# Mapping suitability for Buruli ulcer at fine spatial scales across Africa: a modelling study

Hope Simpson MSc[1]*, Earnest Njih Tabah PhD[2], Richard Phillips MD[3], Michael Frimpong PhD[3], Issaka Maman PhD[4], Joseph Timothy PhD[1], Paul Saunderson MD[5], Rachel L Pullan PhD[1], Jorge Cano PhD[1]

*corresponding author: hope.simpson@lshtm.ac.uk

**Author Affiliations**

[1]London School of Hygiene and Tropical Medicine, United Kingdom
[2] National Yaws, Leishmaniasis, Leprosy and Buruli ulcer Control Programme, Cameroon
[3] School of Medical Sciences, Kwame Nkrumah University of Science and Technology, Ghana
[4] National Reference Laboratory for Buruli Ulcer Disease in Togo; Ecole Supérieure des Techniques Biologiques et Alimentaires (ESTBA), Laboratoire des Sciences Biologiques et des Substances Bioactives, Université de Lomé, Lomé, Togo
[5]Accelerating Integrated Management (AIM) Initiative, Accra, Ghana.

## Abstract

### Background

Buruli ulcer (BU) is a disabling and stigmatising neglected tropical disease (NTD). Its distribution and burden are unknown because of underdiagnosis and underreporting. It is caused by *Mycobacterium ulcerans*, an environmental pathogen whose environmental niche and transmission routes are not fully understood. Active BU case searches can limit morbidity by identifying cases and linking them to treatment, but these are mostly restricted to well-known endemic areas. A better understanding of environmental suitability for environmental reservoirs of *M. ulcerans* and BU disease would advance understanding of the disease's ecology and burden, and could inform targeted surveillance.

### Methodology/Principal Findings

We used previously compiled point-level datasets of BU and *M. ulcerans* occurrence, evidence for BU occurrence within national and sub-national areas, and diverse environmental datasets. We fitted relationships between BU and *M. ulcerans* occurrence and environmental predictors by applying regression and machine learning based algorithms, combined in an ensemble model to characterise the optimal ecological niche for the disease and bacterium across Africa at a resolution of 5km x 5km. Climate and atmospheric variables were the strongest predictors of both distributions, while indicators of human disturbance including damming and deforestation, drove local variation in suitability. We identified patchy foci of suitability throughout West and Central Africa, including areas with no previous evidence of the disease. Predicted suitability for *M. ulcerans* was wider but overlapping with that of BU. The estimated population living in areas predicted suitable for the bacterium and disease was 29.1 million.

### Conclusions/Significance

These maps could be used to inform burden estimations and case searches which would generate a more complete understanding of the spatial distribution of BU in Africa, and may guide control programmes to identify cases beyond the well-known endemic areas.

## Author summary

Like many neglected tropical diseases primarily affecting the rural poor, Buruli ulcer (BU) is under-detected and under-reported within routine health information systems. As such, the burden and distribution are not fully known, impeding appropriate targeting of health resources, control, and care for people affected. Having previously evaluated and mapped the existing evidence for BU and its causative agent *M. ulcerans*, we concluded that the disease was likely to occur beyond the range of known endemic areas. However, we were left with the question of where exactly these undetected cases might be occurring. Answering this question required a more fine-scale approach: BU is highly focal, presumably due to local variation in the environmental factors which determine suitability for *M. ulcerans* survival and transmission to humans. We used the compiled evidence and geographical datasets to build statistical models representing the relationship between environmental factors and previously reported cases. This allowed us to define the ecological niche of BU, and subsequently to identify areas across Africa where this niche was met, providing suitable conditions for the disease. We constructed separate models of suitability for *M. ulcerans*, using locations where its DNA had been detected in environmental sources. Unsurprisingly, suitability for *M. ulcerans* was predicted to be wider than, but geographically overlapping with that for BU. This implies that beyond the conditions necessary for survival of the bacterium, additional factors are required for transmission to humans. The high-resolution suitability maps we present are intended to guide case search activities which may identify endemic areas beyond the known endemic range. Data on the true prevalence of BU from targeted case searches within predicted-suitable areas will also allow us to validate and refine the models, and potentially predict the actual probability of cases occurring within predicted suitable areas.

