

Digitizing the Culicidae collection of Naturalis Biodiversity Center, with a special focus on the former Bonne- Wepster subcollection

Pasquale Ciliberti¹, Astrid Roquas¹, Becky Desjardins¹, Bibiche Berkholst¹, Frank Loggen¹,
Menno Hooft¹, Gideon Gijswijt¹ and Dick de Graaff¹

¹ Naturalis Biodiversity Center, Darwinweg 2 2333 CC Leiden

Corresponding author: e-mail:pasquale.ciliberti@naturalis.nl

Abstract

Natural history collections contain a huge amount of information on diversity, distribution and ecology of species. Unfortunately, this valuable information is not always available to researchers due to historical and practical constraints. The project we describe aims especially to unlock data handwritten in notebooks owned by Culicidae researcher Johanna Bonne-Wepster. The handwritten notes refer to specimens labeled only with a number. The notebooks were scanned and entered in a Google-Spreadsheet. The specimens were provided with a unique identifier, labeled with the information contained in the notebooks and the data exported to GBIF. In addition, type specimens were photographed.

Beside the collection owned by Johanna Bonne-Wepster, mosquitoes from the former Rijksmuseum voor Natuurlijk Historie collection and the former Zoölogisch Museum Amsterdam Nederland collection were digitized. All specimens are now housed at Naturalis Biodiversity Center museum in Leiden.

This paper describes the efforts made to mobilize the data and the problems we encountered.

Keywords: field books, scans, Google spreadsheet, mobilizing data, coordinates, Culicidae

Data description

We describe a dataset of 55706 records of mosquitoes, mainly collected and observed in the former Dutch colonies of Indonesia and Suriname. The dataset refers primarily to mosquitoes identified and owned by Johanna Bonne-Wepster. This collection is now maintained by Naturalis Biodiversity Center (NBC), in Leiden, The Netherlands.

The Bonne-Wepster mosquitoes were collected by staff of the Centraal Militair Geneeskundig Laboratorium (CMGL), by fellow mosquitoes taxonomists and by Bonne-Wepster herself. In addition, many mosquitoes were collected by the South East Asian Mosquito Project (SEAMP), an international collaboration of the American and Dutch armies.

The mosquitoes collected within the SEAMP project were maintained and curated at the Instituut voor Tropische Hygiëne (ITH) in Amsterdam, the Netherlands. The collecting dates range from the year 1917 until the year 1961. The Bonne-Wepster collection was transferred during the seventies to the Rijksmuseum voor Natuurlijk Historie in Leiden (RMNH).

The information about the specimens, including locality data and identifications were stored in 8 field books, 5 from CMGL, and 3 from ITH. These field books used the same numerical order, so a given registration number could refer to two separate specimens. In the collection, the specimens were labeled only with a number. Due to the age of the field books

we feared losing the information they contained. Therefore the field books were scanned, specimens were provided with labels and the data was entered in a database. Before we started our project, we found that some mosquitoes with a Bonne-Wepster number were already provided with a locality label and species label. Most of the CMGL mosquitoes are missing, therefore those specimens could only be digitized as observation records (40705). In addition 1216 mosquitoes of the former Zoölogisch Museum Amsterdam Nederland (ZMAN) collection and 2388 mosquitoes of the former RMNH collection were digitized as well. All specimens belong to the NBC collection (table 1).

Former owner	Structure of collection	Number of specimens	Historical notes
Bonne-Wepster	Formed by former CMGL collection and former ITH collection and her own sampled specimens.	52102 of which 40705 records are observation only for the missing CMGL specimens.	Since the 70's of last century part of the RMNH collection and since 2011 part of the NBC collection
RMNH		2388	Since 2011 part of NBC collection
ZMAN		1216	Since 2011 part of NBC collection

Table1. Composition of the specimens digitized during this project. For abbreviations see text.

Context

Natural history collections are a rich source of data that can be used for scientific research, education and for the general public [1]. Traditionally specimens were collected for private use in “cabinets of curiosity” and subsequently adopted by musea where they became the focal point of taxonomic research [2].

In recent years the use of Natural history collections has embraced other fields of the biological sciences in addition to taxonomy. Despite sampling biases Natural History Collections can be used to model current and past distributions [3], build molecular libraries [4], to analyze biodiversity for example for conservation purposes [5], to assess hybridization and speciation events [6] and to predict reemergence of diseases [7].

The Natural History Collection of NBC contains about 42 million objects sampled over the past 200 years [8]. A small part of this collection is formed by mosquitoes. Mosquitoes, family Culicidae, are among the most feared insects for their vector role in transmitting a wide array of pathogens, the most notorious being protozoans of the genus *Plasmodium*, the causative agents of malaria [9]. According to the World Health Organization 619.000 people died of malaria in 2021 [10].

