

## 1 Title

2 A geopositioned, evidence-graded, pan-species compendium of Mayaro Virus occurrence

## 3 Authors

4 Michael Celone<sup>1\*</sup>, Alex Potter<sup>2,3</sup>, Barbara A. Han<sup>4</sup>, Sean P. Beeman<sup>1</sup>, Bernard Okech<sup>1</sup>, Brett Forshey<sup>5</sup>,  
5 James Dunford<sup>1</sup>, George Rutherford<sup>6</sup>, Neida Karen Mita-Mendoza<sup>7</sup>, Elizabet Lilia Estallo<sup>8</sup>, Ricardo  
6 Khouri<sup>9</sup>, Isadora Cristina de Siqueira<sup>9</sup>, Kyle Petersen<sup>1</sup>, Ryan C. Maves<sup>10,11</sup>, Assaf Anyamba<sup>12</sup>, Simon  
7 Pollett<sup>11,13</sup>

- 8
- 9 1. Uniformed Services University of the Health Sciences, F. Edward Hébert School of Medicine,  
10 Department of Preventive Medicine & Biostatistics, Bethesda, Maryland 20814, USA
  - 11 2. Department of Entomology, Walter Reed Army Institute of Research, Silver Spring, Maryland 20910,  
12 USA
  - 13 3. Walter Reed Biosystematics Unit, Suitland, Maryland 20746, USA
  - 14 4. Cary Institute of Ecosystem Studies, Millbrook, New York 12545, USA
  - 15 5. Armed Forces Health Surveillance Division, Silver Spring, Maryland 20904, USA
  - 16 6. Institute for Global Health Sciences, University of California San Francisco, San Francisco, California  
17 94158, USA
  - 18 7. New York State Department of Health, NY, USA
  - 19 8. Instituto de Investigaciones Biológicas y Tecnológicas, Consejo Nacional de Investigaciones Científicas  
20 y Técnicas (CONICET)-Universidad Nacional de Córdoba, Córdoba, Argentina
  - 21 9. Instituto Gonçalo Moniz-Fiocruz, R. Waldemar Falcão, Salvador-BA, Brazil
  - 22 10. Section of Infectious Diseases, Wake Forest University School of Medicine, Winston-Salem, NC, USA
  - 23 11. Infectious Disease Clinical Research Program, Department of Preventive Medicine and Biostatistics,  
24 Uniformed Services University of the Health Sciences, Bethesda, MD, USA
  - 25 12. Geospatial Science and Human Security Division, Oak Ridge National Laboratory, One Bethel Valley  
26 Road, Oak Ridge, Tennessee 37830, USA
  - 27 13. Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, MD, USA

28  
29 Corresponding author: Michael Celone ([Michael.celone@usuhs.edu](mailto:Michael.celone@usuhs.edu))

## 30 Abstract

31 Mayaro Virus (MAYV) is an emerging health threat in the Americas that can cause febrile illness as well  
32 as debilitating arthralgia or arthritis. To better understand the geographic distribution of MAYV risk, we  
33 developed a georeferenced database of MAYV occurrence based on peer-reviewed literature and  
34 unpublished reports. Here we present this compendium, which includes both point and polygon  
35 locations linked to occurrence data documented from its discovery in 1954 until 2022. We describe all  
36 methods used to develop the database including data collection, georeferencing, management and  
37 quality-control. We also describe a customized grading system used to assess the quality of each study  
38 included in our review. The result is a comprehensive, evidence-graded database of confirmed MAYV  
39 occurrence in humans, non-human animals, and arthropods to-date, containing 266 geo-positioned  
40 occurrences in total. This database - which can be updated over time - may be useful for local spill-over  
41 risk assessment, epidemiological modelling to understand key transmission dynamics and drivers of  
42 MAYV spread, as well as identification of major surveillance gaps.

## 43 **Background & Summary**

44 First detected in Trinidad in 1954<sup>1</sup>, Mayaro virus (MAYV) is a mosquito-borne *Alphavirus* that causes  
45 periodic outbreaks of febrile illness in several Latin American countries<sup>2-5</sup>. Serological surveys conducted  
46 in the Americas suggest widespread circulation of the virus throughout the region<sup>6-8</sup>. Although MAYV  
47 may cause debilitating arthralgia or arthritis, it often presents with non-specific symptoms that may be  
48 clinically indistinguishable from other vector-borne diseases such as dengue or Zika<sup>9</sup>. No licensed  
49 vaccine or antiviral treatment currently exists, and the current standard of care for MAYV infections is  
50 limited to supportive care<sup>9,10</sup>.

51  
52 Collating and evaluating the current evidence regarding the distribution of MAYV occurrence is a critical  
53 step in characterizing its transmission potential and identifying the communities at greatest risk for  
54 MAYV outbreaks. Such compendium databases have proven valuable for the study and prevention of  
55 other emerging and re-emerging pathogens and diseases, such as Middle East respiratory syndrome  
56 coronavirus (MERS-CoV), leishmaniasis, Crimean-Congo hemorrhagic fever, and dengue viruses<sup>11-14</sup>.  
57 Previously published reviews have explored MAYV occurrence in the Americas<sup>15,16</sup>. This review seeks to  
58 fill gaps in the current literature by providing the highest possible level of geographic resolution for the  
59 available MAYV occurrence data across humans, non-human animals, and arthropods, coupled with an  
60 evidence-based grading for each of those occurrences. In addition to drawing from occurrences  
61 identified in a human systematic review, this compendium database includes occurrences identified in  
62 another systematic review<sup>17</sup> evaluating the distribution of MAYV occurrence among non-human animal  
63 and arthropod species in the Americas. Collectively, these two systematic reviews provide a  
64 comprehensive compendium of MAYV occurrence to permit ecological and epidemiological risk  
65 prediction and forecasting. Further, these two systematic reviews put forward joint citable frameworks  
66 for the field to evaluate the quality of studies that propose non-human animal, arthropod, and human  
67 MAYV occurrence.

68  
69 This georeferenced, evidence-graded MAYV database contains 266 unique localities across 15 countries  
70 published between 1954 and 2022. The methods described below are adapted from previously  
71 published disease occurrence compendiums<sup>11-14</sup>.

## 72 **Methods**

### 73 **Data Collection**

74 Data collection and abstraction for non-human animal and arthropod data was described in a previously  
75 published study<sup>17</sup>. We followed a similar data collection strategy for human MAYV occurrence data. We  
76 first searched Embase, Web of Science, PubMed/MEDLINE, and SciELO databases for English, Spanish,  
77 and Portuguese language articles published between January 1954 and January 2021 using the search  
78 term “Mayaro”. Furthermore, a PubMed/MEDLINE alert using the search term “Mayaro” captured five  
79 additional eligible studies reporting human MAYV occurrence that were published between the initial  
80 search and May 2022.

