ARPHA Preprints

PENSOFT

PREPRINT

Author-formatted, not peer-reviewed document posted on 01/04/2024

DOI: https://doi.org/10.3897/arphapreprints.e124163

Low-cost, high-volume imaging for entomological digitization

Dirk Steinke, Jaclyn McKeown, Allison Zyba, Joschka McLeod, Corey Feng, Paul Hebert

For submission to Zookeys

Low-cost, High-volume Imaging for Entomological Digitization

Steinke D[1,2*], McKeown JTA[1*], Zyba A[1], MacLeod J[1], Feng, C[1], Hebert PDN[1,2]

Affiliations:

[1]Centre for Biodiversity Genomics, University of Guelph, 50 Stone Road East, Guelph, Ontario, N1G 2W1, Canada

[2]Department of Integrative Biology, University of Guelph, 50 Stone Road East, Guelph, Ontario, N1G 2W1, Canada

*these authors contributed equally to the work

Corresponding author: Dirk Steinke (dsteinke@uoguelph.ca)

Abstract

Large-scale digitization of natural history collections requires the automation of image acquisition and processing. Reflecting this fact, various approaches, some highly sophisticated, have been developed to support the imaging of museum specimens. However, most of these systems are complex and expensive, restricting their deployment.

Here we describe a simple, inexpensive technique for imaging arthropods larger than 5 mm. By mounting a digital SLR camera on a CNC (computer numerical control) motor-drive rig, we created a system that captures high-resolution z-axis stacked images (6960 x 4640 pixels) of 95 specimens in 30 minutes.

This system is inexpensive ($1000 USD without camera), easy to set up, and to maintain. By coupling low cost with high production capacity, it represents a solution for the digitization of any natural history collection.

Introduction

Advances in computational and imaging technologies have stimulated the digitization of specimens in natural history collections (Beaman and Cellinese 2012; Berents et al. 2010; Hedrick et al. 2020; Holovachov et al. 2014; Hudson et al. 2015; Mantle et al. 2012; Mathys et al. 2015; Mertens et al. 2017; Moore 2011; Vollmar et al. 2010). Because the largest collections contain millions of specimens, comprehensive digitization can be a challenge (Blagoderov et al. 2012). This is particularly true for insects as they dominate most zoological collections (Tegelberg et al 2014). Consequently, many digitization projects have only captured high-resolution images of a few representative specimens of each species (deWaard et al. 2019).

Projects which seek to digitize entire insect collections require automated image acquisition and processing. Because of the effort in handling individual specimens and the risk of damaging them, some digitization programs have imaged drawers of specimens (Blagoderov et al. 2012; Mantle et al. 2012; Holovachov et al. 2014). This approach has three limitations. First, resolution is often insufficient to allow examination of some morphological traits. Second, dorsal images are captured, so characters only visible with a lateral or ventral view are inaccessible. Third, this approach brings informatics challenges as the drawer image must be decomposed into its component specimen images (Blagoderov et al. 2012; Holovachov et al. 2014). In practice, most of the time required in these digitization projects is spent on image selection and metadata capture (Blagoderov et al. 2012).

Recently, several approaches have been developed to digitize individual specimens in museum collections (Heerlien et al. 2015; Ströbel et al. 2018; Tegelberg et al. 2017) or as part of community sampling and sorting procedures (Ärje et al. 2020; Wührl et al. 2021). Some of these systems generate several images per specimen to facilitate 3-D modelling (Ströbel et al. 2018) or include

robotic handling of specimens to accelerate processing (Ärje et al. 2020; Wührl et al. 2021). At this time, most of these systems are elaborate and expensive.

Optimal high-throughput specimen digitization requires combining technologies in novel workflows and is largely driven by purpose (collection digitization versus one component of a multifaceted workflow). This study introduces an imaging system developed to support the specimen-centric workflow employed by the Centre for Biodiversity Genomics (CBG) to gather DNA barcode records. Because images are essential to validate DNA barcodes, the CBG photographs every specimen. Small specimens (<5 mm) are each placed into a well in a 96-well plate and are imaged with a high-resolution automated microscope system (Steinke et al. in prep) before entering molecular analysis. Larger individuals are pinned, arrayed in Schmidt boxes, and then imaged using the SLR rig described here.