Most of the mosquito specimens now housed at NBC were owned and identified by Johanna Bonne-Wepster. The collection was used to study the morphological features that characterize mosquito species. The specimens were provided with labels having a handwritten number under it (fig.1). These numbers were linked to information contained in 8 different field books (fig.2). The data was organized in columns indicating the species name, a description of the collecting place, the collector and sometimes short taxonomical notes (fig.3). The field books of Johanna Bonne-Wepster are very old and we feared losing the data and consequently losing the value of the specimens.

In fact as beautifully stated by Lane [11], a specimen separated from a label has no scientific value:

"Together, a preserved organism and its label are a scientific specimen that has great intrinsic value. Separately, the label is a piece of paper with meaningless inscriptions upon it, and the plant, spider, microbe, mushroom, or bird, though carefully preserved, is just so much dead organic matter."
(Lane, 1996: 536)

Hence we strongly felt the need to “rescue” the information in the field books and to link this information to the specimens.

Johanna Bonne-Wepster

Johanna Wepster was born in 1892 in The Hague. She was trained as a teacher. She married Cornelis Bonne, who was two years older and trained as a physician, specializing as a parasitologist and with an interest in tropical medicine. Together they went to the Dutch overseas colonies.

Bonne worked in Surinam first, and in 1927, they headed for the Dutch East Indies. He studied mosquitoes, and especially their role in spreading pathogens.

Mrs. Bonne-Wepster started collecting mosquitoes at a large scale, and learned to identify many several species of Culicidae. In both Surinam and the Dutch East Indies, she did thorough investigations about this family of blood sucking insects. Her main goal was to give non-taxonomists the means to recognize vector species.

Mrs. Bonne-Wepster wrote and co-wrote several publications on the taxonomy of Culicidae [12,13,14,15,16,17]. During the twenties and thirties she collected thousands and thousands of specimens. Many adults, but also the eggs and larvae she kept and identified.

In addition the couple worked as contributors for the The South East Asian Mosquito Project (SEAMP), an international collaboration between American and Dutch militaries. It was an important research project that lasted until after the Second World War. After their return to The Netherlands in 1948, Cornelis Bonne died. Johanna continued researching Culicidae primarily at the University of Amsterdam. Even though she was not academically trained, her enormous contribution to the field of Culicidae taxonomy did not go unnoticed. She received, an honorary doctorate from the University of Amsterdam in 1951.

In the seventies her collection was transferred to RMNH in Leiden (table 1). Initially the museum obtained only the sampled specimens, not the field books with elaborate fieldnotes.

117 The field books appeared to be lost until a curator of the RMNH made a visit to the elder
118 Mrs. Bonne-Wepster and managed to retrieve them.
119 Mrs. Bonne-Wepster passed away in 1978.

120

121 MATERIAL AND METHODS

122 **Purpose:** The main purpose of the project was to mobilize data contained in the field books
123 of Bonne-Wepster and corroborate it with the associated specimens. As we proceeded with
124 the project, mosquitoes of the former ZMAN collection and the RMNH collection were
125 digitized as well.

126

127 **Sampling description:** All specimens were carefully checked, investigated and provided
128 with unique registration numbers (fig.4). This project produced a total of 55706 records.
129 52102 records refer to the former Bonne-Wepster collection of which 40705 are the missing
130 CMGL specimens which were digitized as observation records. 2388 records refer to
131 mosquitoes of the RMNH Culicidae collection, 1216 records refer to specimens of the former
132 ZMAN Culicidae collection (table 1).

133

134 **Step description:**

135

- 136 1) First step was to enter the data of the field books in a Google Spreadsheet.
- 137 2) After the data was entered in the spreadsheet the field books were scanned and
138 stored according to the Naturalis Archive protocol.
- 139 3) The entered records were georeferenced using the Point Radius method as
140 described by Wieczorek et al. [18] (see Coordinates).
- 141 4) Unique registration numbers with a corresponding QR-code were added to the
142 specimens. When the specimens were only labeled with a Bonne-Wepster number,

additional labels were printed out and added. Those labels were a) a locality label, b) a species name label and c) an “ex. coll. Bonne-Wepster” label (fig.4).

5) The collection contained about 50 (holo)types, which were photographed (fig.5) using a Zeiss Discovery.V 12 modular stereo microscope equipped with an AxioCam Mrc 5 camera. Depending on the state of the specimen photographs were taken of the habitus (dorsal and lateral views). A scale bar in micrometers was added to the photo. The labels accompanying the (holo)types were photographed using a Nikon camera D600 equipped with an AF Micro-Nikkor 60mm f/2.8D lens. An overhead camera setup was used so the labels could be photographed from above.

6) The Google spreadsheet was converted into a standardized sheet-format, applicable for the import into the NBC database. In the following table we present an overview of the specimen specific information included in this sheet.

Number Type	Indicates the origin of a certain number. This translates to 'old registration number' for all records.
Number Value	Value of the Bonne-Wepster field number, RMNH collection and ZMAN collection
Register	Differentiates between the 8 different field books, for the ITH as well as the CMGL collections .
RMNH-reg.nr.	Unique registration number, added during this project.
BE.-nr.	Registration number of the current storage unit, added during this project.
Present_in_coll.	Presence of the specimen. 'Yes' if present. No/Missing if specimen was not found.
General Remarks	1) Verbatim information from the field books not suited for registration in other fields. 2) Comments from collaborators handling the specimen, such as 'Genitals present on pin'.