81  
82 This database search was extended using the bioRxiv (<https://www.biorxiv.org/>) and medRxiv  
83 (<https://www.medrxiv.org/>) pre-print databases. We complemented these database search results with  
84 ‘grey literature’, including hand-searched bibliographies of the included articles and MAYV review  
85 articles (including systematic reviews), the Pan American Health Organization (PAHO) Institutional  
86 Repository for Information Sharing database ([iris.paho.org](https://iris.paho.org/)), the GIDEON database  
87 (<https://www.gideononline.com/>), and GenBank [29] (<https://www.ncbi.nlm.nih.gov/genbank/>). We

88 also searched for dissertations from several Brazilian university repositories including Instituto Evandro  
89 Chagas (<https://patua.iec.gov.br/>), Universidade de São Paulo (<https://teses.usp.br/>), Universidade  
90 Federal de Goiás (<https://repositorio.bc.ufg.br/>), Pontifícia Universidade Católica do Rio Grande do Sul  
91 (<http://tede2.pucrs.br/>), Universidade Federal de Mato Grosso (<https://bdm.ufmt.br/>), Fundação  
92 Oswaldo Cruz (<https://portal.fiocruz.br/>), and Universidade Federal do Pará  
93 (<http://repositorio.ufpa.br/>). In addition, we searched conference handbooks that are available online  
94 (2004-2019) from the American Society of Tropical Medicine and Hygiene  
95 (<https://www.astmh.org/annual-meeting/past-meetings>). Articles were considered for eligibility if they  
96 reported original research studies on MAYV occurrence in humans including serological surveys,  
97 outbreak investigations, case reports, or surveillance studies. After the initial literature search, we  
98 conducted a secondary search to identify any relevant articles describing the occurrence of Uruma virus.  
99 Initially described as a novel human pathogen in 1959<sup>18</sup>, Uruma virus is now considered a strain of  
100 MAYV<sup>19</sup>. Therefore, it was decided that Uruma virus records would be included in our systematic  
101 review.

102  
103 Two reviewers independently screened all titles and abstracts to determine articles that could  
104 immediately be discarded based on relevancy and articles that should be included in the second stage of  
105 review. Results were compared to reconcile any differences between the reviewers. Each reviewer then  
106 independently read the full text of potentially eligible articles identified through screening and identified  
107 articles that were candidates for inclusion in the study. Results were compared to reconcile any  
108 differences between the two reviewers. A third-party reviewer adjudicated if consensus was not  
109 reached between the reviewers. From those studies deemed eligible, data was extracted from articles  
110 by one reviewer using a predetermined data abstraction tool in Microsoft Excel. Five percent of entries  
111 were randomly selected and reviewed by a second reviewer.

112  
113 Overall, 144 research items (including journal articles, dissertations, news articles, GenBank entries, etc.)  
114 were deemed eligible. All eligible articles are included in the Supplementary materials (Data Citation 1).  
115 No additional articles were deemed eligible following the secondary search for Uruma virus.

### 116 117 **Grading quality of evidence**

118 For human occurrences, we developed a customized grading system to assess the quality of each study  
119 included in our review. This followed a similar framework we developed for evaluating the quality of  
120 each study included in the published systematic review on MAYV occurrence in non-human animals and  
121 arthropods<sup>17</sup>. We assigned each study in our systematic review a grade for each of four quality items:  
122 clarity of research question/objective; description of study methods; description of sampling methods;  
123 and validity of diagnostic tests. For each quality item, eligible studies were assigned a score of 3 (strong  
124 evidence), 2 (moderate evidence), 1 (weak evidence), or unable to judge. Studies were deemed unable  
125 to judge if the information provided was insufficient to assign quality scores (e.g., a single GenBank  
126 entry, conference abstract, or report). Table 1 refers to a description of the four quality grading  
127 domains.

128  
129 Two reviewers independently graded the evidence quality for each study and results were compared to  
130 reconcile any differences between the two reviewers. A third-party reviewer adjudicated if the two  
131 initial reviewers did not reach consensus.

132

### 133 **Geo-positioning of the MAYV Occurrence Data**

134 All available location information from each confirmed MAYV case was extracted from each article and  
135 added to the database based on methods that have been described previously<sup>11-14</sup>. Each occurrence of  
136 MAYV was designated as either a point or polygon location according to the spatial resolution provided  
137 in the article. When specific latitude and longitude coordinates were provided, they were verified in  
138 Google Maps (<https://www.google.com/maps>) and added to the database as a point location and  
139 coordinates were converted to decimal degrees. If a neighborhood, town, village, or small city was  
140 explicitly mentioned in the article and the uncertainty of the location was less than 5 kilometers (km), it  
141 was entered into the database as a point location and its centroid coordinates were recorded. An online  
142 gazetteer ([www.geonames.com](http://www.geonames.com)) as well as Google Maps or ArcGIS (ESRI 2011. ArcGIS Pro: Release 2.6.0.  
143 Redlands, CA: Environmental Systems Research Institute), along with contextual information, were used  
144 to verify site locations.

145  
146 For studies that reported a less precise spatial resolution such as states or counties, first level (ADM1),  
147 second level (ADM2) or third level (ADM3) administrative divisions were recorded as polygons and  
148 coded according to the global administrative unit layers (GAUL) from the Food and Agriculture  
149 Organization<sup>20</sup>. If the uncertainty of a specific named location was greater than 5 km (e.g., the city of  
150 Manaus, Brazil), the occurrence was assigned to a custom polygon created in ArcGIS that encompassed  
151 the extent of the city. In the rare case that no specific intra-country location was provided, the record  
152 was assigned to its country of occurrence (ADM0). When place names were duplicated (i.e., the ADM1  
153 and ADM2 units had the same name), the coarsest spatial resolution was used. For example, if the  
154 MAYV case location was reported as “*Cusco, Peru*,” with no additional information provided, the record  
155 was assigned to the Cusco ADM1 polygon. However, if the study specified that the case occurred in the  
156 “*City of Cusco*”, the record was assigned to a custom polygon that encompassed the City of Cusco. The  
157 centroid coordinates of ADM1, ADM2, and ADM3 polygons, or custom polygons were retrieved from the  
158 GeoNames gazetteer whenever possible. If centroid coordinates were not available in GeoNames, they  
159 were estimated using Google Maps. The coordinates for each georeference and the methods and source  
160 used to obtain the coordinates were documented in the compendium.

161  
162 Uncertainty was measured in kilometers for each MAYV occurrence point or polygon<sup>21</sup>. For polygons,  
163 uncertainty was calculated as the distance from the polygon centroid coordinates to the polygon’s  
164 furthest boundary. For point locations with well-defined boundaries, the same procedure was followed,  
165 whereby the uncertainty encompassed the extent of the location’s area. When locations did not have  
166 well defined boundaries, uncertainty was calculated as half the distance to the nearest named place  
167<sup>21,22</sup>. Calculation of uncertainty was completed using measurements from Google Maps. When authors  
168 provided exact coordinates gathered using GPS, uncertainty was calculated using a georeferencing  
169 calculator<sup>23</sup>. Exact coordinates were only used if authors provided a high level of precision (e.g.,  
170 precision higher than “minutes” in degrees-minutes-seconds format and similarly high precision for  
171 coordinates in decimal degrees). When coordinates were provided at a low precision, we georeferenced  
172 the named place instead.