Material and Methods

System hardware

The SLR rig (Figure 1) employs a Canon 90D camera (24.5 megapixel) with an EF-S 60mm f/2.8 Macro USM lens (Canon Corp, Irvine, CA, USA). The camera is attached to an OpenBuilds Acro 1010 40” x 40” motor drive rig (OpenBuilds, New York, NY, USA). Most components for the OpenBuilds Acro 1010 were purchased, but some components were printed (i-Fast; QIDI Technology Official, China) using 3D-models available on the OpenBuilds website. The OpenBuilds Acro 1010 frame is screwed onto a wooden base which is placed on a sturdy table (79 cm from floor). Two 60x60 cm softbox lights (Neewer 24x24, Shenzhen) are stationed to the left and right of the SLR camera motor-drive rig at a height of 165 cm on their stands. A LED light strip is positioned along the circumference of the Acro system facing inwards (Daylight White LED Strip Light; Shenzhen Intellirocks Tech Co. Ltd., Shenzhen, China). The SLR camera is

mounted 16 cm above the table's surface and a Kimaru ACK-E6 DR-E6 DC Coupler LP-E6N Dummy Battery AC Power Adapter Kit is used to provide constant power. A USB-A to micro-USB cable connects the camera to the computer (iMac 27-inch, 8GB Ram. 3.4GHz Quad-core Intel-Core i5).

Software control

The SLR rig is controlled by a program that employs both a python script (Supplementary File 1) and a G-code script (Supplementary File 2) through an Apple iMac operating system. The system is manually calibrated using OpenBuilds Control (OpenBuilds) integrated into the python script. The camera is controlled using the command line tool Gphoto2 (gphoto2.org). It is set to an ISO value of 100, an aperture value of f/8, and a shutter speed of 1/8s. Focus bracketing to allow Z-axis stacking is set to take nine images at different levels of focus. The lens is set to autofocus, so the first image captures the uppermost of a specimen before eight more images are automatically captured at lower focal planes to allow z-axis stacking. The nine images are combined to generate a single composite in-focus image using Helicon Focus 8 (Helicon Soft Ltd, Kharkov, Ukraine).

Operation

Pinned arthropods are loaded into the SLR rig in batches of 95 after being transferred to a 75 cm by 47 cm foam platform (Figure 2). This platform has 95 positions for specimen loading, each a slot with a depth of 1.8 cm, a length of 6 cm, and a width of 5 cm. This count ensures that all specimens in each 96 well micro-plate (95 specimens, 1 negative control) are processed as a batch. The platform is split into 8 rows, each with 12 slots (Figure 2). Each pinned specimen is placed centrally in a slot where it can be positioned for dorsal or lateral imaging (Figure 2). Each row has

a foam strip (height = 2 cm) that facilitates lateral imaging. Specimens stored in envelopes are removed from them and placed centrally at the base of a slot.

Data handling

Gphoto2 is used to transfer images to the computer for further processing. Z-stacked images are cropped to standardised dimensions with a 4x3 aspect ratio using a machine-learning-based cropping tool (Gharaee et al. 2023). A scale bar is added to each image using the batch action tool of Photoshop 2023 (Adobe Inc, San Jose, USA). Once edited, images are uploaded to the Barcode of Life Database System (BOLD) (Ratnasingham & Hebert 2007). A python script (Supplementary File 3) generates a metadata file and compresses it together with packets of 95 images into a zip folder to meet BOLD's requirements for image upload.

Results/Discussion

Performance and costs

Figure 3 shows a selection of specimens and their sizes. When using a Canon 90D with the described settings, the resulting image is 6960 x 4640 pixels before cropping. This translates into an average size of 9.5MB for a jpg-file.