Collector	Collector of the specimen, verbatim.
Datum tekstueel	Collection Date, verbatim.
Coll. Date Start	Collection Date, template dd/mm/yyyy.
Coll. Date End	Collection end date, if collection date covers a range of dates, template dd/mm/yyyy.
Historical Owner	Indicates if the specimen was part of a significant historical collection (e.g. J. Bonne-Wepster).
Sex	Sexe of the specimen, if known: male/female.
Phase or Stage	Phase or stage of the specimen, primarily adult/larva.
Country	Geographic information on the gathering site of the specimen. If necessary, historic 'Country', 'State/Province', 'Island' and other location names are converted into current names.
StateProvince	
Island	
Locality	
Full Locality Tekst	All geographic information on the gathering site of the specimen, verbatim.
Altitude	Altitude in meters.
Order	Taxonomic identification of the specimen.
Family	
Genus	
Species	
Variety rank	
Variety	
Author	

Certainty	Certainty of the identification of the specimen. 'Uncertain' if the identification was doubtful.
Name comments	Additional comments regarding the identification of the specimen, such as specific information on which the specimen was identified. Also, discrepancies between identification in the field books and on the pin were registered in this field.
TYPEstatus	If applicable, type-information of the specimen, verbatim.
Determinator	Name of the identifier of the specimen, verbatim.
Det. Date	Date of the determination template dd/mm/yyyy.
Latitude	Georeference information: Coördinates
Longitude	
Uncertainty	Margin of error of coordinates (in meters)
Method	Point radius method (see Coordinates)
Remarks_coordinates	Additional comments regarding georeference information. E.g. remarks when a locality was not found, or the margin of error proved too high (>100km).

Table 2. Fields of the Google datasheet and their description

Coordinates: For the indication of the coordinates we have used Google Earth and the Georeferencing calculator [19]. The used technique was the 'point radius method'. The point radius method delineates a locality as a pair of points and a distance, with the distance being a radius that describes a circle around the points [18].

The field data of the Bonne-Wepster collection did not contain any coordinates. Georeferencing an 'old' collection has some issues. The locality descriptions are not always very specific. When the locality name was not specific enough (for instance The Lawa River, Suriname) or when unknown locality names were used we did not assign any coordinates. The maximum length of the radius was set at 100 km.

165 Because the gathering area of this collection is so widespread, from Sydney, Australia to
166 Whitehorse, Canada and from Transvaal, South Africa to Pampanga, Philippines an indication
167 of the minimum and maximum latitude and longitude is considered pointless.

168

169 **Data validation:** The field books were handwritten, so the interpretation of the locality
170 names and species names was not always easy.

171 Some of the mosquitoes with a Bonne-Wepster number were already provided with a locality
172 and identification label. We always checked if the locality on the label coincided with that in
173 the field book. When we found a mismatch between the two we reported that in the section
174 General Remarks of the Google spreadsheet. The same was done when a mismatch
175 concerned a species name, however this mismatch was annotated in the Name comment
176 section of the Google Spreadsheet.

177 To preserve the historical character of the collection as much as possible we decided not to
178 synonymize the species names. Therefore the dataset is provided with the names as
179 originally given by Bonne-Wepster. However, spelling mistakes were corrected using the
180 taxonomic checklist Culiclopedia [20]. The same checklist was used to interpret genus and
181 species names which were abbreviated in the field books. This was done to ensure that the
182 names imported in the NBC Database and ultimately exported to the GBIF platform were
183 free of errors.

184 We never removed any of the labels under a specimen, and it was difficult to establish how
185 discrepancies came about. It could be that a mosquito expert later redetermined some
186 specimens or that there were some mistakes made during the previous labeling process, but
187 that is merely speculation.

188 Sometimes the same number could refer to two specimens e.g. no.234, one from the former
189 CMGL collection or one from the former ITH collection. This could be a problem in
190 establishing the collecting event. By working intensively with the specimens we noticed a

difference in the way the two subcollections were labeled. The ITH collection had numbers written vertically (fig.6) while the CMGL collection had numbers written horizontally (fig.6).

Geographic coverage

Description

Indonesia, India, Iran, Sri Lanka, Yemen, Philippines, China, Thailand
Suriname, Brazil, Panama, Guatemala, Peru, Costa Rica, Cuba, Jamaica, Nicaragua,
Bahamas, Trinidad & Tobago, Dominican Republic, El Salvador, Ecuador, Guadalupe, Virgin
Islands, France, Belize, Puerto Rico, Colombia, Canada, Mexico, U.S.A., Netherlands,
Belgium, Russia, Croatia, Spain, North Macedonia, Bulgaria, Romania
Israel, Palestine, Italy, France, Portugal, Australia, Papua New Guinea, Congo, South Africa,
The Gambia, Republic of Guinea, Mozambique

The vast majority of the specimens were collected in Indonesia. Surinam ranks second in numbers collected.