## Introduction

Buruli ulcer (BU) is a chronic necrotizing disease of the skin and soft tissue, which causes debilitating symptoms and sequelae, associated with a high burden of morbidity and stigma for patients and economic costs for affected households [1-3]. These impacts are felt particularly strongly in impoverished rural communities with poor access to health services [3, 4]. The infectious agent is *Mycobacterium ulcerans*, a slow-growing environmental bacterium which may be transmitted from aquatic environments to humans via inhalation or ingestion, or by penetration of the skin [1, 5, 6]. The main control strategy is active case finding in endemic areas to promote early case detection and effective treatment, which limits disease progression [7, 8]. BU occurs mostly in tropical and subtropical areas of West and Central Africa, with smaller foci in parts of Asia, South America, the Western Pacific and Australasia [9]. However, the disease is recognised to be underdiagnosed and under-reported, and may occur undetected in other parts of the world [9-12].

The distribution of BU is presumably linked to environmental suitability for *M. ulcerans* survival and replication, as well as to human and environmental factors favouring transmission [13]. On a continental scale, BU appears to be limited by climatic factors: it is restricted to tropical and subtropical regions and absent from arid areas [14]. Within endemic areas, the disease shows a highly focal distribution [15-17], but reasons for this are not well understood, since the precise niche and transmission routes of *M. ulcerans* have been difficult to characterise [18]. The pathogen has only been cultured from environmental and animal samples a handful of times [19-21], although it has been detected by PCR in aquatic environments of endemic and non-endemic areas, and a in wide range of potential hosts including mammals, fish, amphibians, and aquatic and terrestrial insects [22-26]. Consistent with the ecology of an environmental pathogen, the distribution of *M. ulcerans* in the environment appears to be wider than that of BU, suggesting that factors beyond environmental suitability for *M. ulcerans* are required for transmission [13, 14, 27].

Our understanding of the pathways of BU infection is also limited, partly by its long and variable incubation period, which makes it difficult for patients and clinicians to attribute particular events or activities to disease acquisition [28]. Local spatial analysis has identified several environmental variables associated with increased BU incidence, primarily proximity to rivers, as well as environmental disturbance and land-use changes including deforestation, urbanisation, agriculturalization and mining [17, 29]. Case control studies have identified contact with unprotected waterbodies as a risk factor for disease [30], suggesting that activities which bring people into contact with water sources harbouring *M. ulcerans* increase the risk of disease acquisition [31-33].

Given the recognised scale of BU under-detection and under-reporting, it is likely that BU occurs beyond the known range of reported cases. A better understanding of potential suitability for the pathogen in the environment and the disease in humans would help to improve its surveillance and control in countries where is known to be endemic. Furthermore, characterisation of the environmental factors linked to suitability for *M. ulcerans* and BU may reveal areas at risk for disease emergence, or places harbouring unrecognised cases.

In this investigation, we aim to identify environmental factors which characterise the environmental niche of *M. ulcerans* and BU disease in humans, and to model their respective relationships with BU. These analyses will be used to identify areas of continental Africa which may be suitable for *M. ulcerans* or BU based on their environmental characteristics.

# Methods

## Data on Buruli ulcer and *M. ulcerans* distribution

We used previously compiled spatial datasets of point locations of recorded occurrences of BU disease in humans, and of detection of *M. ulcerans* genetic material in biotic and abiotic environmental samples [9, 34]. Data for the final models was extracted from the database on 03/01/2020.

BU occurrence locations were restricted to those where BU infection was confirmed by a positive result for PCR targeting IS2404, or histopathology consistent with BU disease. To explore the model's sensitivity to the case definition, we repeated the analysis using all locations where clinically diagnosed BU had been reported. We hereon refer to the two datasets as '*confirmed occurrences*' and '*all occurrences*' respectively.

