL	*G. mexicana*	Mexico	30	Species test set
Santa Anna	Stan	*G. mexicana*	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	*G. rostochiensis*	0.00	0.24	0.23
Lind	*G. pallida*	1.00	1.00	0.86
Nimes	*G. rostochiensis*	0.06	0.09	0.20
P260	*G. pallida*	0.96	0.92	0.80
P273/2016	*G. pallida*	0.83	0.68	0.76
P282	*G. rostochiensis*	0.10	0.20	0.36
P285	*G. pallida*	1.00	0.96	1.00
P308	*G. pallida*	1.00	1.00	0.93
P312	*G. rostochiensis*	0.00	0.14	0.10
Rook	*G. pallida*	0.96	0.95	0.90
Total on training set		*Accuracy: 0.96*	*Accuracy: 0.88*	*Accuracy: 0.83*
Chav	*G. pallida*		0.83	0.96
Ecos	*G. rostochiensis*		0.12	0.53
Total on testing set			*Accuracy: 0.85*	*Accuracy: 0.71*

Correct prediction at population scale

Incorrect prediction at population scale