215 Taxonomic coverage

216 Taxa included:

217

Rank	Scientific Name	Common Name
Kingdom	Animalia	Animals
Phylum	Arthropoda	Arthropods
Class	Insecta	Insects
Order	Diptera	Two-winged insects
Family	Culicidae	Mosquitoes
genus	Aedes, Aedimorphus, Anopheles, Armigeres, Banksinella, Bironella, Cancraedes, Chaoborus, Chironomus, Coquilletia, Coquilletidia, Corethra, Corethrella, Culex, Culiciomyia, Culiseta, Deinocerites, Dendromyia, Desvoidya, Ficalbia, Finlaya, Goeldia, Haemagogus, Harpagomyia, Heizmannia, Hodgesia, Howardina, Joblotia, Larrainea,	

	Leicesteria, Leucomyia, Lophoceratomyia, Limatus, Lutzia, Machlonyx, Malaya, Mansonia, Mansoniodes, Megarhinus, Miomyia, Mimomyia, Mochthogenes, Mucidus, Myzomyia, Ochlerotatus, Orthopodomyia, Pardomyia, Pseudoskusea, Psorophora, Rachionotomyia, Rachiosura, Runchomyia, Sabethes, Sabethoides, Simulium, Skusea Stegomyia, Taeniorhynchus, Theobaldia, Tipulida, Topomyia, Toxorhynchites, Trichoprosopon, Tripteroides, Uranotaenia, Verrallina, Wyeomyia	
--	---	--

218

219

220

221

222 Re-Use Potential

223 The data made available through this project is valuable because it describes historical and
224 recent records of mosquitoes and can be used in different areas of research including
225 estimates of spatial distribution, modeling current and future distributions through the tool of
226 Ecological Niche Modeling, systematics and for vector control programs.

227 The species names attached to the records are the original one and were not synonymized.

228

229 DATA AVAILABILITY

230 The dataset is available in the GBIF repository [21].

231

232 ETHICAL APPROVAL

233 Not Applicable

234

235 CONSENT FOR PUBLICATION

236 Not Applicable

237

238 COMPETING INTERESTS

239 The author(s) declare they have no competing interests

240 Funding

241 This work was co-funded by NLBIF, nlbif2020.007

242 Author contributions

243 **Pasquale Ciliberti**, assigned coordinates, digitized specimens, took pictures of the types,
244 quality control of the spreadsheet, wrote the paper, applied for funding
245 **Astrid Roquas**, entered the field books data in the spreadsheet, digitized specimens, wrote
246 the paper
247 **Becky Desjardins**, wrote the paper
248 **Bibiche Berkholst**, digitized specimens, took pictures of the types, quality control of the
249 spreadsheet, wrote the paper
250 **Frank Loggen**, assigned coordinates, digitized specimens, curated and prepared the
251 spreadsheet, quality control of the spreadsheet, wrote the paper
252 **Menno Hooft**, scanned the field books
253 **Gideon Gijswijt**, curated and imported the spreadsheet into the NBC database
254 **Dick de Graaff**, assigned coordinates, digitized specimens, quality control of the
255 spreadsheet, wrote the paper, applied for funding

256 Acknowledgements

257 We are very thankful to NLBIF for funding the project. Many thanks to Caroline Pepermans
258 for helping coordinate the team at the beginning of the project and to Jeroen Creuwels for
259 his suggestions in how to assign coordinates and his precious help in generating the DOI.
260 Thanks to Maarten Koster for his valuable help in entering the data in the Google
261 spreadsheet on a voluntary basis at the very beginning of the project. We are grateful to
262 Hendrik van Ede, Mark Doeland and Tamara de Reus for their help in assigning the
263 coordinates.
264

References

- 1 **Bradley RD, Bradley LC, Garner HJ & Baker RJ.** Assessing the value of Natural History Collections and addressing issues regarding long-term growth and care. *BioScience*, 2014; 64: 1150-1158
- 2 **Hilton EJ, Watkins-Colwell GJ & Huber SK.** The expanding role of Natural History Collections. *Ichthyology & Herpetology*, 2021; 109 (2): 379-391
- 3 **Ewers-Saucedo C, Allspach A, Barilaro C, Bick A, Brandt A, Fiege D, Fütting S, Hausdorf B, Hayer S, Husemann M, Joger U, Kamcke C, Küster M, Lohrmann V, Martin I, Michalik P, Reinicke G, Schwentner M, Stiller M & Brandis D.** Natural History collections recapitulate 200 years of faunal change. *Royal Society Open Science*, 2021; 8:1-19
- 4 **Yeates DK, Zwick A & Mikheyev AS.** Museum are biobanks: unlocking the genetic potential of the three billion specimens in the world's biological collections. *Current opinion in Insect Science*, 2016; 18: 83-88
- 5 **Lister AM & Climate Change Research Group.** Natural history collections as source of long-term datasets. *Trends in Ecology and Evolution*, 2011; 26 (4): 153-154
- 6 **Gaubert P, Papes M, Peterson AT.** Natural history collections and the conservation of poorly known taxa: Ecological niche modeling in central African rainforest genetids (*Genetta* ssp.). *Biological Conservation*, 2006; 1 (130): 106-117