The environmental dataset was restricted to locations where *M. ulcerans* DNA had been identified and distinguished from that of other mycobacteria: either by multiplex qPCR assays quantifying the relative copy numbers of IS2404, IS2606 and the KR-B domain [35]; byvariable nucleotide tandem repeat (VNTR); or mycobacterial interspersed repetitive unit (MIRU) typing [36, 37]. We hereon refer to this dataset as '*environmental occurrences*'.

All records were restricted to locations with reliable geographical coordinates and deduplicated by geographical location. Human and environmental locations were weighted by the year and the specificity of confirmatory tests reported (S1 Text).

## Environmental datasets used in ecological modelling

We assembled gridded datasets of 51 environmental variables considered relevant to the ecological niche of *M. ulcerans* [18]. All variables and their sources are shown in Table 1 and full details are provided in S3 Text.

## Variable selection

We compiled the gridded predictor variables at a resolution of 5km x 5km within a rectangular area of West Africa from latitude -13.57195, longitude-4.11032, to lat. 16.67107, long. 14.493. This area contained 94% of all BU occurrence locations, 95% of confirmed BU occurrence locations, and all environmental occurrence locations. We extracted the values of predictor variables at the locations of BU cases (all occurrences) and environmental occurrences of *M. ulcerans* DNA.

We used principal components analysis (PCA) to identify the minimum set of variables that best characterized the environment at observations of BU and of *M. ulcerans* from the global environment. We undertook separate PCAs on the human (confirmed cases) and environmental datasets, selecting variables that contributed most strongly to the minimum number of principal components collectively accounting for at least 80% of the total dataset variance.

## Pseudoabsence and Background data

The occurrence data used in the modelling framework were supplemented with systematically generated pseudo-absence and background data. The use of artificial absence data is a common approach in species distribution modelling, designed to account for geographically biased presence data and sparse absence data [38]. The terms background and pseudoabsence are often used interchangeably to describe artificial absence data, though Elith and Hijmans distinguish them on the basis that background data are intended to characterise the 'environmental domain' of the study-helping to account for geographical bias, while pseudoabsence data represent areas assumed to be unsuitable for the species and are intended to capture the environment in these areas [39]. We hereon use the terms in this sense.

Pseudoabsence and background points were both selected from a restricted geographical extent around occurrence points, defined by the spatial structure of the occurrence predictors. This has been recommended by previous authors as way to provide an ecologically meaningful definition of the study range [40]. Pseudoabsence points were sampled at higher density in areas of weaker evidence according to a systematic review of the geographical distribution of BU [41]. Background points were sampled at higher density around recorded occurrence points. More details on the generation of pseudoabsence and background points are provided in S2 Text.

Pseudoabsence and background weights were uniform within datasets and assigned so their sum was equal to the sum of occurrence weights in each model [42]. Human background points were restricted to a minimum distance of 10km from any occurrence location, and environmental background points were restricted to 10km from any human or environmental occurrence location.

The distribution of BU and *M. ulcerans* occurrences, pseudoabsence and background points are shown in S1 Figure and S2 Figure.

**Ensemble modelling**

The environmental factors selected through PCA were used as predictor variables. We used the *biomod2* package in R [43, 44] to implement seven algorithms: generalized linear models (GLM), generalized additive models (GAM), generalized boosted regression models (GBM), artificial neural networks (ANN), multiple adaptive regression splines (MARS), maximum entropy (MaxEnt) and random forest (RF).

Individual model algorithms were each run 20 times with a random sample of 80% of data points, and evaluated with the remaining 20%. For each algorithm we calculated the mean true skill statistic (TSS), the mean positive correctly classified (PCC) and the mean area under the curve (AUC) of the receiver operation characteristic (ROC) [45]. Models with mean AUC above 0.8 were integrated in an ensemble using committee averaging to attribute higher weight to better performing models.