173  
174 Several articles reported the diagnosis of MAYV in human blood samples at urban hospitals. If no  
175 relevant information was provided on the study participants (e.g., place of residence), we georeferenced  
176 the ADM2 unit in which the hospital was located.

### 177 **Data and Metadata Records**

178 This database is available in the Supplementary materials (Data Citation 1). Each of the 280 rows  
179 represents a unique occurrence of MAYV in a human, non-human animal, or arthropod. Location IDs for

180 points and polygons were assigned to each unique location. The MAYV occurrence database contains  
181 the following fields, following best-practice nomenclature as previously documented in georeferenced  
182 compendiums of other pathogens<sup>11-14</sup>:

- 183 1) **Location\_ID**: A unique identifier was assigned to each georeference. The prefix used in the  
184 location ID denoted the georeference type: *ADM 0, 1, 2, or 3* for administrative units, *CP* for  
185 custom polygons, and *P* for point locations. Separate studies with duplicate georeferences were  
186 assigned the same Location ID, and duplicates were removed according to the methods  
187 described above.
- 188 2) **Author\_Year**: The first author and publication year for each record.
- 189 3) **Ref\_Number**: A reference identification number was documented when applicable. A PubMed  
190 ID number was recorded for all published studies. If this was not available a DOI, GenBank locus,  
191 URL, ProMED identifier, etc., was captured.
- 192 4) **Year\_MAYV\_Start**: The earliest year that MAYV infection was detected within the publication  
193 was recorded if available. If studies only included a range of years and did not specify the precise  
194 year that MAYV was found, this range was documented. Note that this variable refers to  
195 infection detection and doesn't infer the onset of infection (particularly in the case of  
196 serological-based occurrence studies).
- 197 5) **Year\_MAYV\_End**: The latest year that MAYV infection was detected within the publication was  
198 recorded if available. If studies only included a range of years and did not specify the precise  
199 year that MAYV was found, this range was documented. We followed the methods of Hill et  
200 al.,<sup>24</sup> when studies did not report any year (i.e., an assumption was made that the case was  
201 detected three years before publication) .
- 202 6) **Host\_Type**: One of three host types was documented for each occurrence: human, non-human  
203 animal, or arthropod. If multiple host types were detected with MAYV in the same location, a  
204 separate row was included for each host type.
- 205 7) **Location\_Description**: We documented relevant information related to the location of the  
206 occurrence record. This field included the decisions made during the georeferencing process to  
207 reach the final determination regarding the location of each record.
- 208 8) **Adm0**: The country where MAYV occurrence was detected.
- 209 9) **Adm1**: The first level administrative unit where MAYV occurrence was detected (if available).
- 210 10) **Adm2**: The second level administrative unit where MAYV occurrence was detected (if available).
- 211 11) **GAUL\_code**: When a MAYV occurrence was georeferenced as an ADM1 or ADM2 administrative  
212 polygon, the GAUL code was included.
- 213 12) **Finer\_Res**: If finer spatial resolution was documented (e.g., a town, city, or exact coordinates)  
214 this was recorded.
- 215 13) **Location\_Type**: Each occurrence was documented as either a point or polygon location type,  
216 depending on the spatial resolution that was provided. Custom polygons are available as  
217 shapefiles in the Supplementary materials (Data Citation 1). These can be opened in GIS  
218 software or using statistical packages that handle spatial data.
- 219 14) **Admin\_Level**: The administrative level for each polygon location was recorded as either 0  
220 (country level), 1 (first level administrative division), 2 (second level administrative division), or 3  
221 (third level administrative division) depending on the spatial resolution that is provided. If the  
222 occurrence was georeferenced as a point location or custom polygon, -999 was recorded.
- 223 15) **Y\_Coord**: The longitude coordinate was recorded in decimal degrees. The coordinates were  
224 taken verbatim from the article when available. Otherwise, the polygon centroid was recorded.
- 225 16) **X\_Coord**: The latitude coordinate was recorded in decimal degrees. The coordinates were taken  
226 verbatim from the article when available. Otherwise, the polygon centroid was recorded.

- 227 17) **Coord\_Source**: This field describes how the coordinates were determined. Possibilities include  
228 the following:
- 229 a. Exact coordinates provided in the article.
  - 230 b. Polygon centroid coordinates retrieved from GeoNames.
  - 231 c. Location was determined based on the details provided in the article (e.g., a specific  
232 neighborhood was mentioned), and centroid coordinates were subsequently  
233 determined using Google Maps.
- 234 18) **Uncertainty\_km**: The amount of uncertainty associated with the record, measured in km.
- 235 19) **Uncertainty\_Description**: The method used to calculate uncertainty for each georeference. For  
236 example, the uncertainty of polygons was measured as the distance from the polygon centroid  
237 to the furthest polygon boundary. For certain point and polygon locations where the boundary  
238 was not clear, uncertainty was measured as half the distance to the nearest named place.  
239 Uncertainty calculations were based on previously published methods <sup>21,22</sup>.
- 240 20) **Diagnostic\_Type**: The specific diagnostic test (e.g., polymerase chain reaction [PCR],  
241 neutralization test [NT]; hemagglutination inhibition [HI]; enzyme-linked immunosorbent assay  
242 [ELISA]) was documented.
- 243 21) **Positive\_n**: The number of positive MAYV cases that were reported.
- 244 22) **Denominator**: The total number of humans, non-human animals, or arthropod pools in the  
245 study.
- 246 23) **Language**: The language of the article, either English, Portuguese, or Spanish.

## 247 **Technical Validation**

248 All georeferencing was completed by one study author and validated by a second author. In the case of  
249 a disagreement or discrepancy between the two authors, a third author adjudicated. A location  
250 identification was assigned to each unique georeference in the dataset. All final georeferences were  
251 plotted in ArcGIS for visual inspection. All data points were checked to ensure they fell on land and  
252 within the correct country. In the case of duplicate georeferences, we retained the record with the  
253 highest quality score; if quality scores were identical, we retained the more recent record.

## 254 **Usage Notes**

255 We identified 144 eligible references for inclusion in our study (see flowchart in Fig 1). The resulting  
256 database 266 unique geo-positioned MAYV locations worldwide, including 96 unique points and 170  
257 unique polygons. Therefore, each row in the database represents a unique location where MAYV was  
258 detected in humans, non-human animals, or arthropods. Duplicate georeferences from the same host  
259 type were removed from the main database (see the Duplicates table in Supplementary material in Data  
260 Citation 1) following the approach specified in the methods section. Some duplicate georeferences were  
261 included if multiple host types (e.g., human and arthropod) were found with MAYV at the same location.  
262 For example, Hoch et al., 1981 <sup>25</sup> detected MAYV in both humans and arthropods in the ADM2 unit of  
263 Belterra, Brazil. Two separate rows (one row for humans and one for arthropods) are included in the  
264 compendium with the same georeference.; therefore, the database includes 280 rows, each  
265 representing a unique occurrence of MAYV in a human, non-human animal, or arthropod. Of these 280  
266 rows, 222 (79%) were in humans, 34 (12%) were in non-human animals, and 24 (9%) were in arthropods.  
267 MAYV was reported in 15 countries overall, with the majority occurring in Brazil (n = 135). According to  
268 our review, MAYV occurrences are limited to the region between latitude 35S and 12N of tropical South  
269 America. One article <sup>26</sup> reported MAYV occurrence in Zambia; this occurrence was not included in our  
270 georeferenced database due to the lack of evidence supporting MAYV circulation outside of the  
271 Americas and the potential cross-reactivity with antibodies of other alphaviruses in the Semliki Forest  
272 serocomplex <sup>27</sup>.