- 291 7 **DiEuliis D, Johnson KR, Morse SS & Schindel DE.** Specimens collections should
292 have a much bigger role in infectious disease research and response. *PNAS*, 2016;
293 113 (1): 4-7
294
- 295 8 **Monquil-Broersen T & Gassó Miracle E.,** Van onschatbare waarde 200 jaar
296 *Naturalis. Naturalis Biodiversity Center & Walburg Pers, Zutphen*, 2020; pp. 1-352
297
- 298 9 **Becker N, Petric D, Zgomba M, Boase C, Madon M, Dahl C, Kaiser A.** Mosquito
299 and their control. *Springer, Berlin Heidelberg*, 2010; 1-577
300
- 301 10 **World Malaria Report 2022.** Geneva: World Health Organization, 2022; 1-372
302
- 303 11 **Lane MA.** Roles of Natural history collections. *Annals of the Missouri Botanical*
304 *Garden*, 1996; 83: 536-545
305
- 306 12 **van den Assem J & Bonne-Wepster J.** New Guinea Culicidae: a synopsis of
307 vectors, pests and common species. *Zoologische Bijdragen*, 1964; 6: 1-136
308
- 309 13 **Bonne C & Bonne-Wepster J.** Mosquitoes of Surinam: a study on Neotropical
310 mosquitoes. *Mededeeling Afdeeling Tropische Hygiëne*, 1925; 13: 1-558
311
- 312 14 **Bonne-Wepster J & Brug SL .** The subgenus *Stegomyia* in Netherlands India.
313 *Geneeskundig Tijdschrift voor Nederlandsch Indië*, 1932; 72: 39-119.
- 314 15 **Bonne-Wepster J & Brug SL.** Nederlandsch-Indische Culicinen. *Geneeskundig*
315 *Tijdschrift voor Nederlandsch Indië*, 1937; 77: 1-105
- 316 16 **Bonne-Wepster J & Swellengrebel NH.** The anopheline mosquitoes of the Indo-
317 Australian Region, 1953; J.H. de Bussy, Amsterdam: 1-504.

318 17 **Bonne-Wepster J.** Synopsis of a hundred common non-anopheline mosquitoes of
319 the greater and lesser Sundas, the Moluccas and New Guinea. *Special publication,*
320 *Department of Tropical Hygiene & Geographical Pathology*,1954; pp.1-147.

321 18 **Wieczorek J, Guo Q & Hijmans RJ.** The point-radius method for georeferencing
322 locality descriptions and calculating associated uncertainty. *International Journal of*
323 *Geographical Information Science*, 2004; 18 (8): 745-767
324

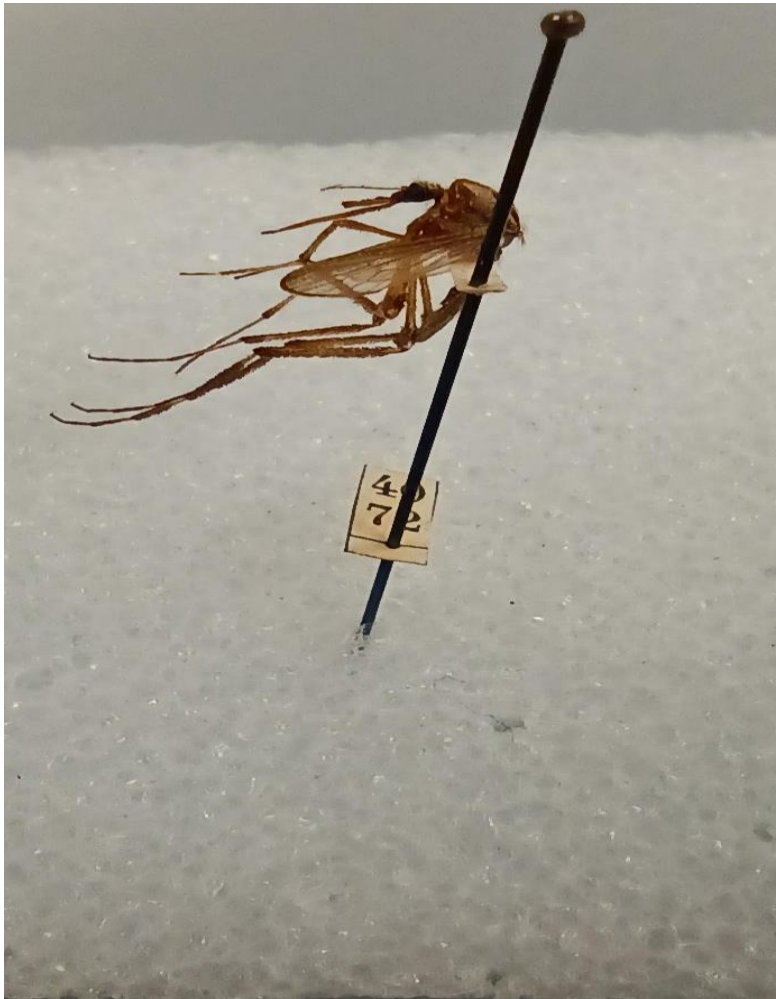
325 19 **Wieczorek C, Wieczorek J.** Georeferencing Calculator, 2021
326 <http://georeferencing.org/georefc/calculator/gc.html>. Accessed [23-2-2023].
327