We plotted the importance values representing each variable's contribution to the model and created marginal effect plots for the modelled covariates in the highest performing model ensemble.

**Estimating total population living in suitable areas**

We calculated the total area suitable for BU, *M. ulcerans*, and the total area suitable for both, and extracted estimates of the population living in each of these areas from a raster representing estimated number of people per $1km^2$ grid square in 2020 [46].

# Results

## Datasets of BU occurrence in humans and *M. ulcerans* DNA detection in the environment

The modelled data included 2,183 unique point locations with reported cases of BU in Africa (Figure 1A). BU was confirmed by PCR or histopathology at 738 unique locations. There were 91 unique locations where *M. ulcerans* DNA had been detected by MIRU, VNTR or qPCR (Figure 1B). The dataset of clinically diagnosed human cases represented 16 countries, mostly in West and Central Africa, with a few in East and southeast Africa. The confirmed cases were restricted to 13 countries. The time period of human case detection was from 1957 to 2019. The median year of diagnosis was 2010. The 91 records of environmental detection of *M. ulcerans* represented three countries: Ghana, Cameroon and Benin, and covered the period from 2006 to 2018 with a median year of detection of 2013.

**Figure 1A**: Selection of BU occurrence points from BU database

**Figure 1B:** Selection of model environmental occurrence points from global database of MU occurrences

## Principal components analysis

Ten principal components (PCs) collectively contributed 86% of variation in the human BU occurrence locations. Seven PCs characterised the environmental locations. The variables selected for each model are shown in S3 Text. The mean contribution of the most important predictors of human BU occurrence dropped sharply at 150km, while the contribution of the predictors of environmental detection of *M. ulcerans* declined at 300km (S4 Text).

## *Environmental suitability for BU*

All individual distribution models performed well with ROC scores above 0.8 (S5 Text). Mean PCC scores were between 79.3 and 91.6% and mean TSS scores were between 0.57 and 0.79 (Figures A and B, S5 Text).RF performed best with a mean PCC of 91.6%, a mean TSS of 0.79 and mean ROC 0.95. The final ensemble model showed an overall mean ROC of 0.96 with sensitivity of 87.1% and specificity of 92.9%. The mean TSS was 0.80 and the mean kappa score was 0.80 (Table 2).

211 Table 2: Validation metrics for ensemble models for BU and *M. ulcerans* suitability

| | | Weighted Mean | Lower CI | Upper CI |
|---|---|---|---|---|
| BU suitability | TSS | 0.797 | 0.797 | 0.794 |
| | ROC | 0.963 | 0.961 | 0.963 |
| | kappa | 0.801 | 0.804 | 0.797 |
| MU suitability | TSS | 0.885 | 0.885 | 0.896 |
| | ROC | 0.96 | 0.96 | 0.961 |
| | kappa | 0.894 | 0.894 | 0.894 |

212

213 The minimum temperature of the coldest month was the strongest contributor to the RF models,
214 followed by PET and distance to dams (Figure A in S6 Text).Environments with minimum
215 temperature below 18°C were unsuitable, but suitability increased sharply at this temperature,
216 decreased a little from 20-21°C, before increasing again, remaining high up to 24°C (Figure A in S7
217 Text). Optimal values of PET were between 1,000 and 1,500mm per month, corresponding to
218 tropical rainforest canopy cover. There was a gradual decline in suitability for BU with increasing
219 distance to the nearest dam up to a distance of 80km, after which suitability dropped rapidly.

220 The overall distribution was constrained to humid tropical areas and local scale variation appeared
221 to be driven by hydrological features and deforestation patterns (Figure 2). The total area predicted
222 to be suitable for BU was 338,500 $km^2$, and the total population living in areas predicted suitable was
223 69.7 million (Table 3). Pockets of suitability for BU totalling >100 $km^2$ were predicted in 16 countries
224 in Africa, including all 12 countries along the west-central African coastline from Liberia to Angola
225 (S2 Table). Angola had the widest area predicted suitable, followed by the Democratic Republic of
226 the Congo, although the patches of suitability predicted in these countries were associated with high
227 uncertainty. Nigeria had the largest population at risk, with 17.8 million predicted to be living in
228 areas suitable for BU, followed by the Democratic Republic of the Congo where 10.8 million were
229 predicted to be living in suitable areas (S2 Table).