273  
274 Studies were included in our systematic review if they reported testing for MAYV occurrence, even if  
275 MAYV was not detected. These negative results are not included in the georeferenced compendium, but  
276 they can be found in the Supplementary materials (Data Citation 1). Many machine-learning models that  
277 are common in the ecological literature are presence-only or presence-background algorithms that rely  
278 on “pseudoabsence” data in lieu of true absences. For this reason, the true absence data presented here  
279 are potentially valuable for disease modelling. However, any reports of disease absence must be  
280 considered carefully as true absence is difficult to establish and false absence data can result in  
281 miscalibration of distribution models<sup>28</sup>. Ideally, representative country-wide surveys should be used to  
282 ascertain “true” absence locations that can be used in subsequent modelling efforts<sup>29,30</sup>.

283  
284 As with other published compendiums<sup>11</sup>, these curated data derived from published sources are  
285 expected to complement and augment other surveillance data used by public health agencies, thereby  
286 increasing our understanding of the distribution of MAYV in the Americas across multiple host types  
287 with a high spatial resolution. These data may also assist in identifying under-sampled regions, and may  
288 assist in identifying priority regions for surveillance. The georeferences can also serve as the basis for  
289 development of epidemiological models or risk maps that characterize the potential suitability for MAYV  
290 occurrence, including the risk of spillover into human populations and the potential influence of climate  
291 change on MAYV distribution. For example, the 2013 compendium of dengue virus (DENV) occurrence<sup>12</sup>  
292 was used as the basis for a highly cited modelling study that estimated the global distribution of DENV  
293 risk<sup>31</sup>. Finally, leveraging the methods and data presented here, this open access database can be  
294 updated as additional studies are published that report MAYV in the Americas.

295  
296 There are several important limitations that must be considered when using this dataset. One significant  
297 limitation is the impact of sampling bias on the detection and public reporting of MAYV occurrence.  
298 Heterogeneity of public health arboviral surveillance systems (including variability in surveillance  
299 infrastructure and competing public health demands) and MAYV research activity may skew MAYV  
300 detection and reporting by geographic region. Therefore, the absence of MAYV occurrence in some  
301 settings may not represent true disease absence, but rather ascertainment bias. This important  
302 limitation must be addressed in subsequent modeling studies in order to reduce the effects of sampling  
303 bias on model accuracy<sup>32</sup>. Some published studies have proposed an evidence consensus score which  
304 quantifies the evidence supporting the presence or absence of a pathogen in a given region<sup>33</sup>. This score  
305 can be calculated using multiple evidence categories (e.g., health organization reporting status or health  
306 expenditure) which may provide useful evidence of disease presence or absence in areas, including  
307 those with more limited arboviral surveillance.

308  
309 Another limitation of our study is the lack of geographic precision associated with MAYV occurrence  
310 records. Many articles did not provide sufficient geographic detail to georeference MAYV records with a  
311 high level of precision. We attempted to capture this uncertainty by assigning polygon locations to these  
312 records. When a greater level of geographic detail was provided by study authors, we were able to  
313 georeference some records as point locations (i.e., locations of MAYV occurrence with less than 5 km of  
314 uncertainty).

315  
316 Finally, an additional limitation is associated with the variable assay validity used to detect MAYV. Some  
317 studies reported MAYV presence based only on positive serological assays such as hemagglutination  
318 inhibition (HI) tests while other studies provided stronger evidence of MAYV occurrence based on  
319 reference neutralization assays or PCR testing. We estimated the strength of evidence of MAYV  
320 occurrence using a custom evidence grade which could be used in other studies. The strength of

321 evidence annotated in these data can be considered in future modeling efforts, with certain low-  
322 evidence records potentially excluded from models as part of sensitivity analyses. Moreover, the  
323 variability of evidence for MAYV occurrence demonstrated here prompts study design considerations for  
324 future MAYV research and public health surveillance.

325  
326 In addition to the main comma-delimited database, three Supplementary Files are included as part of  
327 the file set which can be found online (Data Citation 1). These files include: (i) a document containing  
328 the quality score and a citation for each of the references included in our review, (ii) a list of duplicate  
329 georeferences that were excluded from the database, and (iii) a shapefile of the custom polygons. The  
330 studies that described only negative MAYV results (i.e., those that are not included in the georeferenced  
331 compendium) are indicated by an asterisk in the S2 File.

332

### 333 **Acknowledgements**

334 The contents, views or opinions expressed in this publication or presentation are those of the authors  
335 and do not necessarily reflect official policy or position of Uniformed Services University of the Health  
336 Sciences, the Department of Defense (DoD), or Departments of the Army, Navy, or Air Force. Mention of  
337 trade names, commercial products, or organizations does not imply endorsement by the U.S.  
338 Government.

339 SP was supported by the National Institute of Allergy and Infectious Diseases, National Institutes of  
340 Health, <https://www.niaid.nih.gov/>, under Inter-Agency Agreement Y1-AI-5072, and the Defense Health  
341 Program, U.S. DoD, under award HU0001190002. AP was financially supported by the Armed Forces  
342 Health Surveillance Division – Global Emerging Infections Surveillance (AFHSD-GEIS) award  
343 P0065\_22\_WR. The activities undertaken at the Walter Reed Biosystematics Unit were performed in  
344 part under a Memorandum of Understanding between the Walter Reed Army Institute of Research  
345 (WRAIR) and the Smithsonian Institution, with institutional support provided by both organizations. The  
346 funders had no role in study design, data collection and analysis, decision to publish, or preparation of  
347 the manuscript.

348

### 349 **Author Contributions**

350 M.C., B.A.H, A.A., and S.P. drafted the data collation protocol. M.C., A.P., B.F., J.D., G.R., N.K.M.M., E.L.L,  
351 R.K., I.C.S, K.P, R.C.M, and S.P. assisted with article review, evidence grading, and data extraction. M.C.,  
352 A.P., N.K.M.M, E.L.L, R.K., and I.C.S contributed to georeferencing. C.L.M. checked all extracted data.  
353 M.C. and S.P. wrote the first draft, and all authors contributed to the manuscript.