328 20 **Harbach RE.** Culiclopedia: Species-group, genus group and family group names in
329 Culicidae (Diptera). *CABI Publishing*, 2018; pp 1-396
330

331 21 **Global Biodiversity Information Facility.** <https://doi.org/10.15468/dl.crvd8z>
332
333
334
335
336
337
338
339
340
341
342
343
344

345 **FIGURES:**

346



347

348 Fig.1. Example of a specimen labeled only with a number.

349

350

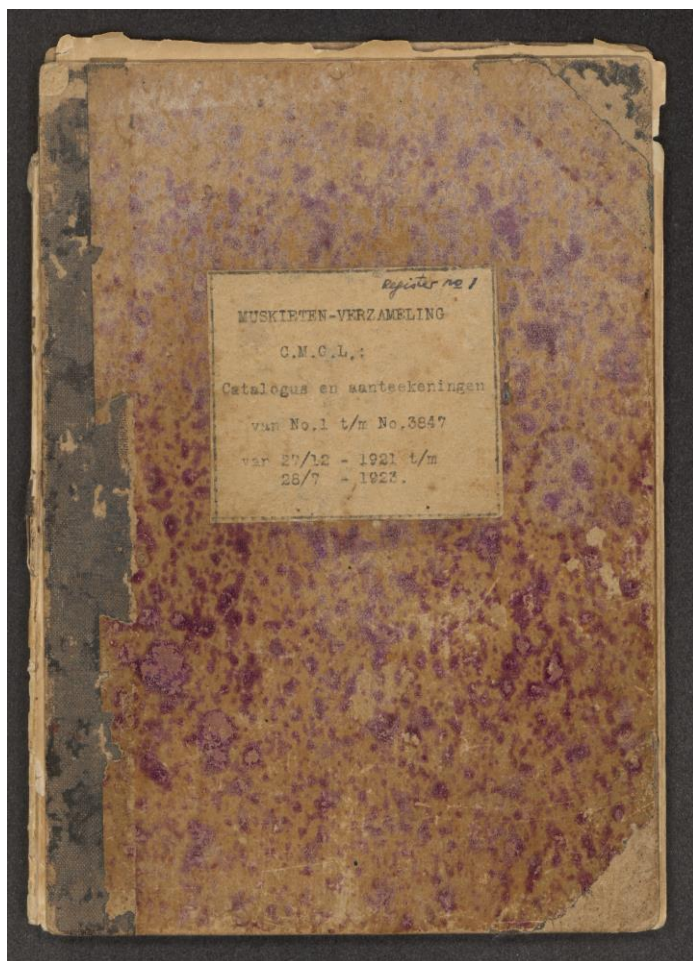


Fig.2. Field book.