230 The model including all cases of BU (S3 Figure) gave similar results to the model including confirmed
231 cases only. The Pearson coefficient of correlation between the two models was 0.86.

232 Figure 2: Predicted environmental suitability for the occurrence of BU disease in humans and
233 associated error of prediction.

234 ***Environmental suitability for M. ulcerans***

235 All models performed well with ROC above 0.8, apart from MAXENT Phillips which was excluded
236 from the ensemble model. Mean PCC varied from 0.80- 0.89 between models and mean TSS was
237 between 0.54 and 0.77 (Figures C and D, S5 Text). RF outperformed other algorithms in predicting

238 the occurrence of *M. ulcerans*. The final ensemble model had a mean TSS score of 0.87, with a
239 sensitivity of 92.3 and specificity of 94.5% (Table 2). The ROC sore was 0.99 and the kappa score was
240 0.87.

241 The minimum temperature of the coldest month was the strongest predictor of *M. ulcerans*
242 occurrence in the RF models, accounting for 24.6% of all variance in the model (Figure B in S6 Text).
243 Distance to deforested areas was also a strong predictor, accounting for 23.7% of the variance.
244 Suitability was low at coldest month minimum temperatures below 18°C, and increased sharply to a
245 peak at 25°C. There was a strong response to the distance to deforested areas, with high suitability
246 at close range to deforested areas, decreasing sharply at 25km (Figure B in S6 Text).

247 The overall distribution appeared to be restricted by suitability for minimum temperature and
248 precipitation in the driest month, with local variation driven by deforestation (Figure 3). The total
249 area predicted to be suitable for *M. ulcerans* was 833,975km$^2$, and the total population living in
250 areas predicted suitable was 71.2 million (Table 3). Pockets of suitability were predicted in 31
251 countries (S2 Table). The DRC had the widest area predicted suitable (184,500 km$^2$) followed by Côte
252 d'Ivoire (120,600 km$^2$).The highest population living in suitable areas was in Nigeria (26.7 million).

253 Figure 3: Predicted environmental suitability for the occurrence of *M. ulcerans* in the environment
254 and associated error of prediction.

255 **Overlap of suitability for BU and *M. ulcerans***

256 The total area predicted to be suitable for both BU and *M. ulcerans* was 126,775 km$^2$, with 29.1
257 million people predicted to be living in areas at risk. There were notable differences in the extents of
258 the areas predicted suitable for BU disease and environmental *M. ulcerans* (Figure 4). There were
259 wide areas in Sierra Leone and Cote d'Ivoire predicted suitable for *M. ulcerans* but not for BU
260 disease, which was restricted to smaller pockets within these countries. Suitability for *M. ulcerans*
261 was also predicted outside the predicted range of BU across central Africa and along the north-west
262 cost of Mozambique, eastern Tanzania and coastal Kenya. In contrast, there were large patches in
263 Central Africa predicted suitable for BU but not suitable for *M. ulcerans*. Ghana followed by
264 Cameroon had the widest area predicted suitable for both BU and *M. ulcerans*. The highest
265 populations living in areas predicted suitable for both BU and *M. ulcerans* were in Nigeria and
266 Ghana, with 11.4 and 8.0 million respectively at risk.