The CNC motor activates at a pre-set time which depends upon the distance between each slot and the dwell time (3.75 seconds) at each stop which allows the camera to take nine images and transfer them to the computer. Operating in this mode, the SLR rig images 95 specimens in 30 minutes and the stacking software requires another 11.5 minutes to process these images, but this usually occurs while the next batch is being photographed. The time required to crop and edit each batch varies (15-30 min) with the type of specimens. Operated by one staff member, the SLR rig

can image 4,000 specimens in a week. The CBG's system has now imaged more than 250,000 specimens and the sole maintenance involved the replacement of a wire leading to one motor.

The SLR rig costs of $4500 are reflecting three main components: 1) CNC machine kit ($1000), 2) Apple computer ($1000), and 3) Canon 90D including lens ($2500). Costs can be reduced by replacing the computer with a raspberry pi ($100), but under heavy usage (40h a week), it will need replacement every six months. Less expensive cameras can be used if they can be controlled with gphoto2. They do have lower resolution (12-20 megapixels) than the Canon 90D (24.5 megapixels), but this resolution is adequate for many applications. However, it is important to select a camera with a depth stacking function such as focus bracketing (e.g., Canon PowerShot G7 X Mark II, $900; Olympus OM-D E-M1 Mark III, $920 for body). By careful selection of components, the overall cost can be reduced to about $2000. Plastic components for the OpenBuilds Acro 1010 are freely available as 3D models so users can modify and 3D-print custom components. One modification made to our SLR rig was the addition of bumpers and a triangular structure to improve wire management during operation (Supplementary files 4 & 5).

Adjustments

The SLR rig can image a wide variety of specimens by adjusting settings as described in this section. The distance between the camera and specimen dictates the size of the image (focal distance from base = 16 cm). The frame size varies by 0.5 cm in both directions depending upon the depth of the focus point determined by the auto focus program. This limits the size range of specimens which can be imaged (5 – 45 mm). As each slot on the platform is designed to fit the camera frame, no specimen should overlap an adjacent space because the cropping tool is likely to malfunction. However, larger specimens can be imaged if the distance between the camera and

specimen is increased as this enlarges the size of the frame. Conversely, reducing the camera-specimen distance decreases the size of the frame, allowing smaller specimens (down to 2 mm using the 60mm macro lens in our set up) to be imaged. Any change in the camera's operating height is difficult with the described setup as it requires remounting the camera at a higher or lower position on its mount or the exchange of the legs mounted to the rig frame. Future optimization could incorporate legs capable of height adjustment.

Background colour and light settings can also be modified to improve image quality. Dark backgrounds improve the contrast for dark specimens, helping to highlight otherwise subtle features and also help to contrast pale specimens that blend into a white background. To make this adjustment, a second platform can be made of dark foam, or dark strips can be temporarily added to the existing platform. As lighting and whitening settings on the camera must be adjusted to accommodate the change in background colour, all 95 slots must have the same background.

The number of images taken of each specimen can be adjusted with the depth stack function on the camera. Increasing the image count expands the depth range in focus, but increases the time required to capture photographs and to process them in Helicon Focus. The dwell time of the CNC motor system does needs to be extended to allow more images to be taken before the camera moves. Conversely, imaging and processing times can be reduced by reducing the number of images taken per specimen. Experimentation with sets of specimens in the target group is the best way to optimize the number of images taken.

Although this CBG's SLR system is primarily used with pinned insects, it is effective in imaging other specimens (e.g., soft-bodied invertebrates in liquid preservative). In the latter case, the foam platform is simply replaced by a grid structure that holds each specimen vial (Mendez et al. 2018). Focal distance and stack depth often need to be adjusted in such case (Mendez et al. 2018).