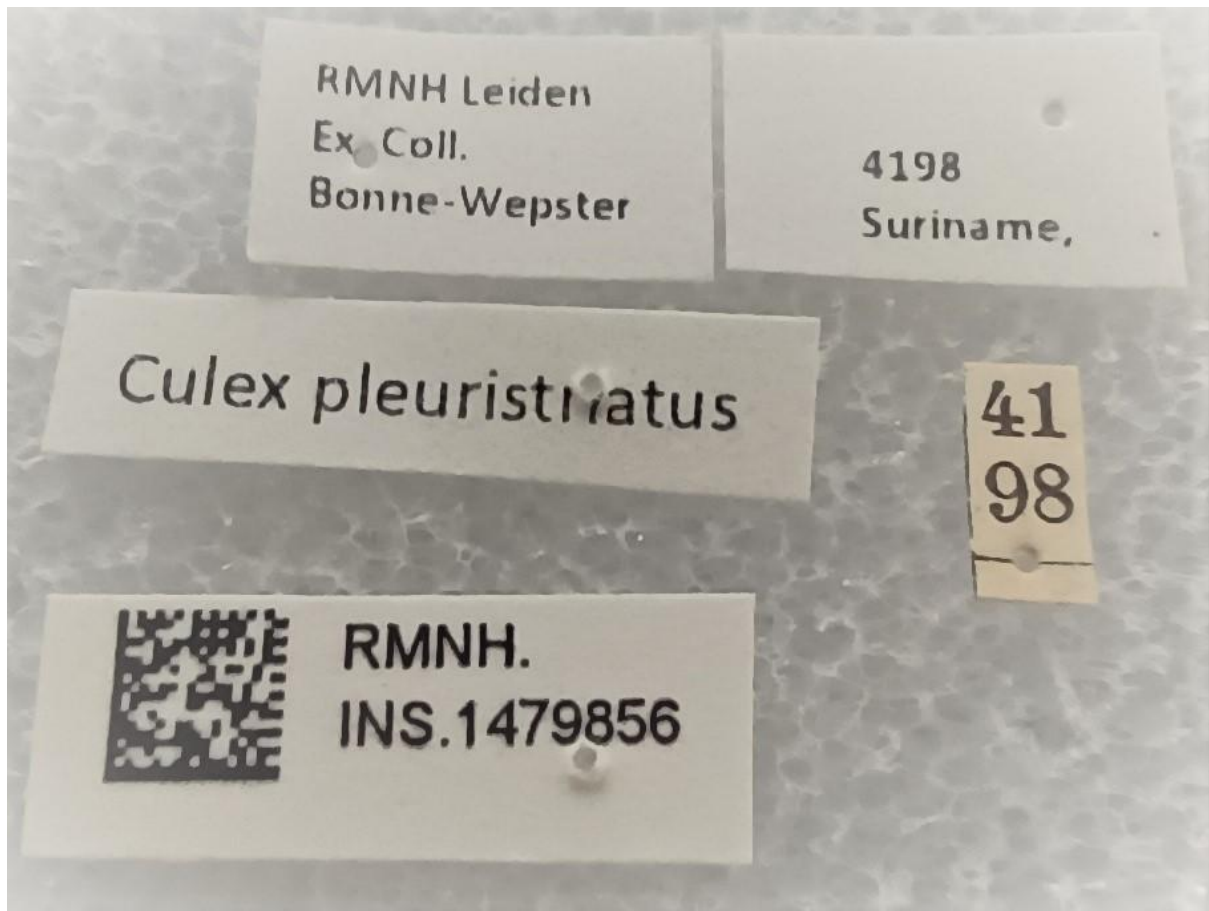


Fig.4. Examples of labels added during this project when only a number was pinned under the specimen. Note the unique registration number with a QR-code added to all specimens.