### 354 **Data Citation**

355 To download data click the following link:

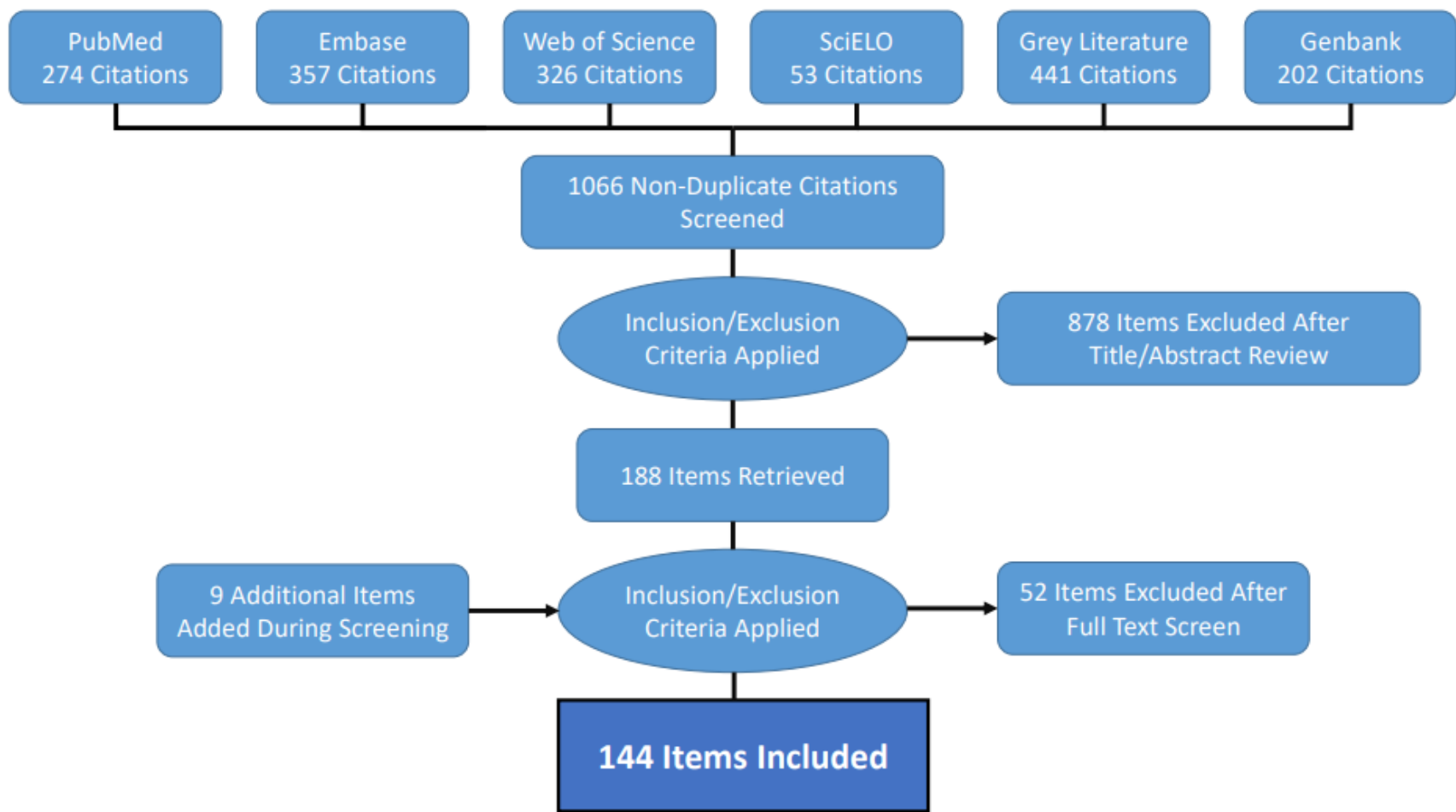
356 <https://datadryad.org/stash/share/2EY752eZN2Rfh8LXL6QHxfU6bxHPEjWQcchikr61ImM>