178

179 *Limitations*

180 Generating an image with enough resolution to allow species identification can be difficult with
181 any automated system given the manifold differences in shape and size of specimens (Blagoderov
182 et al. 2012). Very large individuals that exceed the standard stacking depth can cause the autofocus
183 program to return an out-of-focus image. The autofocus function is also vulnerable to vertical
184 protrusions, especially if they contrast with the background. In such cases, the depth stack may
185 begin above the organism's body plane leading to a blurred image. For winged insects, such as
186 Lepidoptera, variation in wing orientation can lead the wingspan to exceed the range of the image
187 stack. In such cases, the resulting image may show a focused wing with an out-of-focus body or
188 an in-focus body with a blurred wing. In such cases, a slight change in the angle at which a
189 specimen is positioned can greatly improve image quality but a switch from lateral to dorsal view
190 is sometimes required. Because reorienting a few specimens requires recapturing an entire set of
191 images, it is often more time effective to simply accept few imperfect images (Chapman 2005, Ahl
192 et al. 2023).

193 At the CBG, specimens are usually imaged before they are labelled. When labelled specimens
194 are imaged, a small white piece of paper with a slit in the middle is used to cover labels, allowing
195 images of small specimens to remain sharp when cropped. Alternatively, the labels can be removed
196 and reattached to the specimen after photography.

197

198 **Conclusion**

199 The present SLR rig was designed to photograph terrestrial arthropods that were being analyzed
200 to construct DNA barcode reference libraries. About 90% of these specimens are small enough to
201 be imaged within 96-well plates, but the remainder must be pinned. As the CBG currently barcodes

three million specimens annually, it was essential to develop a system capable of imaging the larger specimens in a cost-effective way. This led to the present solution which can be acquired for $2000-4500 depending on choice of camera and controller and generates almost 200 high-resolution specimen images per hour.

As the CBG's SLR rig has performed reliably for 2.5 years of heavy use (12h/day), this system is ideal for deployment in settings remote from technical support. Because of its capacity to rapidly generate large numbers of high-quality digital images for online databases, it is also an asset for any large specimen collection.

Acknowledgements

This study was enabled by awards to PDNH from the Ontario Ministry of Economic Development, Job Creation and Trade, the Canada Foundation for Innovation, and by a grant from the Canada First Research Excellence Fund to the University of Guelph's "Food From Thought" research program. We thank Suzanne Bateson for aid with graphics.

References

Ärje J, Melvad C, Rosenhøj Jeppesen M, Agerskov Madsen S, Raitoharju J, Strandgård Rasmussen M, Iosifidis A, Tirronen V, Gabbouj M, Meissner K, Høye TT (2020). Automatic image-based identification and biomass estimation of invertebrates. Methods in Ecology and Evolution 11(8): 922-931.

Beaman RS, Cellinese N (2012). Mass digitization of scientific collections: New opportunities to transform the use of biological specimens and underwrite biodiversity science. ZooKeys 209: 7-17.

Berents P, Hamer M, Chavan V (2010). Towards demand-driven publishing: Approaches to the prioritization of digitization of natural history collection data. Biodiversity Informatics 7: 113-119.

Blagoderov V, Kitching IJ, Livermore L, Simonsen TJ, Smith VS (2012). No specimen left behind: industrial scale digitization of natural history collections. ZooKeys 209: 133-146.

Chapman A (2005). Principles of Data Quality. Global Biodiversity Information Facility https://doi.org/10.15468/doc.jrgg-a190

deWaard JR, Ratnasingham S, Zakharov EV, Borisenko AV, Steinke D, Telfer AC, Perez KHJ, Sones JE, Young MR, Levesque-Beaudin V, Sobel CN, Abrahamyan A, Bessonov K, Blagoev G, deWaard SL, Ho C, Ivanova NV, Layton KKS, Lu L, Manjunath R, McKeown JTA, Milton MA, Miskie R, Monkhouse N, Naik S, Nikolova N, Pentinsaari M, Prosser SWJ, Radulovici AE, Steinke C, Warne CP, Hebert PDN (2019). A reference library for Canadian invertebrates with 1.5 million barcodes, voucher specimens, and DNA samples. Scientific Data 6: 308.