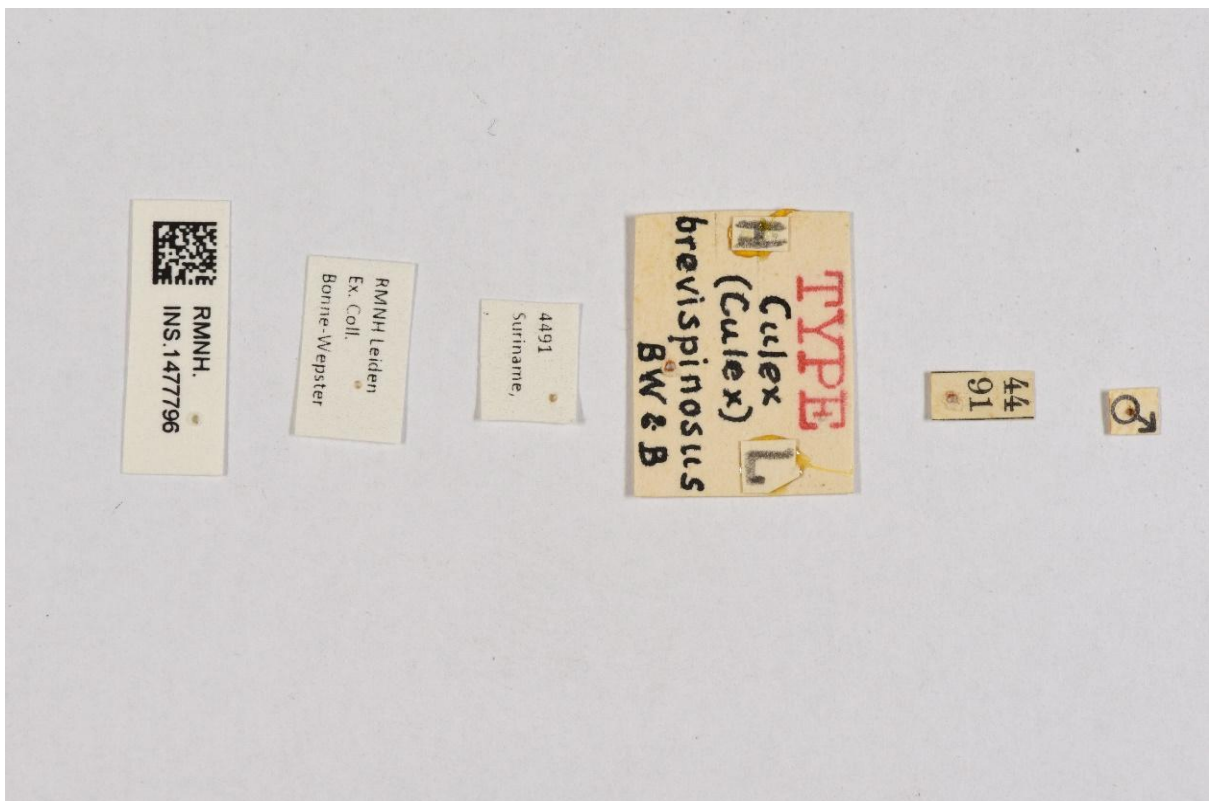
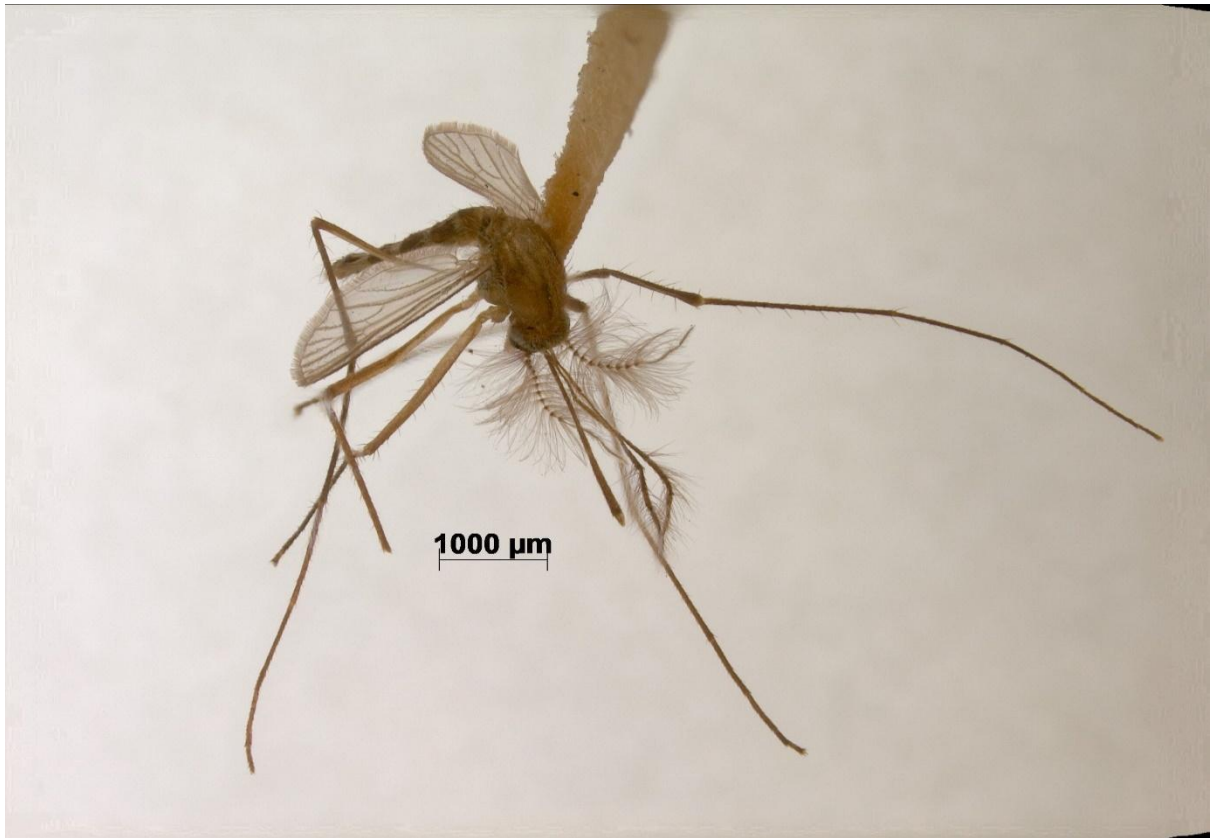
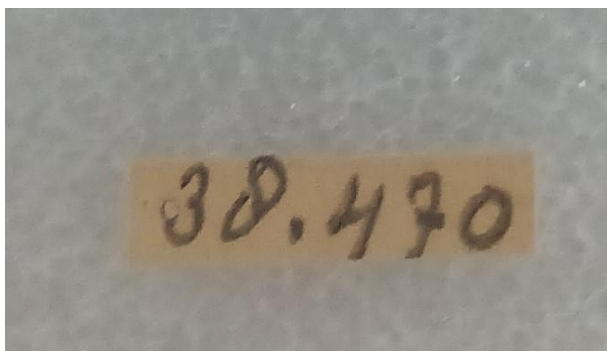


Fig.5. Above. Example of a type specimen (*Culex brevispinosus* Bonne-Wepster & Bonne 1920) Under. Labels of the type specimen *Culex brevispinosus*



371

372



373

374 Fig.6.Above. Example of typical ITH label number

375 Under: Example of typical CMGL label number

376

377