### 357 **Competing Interests**

358 The authors declare no competing financial interests.

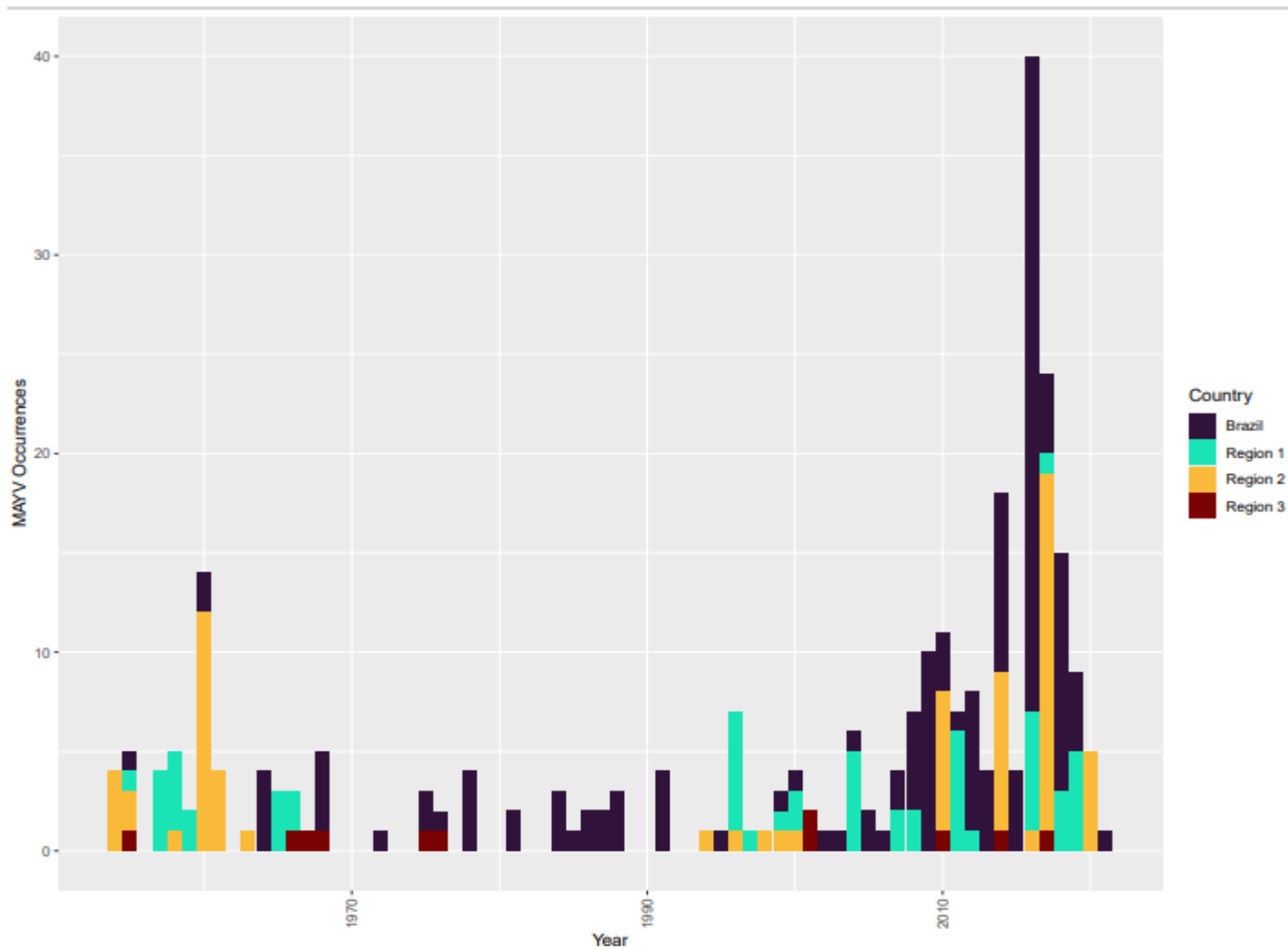
359



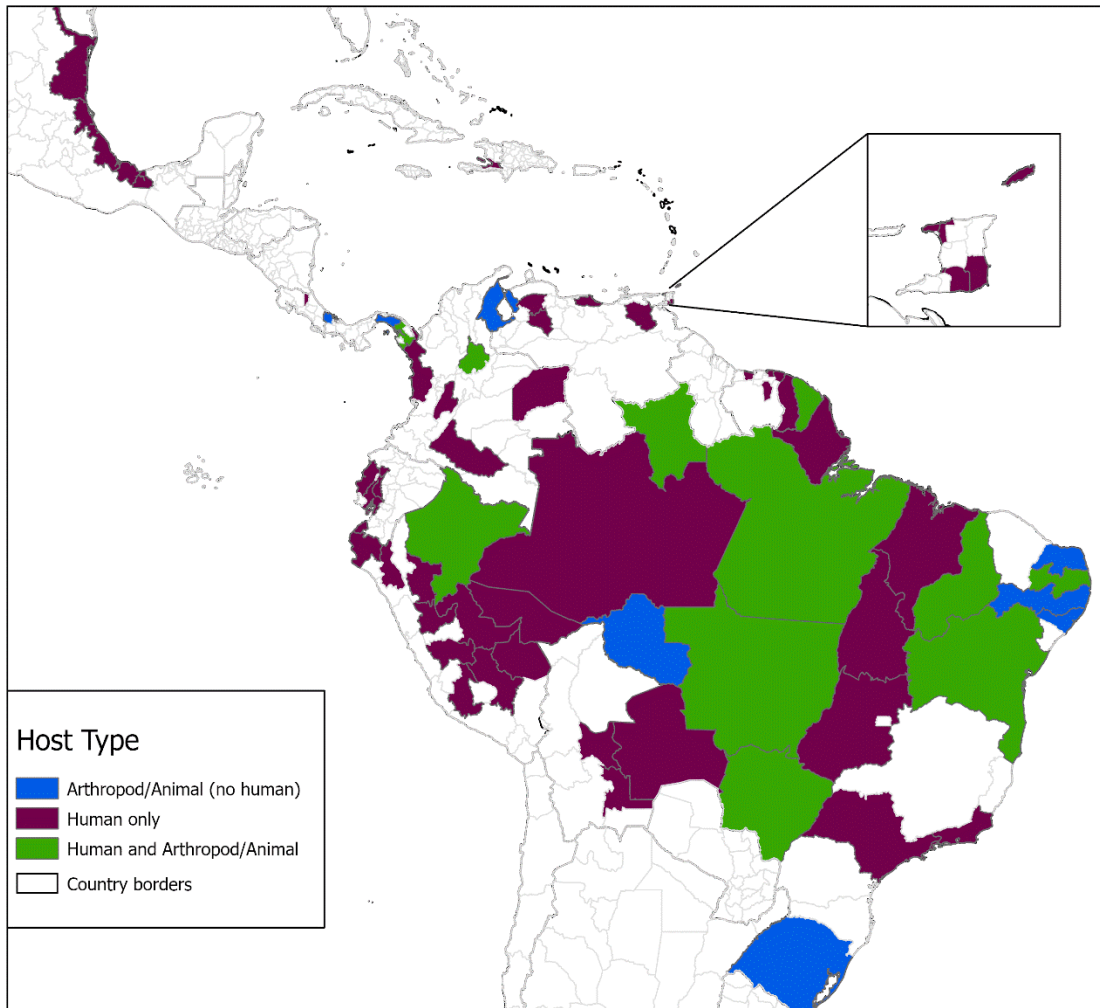


360  
361  
362  
363  
364

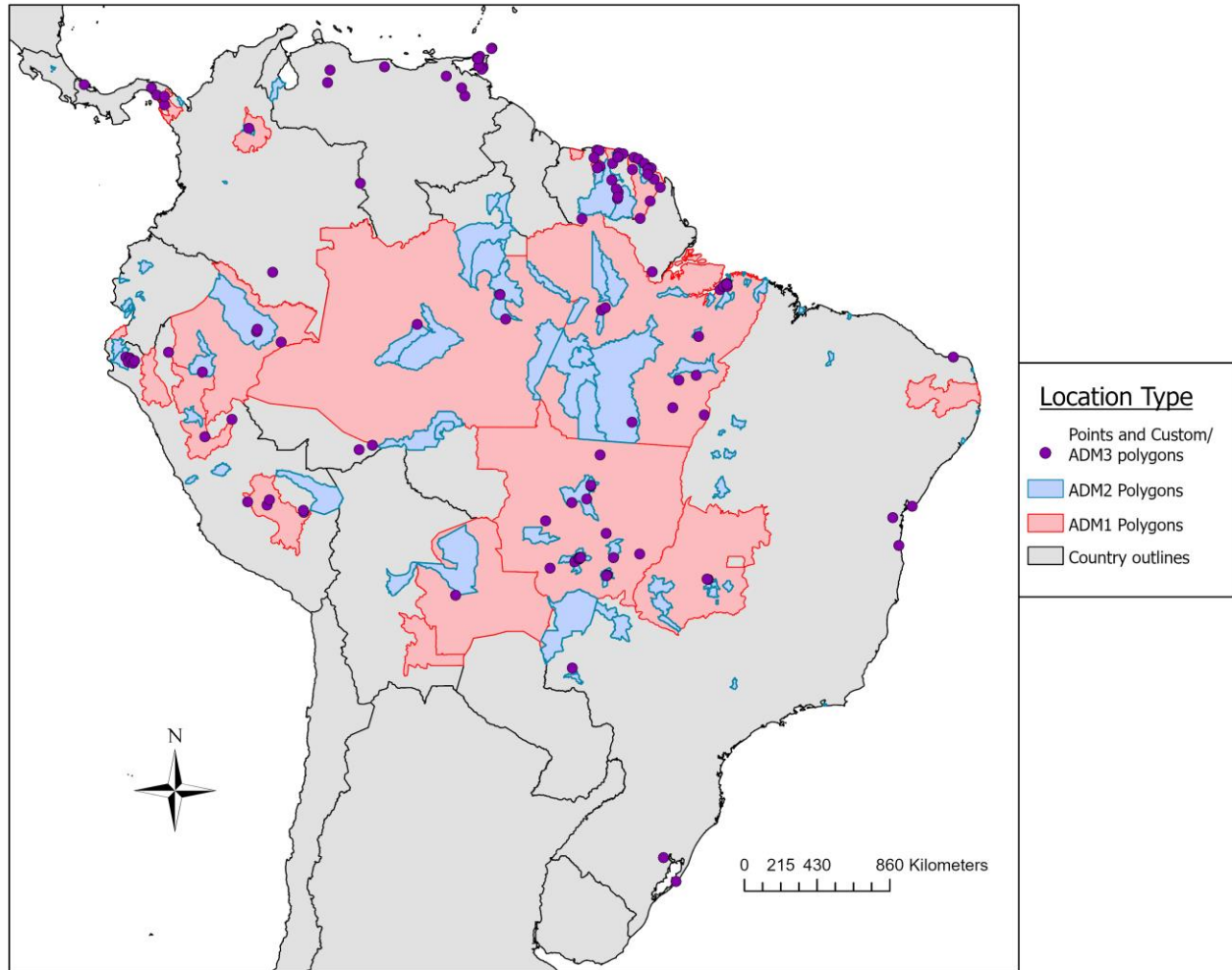
**Figure 1. MAYV literature extraction flowchart for human occurrence.** The flowchart for non-human animal and arthropod occurrence is provided in the previously published systematic review <sup>17</sup>.