Gharaee Z, Gong Z, Pellegrino N, Zarubiieva I, Haurum JB, Lowe SC, McKeown JTA, Ho CCY, McLeod J, Wei YC, Agda J, Ratnasingham S, Steinke D, Chang AX, Taylor GW, Fieguth P (2023). A step towards worldwide biodiversity assessment: the BIOSCAN-1M insect dataset. Advances in Neural Information Processing Systems 37.

Heerlien M, van Leusen J, Schnörr S, de Jong-Kole S, Raes R, van Hulsen K (2015). The natural history production line: An industrial approach to the digitization of scientific collections. ACM Journal on Computing and Cultural Heritage. 8, 1, Article 3.

Hedrick BP, Heberling JM, Meineke EK, Turner KG, Grassa CJ, Park DS, Kennedy J, Clarke JA, Cook JA, Blackburn DC, Edwards SV, Davis CC (2020). Digitization and the future of natural history collections, BioScience, 70(3): 243-251

Holovachov O, Zatushevsky A, Shydlovsky I (2014). Whole-drawer imaging of entomological collections: Benefits, limitations and alternative applications. Journal of Conservation and Museum Studies 12(1): 9.

Hudson LN, Blagoderov V, Heaton A, Holtzhausen P, Livermore L, Price BW, van der Walt S, Smith VS (2015). Inselect: Automating the digitization of natural history collections. PLoS one 10: e0143402.

Ahl LI, Bellucci L, Brewer P, Gagnier P-Y, Hardy HM, Haston EM, Livermore L, De Smedt S, Enghoff H (2023). Digitisation of natural history collections: criteria for prioritization. Research Ideas and Outcomes 9: e114548.

Mantle BL, Salle JL, Fisher N (2012). Whole-drawer imaging for digital management and curation of a large entomological collection. ZooKeys 209: 147-163.

Mathys A, Brecko J, Semal P (2013). Comparing 3D digitizing technologies: What are the differences? Digital Heritage International Congress. Marseille, 201-204.

Mendez PK, Lee S, Venter CE (2018). Imaging natural history museum collections from the bottom up: 3D print technology facilitates imaging of fluid-stored arthropods with flatbed scanners. ZooKeys 795: 49-65.

Mertens JEJ, Roie MV, Merckx J, Dekoninck W (2017). The use of low-cost compact cameras with focus stocking functionality in entomological digitization projects. ZooKeys 712: 141-154.

Moore W (2011). Biology needs cyber-infrastructure to facilitate specimen-level data acquisition for insect and other hyperdiverse groups. ZooKeys 147: 479-486.

Ratnasingham, S., and P. D. N. Hebert. 2007. BOLD: the barcode of life data system (http://www.barcodinglife.org). Molecular Ecology Notes 7:355-364.

Ströbel B, Schmelzle S, Blüthgen N, Heethoff M (2018). An automated device for the digitization and 3D modelling of insects, combining extended-depth-of-field and all-side multi-view imaging. ZooKeys 759: 1-27.

Tegelberg R, Mononen T, Saarenmaa H (2014). High-performance digitization of natural history collections: Automated imaging lines for herbarium and insect specimens. Taxon 63(6): 1307-1313.

Tegelberg R, Kahanpää J, Karppinen J, Mononen T, Wu Z, Saarenmaa H (2017). Mass digitization of individual pinned insects using conveyor-driven imaging. 2017 IEEE 13th International Conference on e-Science (e-Science), 2017, pp. 523-527.

Vollmar A, Macklin JA, Ford LS (2010). Natural history specimen digitization: Challenges and concerns. Biodiversity Informatics 7: 93-113.

Wührl L, Pylatiuk C, Giersch M, Lapp F, von Rintelen T, Balke M, Schmidt S, Cerretti P, Meier R (2021). DiversityScanner: Robotic discovery of small invertebrates with machine learning methods. BioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444523

287 **Figure legends**

288

289 **Figure 1:** The SLR rig is placed on a heavy base table to minimize vibration. The inset shows the
290 actual rig area with specimens on the styrofoam base.

291

292 **Figure 2:** 75 cm by 47 cm foam platform with pinned insects in dorsal and lateral positions.

293

294 **Figure 3**: Panel of example images taken with the SLR rig.

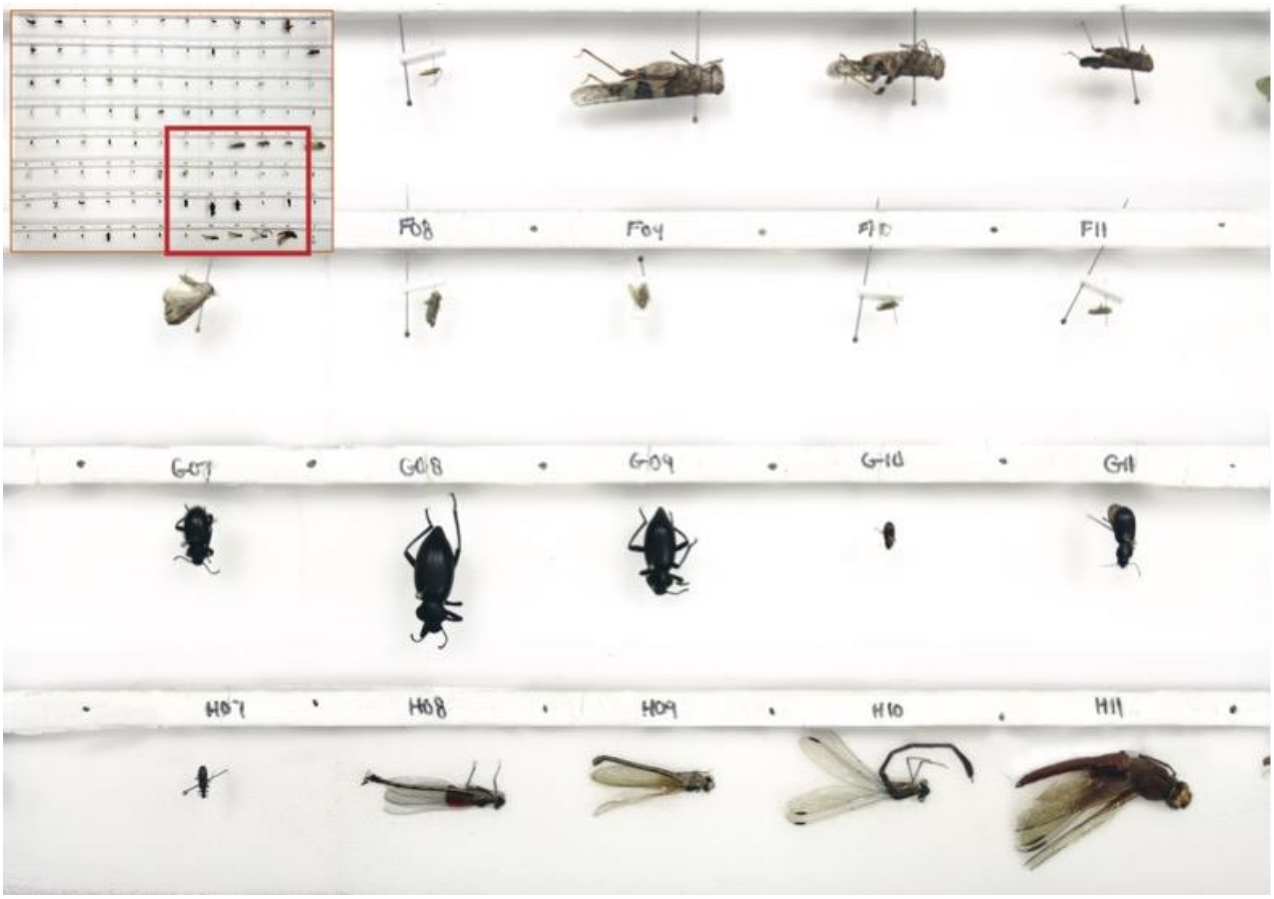
295

296

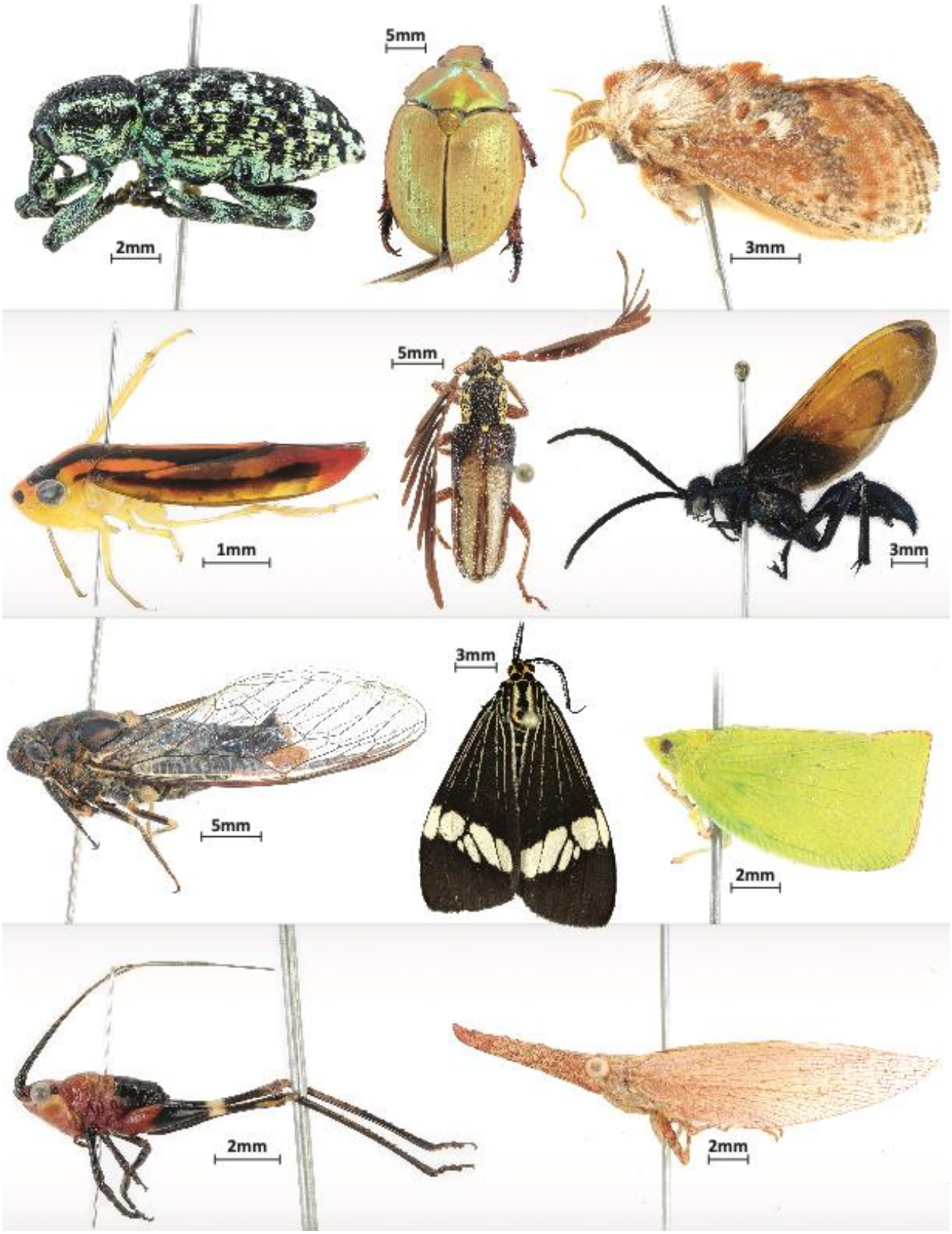
297

Figure 1

Figure 2



305 **Figure 3**



306

307