365  
 366 **Figure 2. MAYV occurrence (all host types) by year and region.** All countries except Brazil were grouped according to geographic region. Region  
 367 1 includes Peru, Bolivia, Ecuador, and Colombia. Region 2 includes French Guiana, Guiana, Suriname, Venezuela, and Trinidad & Tobago. Region  
 368 3 includes Panama, Costa Rica, Mexico, Haiti and Antigua.



369  
370 **Figure 3. Distribution of MAYV occurrence by first level administrative division.** MAYV occurrences are  
371 aggregated to the ADM1 level and presented by host type. Host types include human only, reservoir  
372 only (non-human animal, arthropod, or both), or human and reservoir (human and non-human animal  
373 or arthropod, or all three host types). The inset map shows Trinidad and Tobago.  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384



385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399

**Figure 4. Distribution map of MAYV occurrence by location type.** All unique MAYV occurrences are presented according to the precision of the georeference. Red outlines represent first-level administrative units and blue outlines represent second-level administrative units. Both point locations and custom polygons are represented as purple points. Not visible on the map are two ADM2 polygons in Mexico and one ADM2 polygon in Haiti.

400 **Tables**

401 **Table 1: Evidence quality grading scheme – Human infection**

402

<b>Quality item</b>	<b>3-Strong</b>	<b>2- Moderate</b>	<b>1- Weak</b>	<b>Unable to judge</b>
<b>Was the research question/objective clearly described and stated?</b>	Described in full detail	Described in moderate detail (neither strong nor weak evidence)	Not described at all	Unable to judge based on the information presented
<b>Were the study design (including location and year of study) and participant recruitment presented in a reproducible way?</b>	Described in full detail	Described in moderate detail (neither strong nor weak evidence)	Not described at all	Unable to judge based on the information presented
<b>Was relevant information collected on participants, including travel history, occupation, location of residence, etc.?</b>	Described in full detail	Described in moderate detail (neither strong nor weak evidence)	Not described at all	Unable to judge based on the information presented
<b>Was MAYV exposure positivity measured in a valid way?<sup>1</sup></b>	PCR or viral culture or high-specificity serological method (e.g., PRNT)	Lower -specificity serological assay (i.e., HI and ELISA) without exclusion of sero-cross-reactivity	No confirmatory assay, presumptive exposure only	Unable to judge based on the information presented <sup>2</sup>

403 MAYV: Mayaro virus; PCR: polymerase chain reaction; PRNT: plaque-reduction neutralization test; HI:

404 hemagglutination inhibition; ELISA: enzyme-linked immunosorbent assay

405 <sup>1</sup>When multiple methods were used with varying validity (e.g., HI and PCR), we assigned a score to the most valid  
406 assay that detected MAYV positivity.

407 <sup>2</sup>When no MAYV-positive results were found, we assigned a score of N/A.

408

409 **Table 2. Unique occurrences by country and location type**

<b>Country</b>	<b>Point</b>	<b>Polygon</b>	<b>Total</b>
<b>Brazil</b>	35	100	135
<b>Peru</b>	10	24	34
<b>French Guiana</b>	23	8	31
<b>Suriname</b>	5	9	14
<b>Trinidad &amp; Tobago</b>	11	2	13
<b>Colombia</b>	3	6	9
<b>Venezuela</b>	4	3	7
<b>Panama</b>	4	2	6
<b>Bolivia</b>	1	5	6
<b>Ecuador</b>	0	5	5
<b>Mexico</b>	0	2	2
<b>Haiti</b>	0	1	1
<b>Costa Rica</b>	0	1	1

<b>Antigua</b>	0	1	1
<b>Guyana</b>	0	1	1
<b>TOTALS</b>	<b>96</b>	<b>170</b>	<b>266</b>

410

411 **Table 3. Mayaro virus occurrences by host type and spatial resolution**

	<b>Points</b>	<b>Custom</b>	<b>Admin3</b>	<b>Admin2</b>	<b>Admin1</b>	<b>Admin0</b>
<b>Human</b>	74	24	7	88	19	10
<b>Non-human animal</b>	7	2	0	20	4	1
<b>Arthropod</b>	15	3	1	2	1	2

412

### 413 **Code Availability**

414 No custom code was used to generate or process the data described in the manuscript.

## 415 References

416

417 1 Anderson, C. R., Downs, W. G., Wattley, G. H., Ahin, N. W. & Reese, A. A. Mayaro virus: a new  
418 human disease agent. II. Isolation from blood of patients in Trinidad, B.W.I. *Am J Trop Med Hyg* **6**, 1012-  
419 1016, doi:10.4269/ajtmh.1957.6.1012 (1957).

420 2 Causey, O. R. & Maroja, O. M. Mayaro virus: a new human disease agent. III. Investigation of an  
421 epidemic of acute febrile illness on the river Guama in Para, Brazil, and isolation of Mayaro virus as  
422 causative agent. *Am J Trop Med Hyg* **6**, 1017-1023 (1957).

423 3 LeDuc, J. W., Pinheiro, F. P. & Travassos da Rosa, A. P. An outbreak of Mayaro virus disease in  
424 Belterra, Brazil. II. Epidemiology. *Am J Trop Med Hyg* **30**, 682-688, doi:10.4269/ajtmh.1981.30.682  
425 (1981).

426 4 Schaeffer, M., Gajdusek, D. C., Lema, A. B. & Eichenwald, H. Epidemic jungle fevers among  
427 Okinawan colonists in the Bolivian rain forest. I. Epidemiology. *Am J Trop Med Hyg* **8**, 372-396,  
428 doi:10.4269/ajtmh.1959.8.372 (1959).

429 5 Auguste, A. J. *et al.* Evolutionary and Ecological Characterization of Mayaro Virus Strains Isolated  
430 during an Outbreak, Venezuela, 2010. *Emerg Infect Dis* **21**, 1742-1750, doi:10.3201/eid2110.141660  
431 (2015).

432 6 Forshey, B. M. *et al.* Arboviral etiologies of acute febrile illnesses in Western South America,  
433 2000-2007. *PLoS Negl Trop Dis* **4**, e787, doi:10.1371/journal.pntd.0000787 (2010).

434 7 Jonkers, A. H., Spence, L. & Karbaat, J. Arbovirus infections in Dutch military personnel stationed  
435 in Surinam. Further studies. *Trop Geogr Med* **20**, 251-256 (1968).

436 8 Talarmin, A. *et al.* Mayaro virus fever in French Guiana: isolation, identification, and  
437 seroprevalence. *Am J Trop Med Hyg* **59**, 452-456, doi:10.4269/ajtmh.1998.59.452 (1998).

438 9 Suhrbier, A., Jaffar-Bandjee, M. C. & Gasque, P. Arthritogenic alphaviruses--an overview. *Nat Rev*  
439 *Rheumatol* **8**, 420-429, doi:10.1038/nrrheum.2012.64 (2012).

440 10 Pan American Health Organization / World Health Organization. Epidemiological Alert: Mayaro  
441 Fever. (2019).

442 11 Ramshaw, R. E. *et al.* A database of geopositioned Middle East Respiratory Syndrome  
443 Coronavirus occurrences. *Sci Data* **6**, 318, doi:10.1038/s41597-019-0330-0 (2019).

444 12 Messina, J. P. *et al.* A global compendium of human dengue virus occurrence. *Sci Data* **1**,  
445 140004, doi:10.1038/sdata.2014.4 (2014).

446 13 Messina, J. P. *et al.* A global compendium of human Crimean-Congo haemorrhagic fever virus  
447 occurrence. *Sci Data* **2**, 150016, doi:10.1038/sdata.2015.16 (2015).

448 14 Pigott, D. M. *et al.* Global database of leishmaniasis occurrence locations, 1960-2012. *Sci Data* **1**,  
449 140036, doi:10.1038/sdata.2014.36 (2014).

450 15 Caicedo, E. Y. *et al.* The epidemiology of Mayaro virus in the Americas: A systematic review and  
451 key parameter estimates for outbreak modelling. *PLoS Negl Trop Dis* **15**, e0009418,  
452 doi:10.1371/journal.pntd.0009418 (2021).

- 453 16 Ganjian, N. & Riviere-Cinnamond, A. Mayaro virus in Latin America and the Caribbean. *Revista*  
454 *Panamericana de Salud Pública* **44**, doi:10.26633/rpsp.2020.14 (2020).
- 455 17 Celone, M. *et al.* A systematic review and meta-analysis of the potential non-human animal  
456 reservoirs and arthropod vectors of the Mayaro virus. *PLoS Negl Trop Dis* **15**, e0010016,  
457 doi:10.1371/journal.pntd.0010016 (2021).
- 458 18 Schmidt, J. R., Gajdusek, D. C., Schaffer, M. & Gorrie, R. H. Epidemic jungle fever among  
459 Okinawan colonists in the Bolivian rain forest. II. Isolation and characterization of Uruma virus, a newly  
460 recognized human pathogen. *Am J Trop Med Hyg* **8**, 479-487, doi:10.4269/ajtmh.1959.8.479 (1959).
- 461 19 U.S. Centers for Disease Control and Prevention. *Arbovirus Catalog.* ,  
462 <<https://www.cdc.gov/arbovat/VirusDetails.aspx?ID=293&SID=5>> (
- 463 20 Food and Agricultural Organization of the United Nations. The global administrative unit layers  
464 (GAUL): technical aspects. . (FAO, 2008).
- 465 21 Wieczorek, J., Guo, Q. & Hijmans, R. The point-radius method for georeferencing locality  
466 descriptions and calculating associated uncertainty. *Int. J. Geogr. Inf. Sci* **18**, 745-767 (2004).
- 467 22 Chapman, A. D. & Wieczorek, J. *Guide to Best Practices for Georeferencing.* (Global Biodiversity  
468 Information Facility. Available at: <http://www.gbif.org/resource/80536>, 2006).
- 469 23 Wieczorek, C. & Wieczorek, J. *Georeferencing Calculator,*  
470 <<http://georeferencing.org/georefcalculator/gc.html>> (2021).
- 471 24 Hill, E. *et al.* A database of geopositioned onchocerciasis prevalence data. *Sci Data* **6**, 67,  
472 doi:10.1038/s41597-019-0079-5 (2019).
- 473 25 Hoch, A. L., Peterson, N. E., LeDuc, J. W. & Pinheiro, F. P. An outbreak of Mayaro virus disease in  
474 Belterra, Brazil. III. Entomological and ecological studies. *Am J Trop Med Hyg* **30**, 689-698,  
475 doi:10.4269/ajtmh.1981.30.689 (1981).
- 476 26 Chisenga, C. C. *et al.* Sero-prevalence of arthropod-borne viral infections among Lukanga swamp  
477 residents in Zambia. *PloS One* **15**, e0235322, doi:10.1371/journal.pone.0235322 (2020).
- 478 27 Hassing, R. J., Leparc-Goffart, I., Tolou, H., van Doornum, G. & van Genderen, P. J. Cross-  
479 reactivity of antibodies to viruses belonging to the Semliki forest serocomplex. *Euro Surveill.* **15** (2010).
- 480 28 Escobar, L. E. & Craft, M. E. Advances and Limitations of Disease Biogeography Using Ecological  
481 Niche Modeling. *Front. Microbiol.* **7**, doi:10.3389/fmicb.2016.01174 (2016).
- 482 29 Deribe, K. *et al.* Mapping the global distribution of podoconiosis: Applying an evidence  
483 consensus approach. *PLoS Negl Trop Dis* **13**, e0007925, doi:10.1371/journal.pntd.0007925 (2019).
- 484 30 Deribe, K. *et al.* Predicting the environmental suitability and population at risk of podoconiosis in  
485 Africa. *PLoS Negl Trop Dis* **14**, e0008616, doi:10.1371/journal.pntd.0008616 (2020).
- 486 31 Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496**, 504-507,  
487 doi:10.1038/nature12060 (2013).
- 488 32 Kramer-Schadt, S. *et al.* The importance of correcting for sampling bias in MaxEnt species  
489 distribution models. *Divers. Distrib.* **19**, 1366-1379, doi:10.1111/ddi.12096 (2013).



490 33 Nsoesie, E. O. *et al.* Global distribution and environmental suitability for chikungunya virus, 1952  
491 to 2015. *Euro Surveill* **21**, doi:10.2807/1560-7917.Es.2016.21.20.30234 (2016).  
492