267 Figure 4: Predicted overlap of environmental suitability for BU and of *M. ulcerans* occurrence.

268

269 Table 3: Total area predicted suitable and population in areas at risk for Buruli ulcer, M. ulcerans,

270 and both, in Africa

| | Total area suitable ($km^2$) | *Lower bound* | *Upper bound* | Population in suitable areas | *Lower bound* | *Upper bound* |
|---|---|---|---|---|---|---|
| BU | 338,500 | *283,950* | *425,600* | 69,740,630 | *62,573,091* | *83,050,754* |
| MU | 912,600 | *767,775* | *1,112,275* | 88,542,574 | *75,380,436* | *104,645,747* |
| BU & MU | 126,775 | *94,800* | *177,700* | 29,124,902 | *22,597,071* | *39,371,153* |

271

272 Suitability for BU and *M. ulcerans* is shown by country in S1 Supporting Maps.

## Discussion

We have used ecological niche modelling to identify environmental factors associated with the occurrence of Buruli ulcer and its causative agent *M. ulcerans*, and to predict environmental suitability for the disease and bacterium across continental Africa. Incorporating existing data on BU distribution and a geo-environmental definition of the range of occurrences, the resulting maps represent evidence-based predictions within a relevant spatial context.

There was substantial overlap in the factors contributing to suitability for human cases and environmental occurrence. Both BU and *M. ulcerans* were constrained to particular bioclimatic zones by environmental predictors which varied over large areas, characterising the humid tropical realm where BU is endemic in Africa. Our finding that areas with minimum temperature less than 18°C were unsuitable for BU supports evidence for a different epidemiology of the disease in Africa compared to endemic areas of temperate Australia and Japan [47]. Local-scale variation in factors including the distance to dams, deforestation and hydrology resulted in a patchy distribution of predicted suitability, consistent with our understanding of the epidemiology of BU, which is recognised to be highly focal in endemic settings [48].

We identified pockets of suitability for BU in patchy foci throughout the known-endemic range of the disease, particularly in the tropical zones of countries around the Gulf of Guinea. Throughout this range, suitability was predicted in regions not previously recognised as endemic. For example, two foci of suitability were predicted in Equatorial Guinea, which had no evidence of cases reported in peer-reviewed literature. However, these two areas correspond to the origins of a number of cases diagnosed by an expert in BU between 1995 and 2005 [49, 50].

Some locations in northern Cameroon with previous evidence of PCR confirmed BU were found to be unsuitable for the disease. This discordance may be due to the model's failure to identify suitable environments in areas of lower BU incidence. However, given the great volume of surveillance data collected by the well-established BU control programme in Cameroon, some patients are likely to

298 have been diagnosed outside the region where they acquired the disease [51], and we consider it
299 plausible that some regions where BU has been recorded are not actually suitable for transmission.

300 The suitability maps provide a depiction of areas potentially at risk for BU beyond what is known
301 from the distribution of reported cases, currently the basis for targeting of surveillance and control.
302 Given the recognised scale of underreporting of BU [41], the current approach is likely to exclude
303 cases outside of known disease foci, and we suggest that areas predicted suitable for BU could be
304 considered as targets for case finding activities, with the aim of identifying unrecognised foci and
305 patients not known to the health system. Based on the wide areas of suitability predicted by this
306 work and existing evidence of under-reporting of BU [52], the south of Nigeria would be a key target
307 for case finding activities. There were wide areas predicted suitable in the Republic of Congo, the
308 DRC and Angola, although these predictions were associated with significant uncertainty, which
309 should be considered in the design of any future surveys. Suitable areas of Equatorial Guinea with
310 historical evidence of cases would also be targets for case finding, although in this case the predicted
311 suitable areas are more restricted, potentially necessitating a more stratified approach. A
312 comparable approach has been applied to target malaria elimination efforts to transmission
313 hotspots predicted through geospatial risk mapping [53], employing environmental modelling to
314 impute risk in the absence of full surveillance coverage.

315 The model predictions could also be used to inform the design of cross-sectional surveys for BU,
316 combining exhaustive case searches with environmental modelling to achieve robust estimates of
317 prevalence. In a nationwide survey for podoconiosis in Cameroon, the selection of survey
318 communities was stratified according predicted suitability for the disease based on a model trained
319 mainly using data from Ethiopia [54]. This survey identified higher rates of podoconiosis in
320 communities that were predicted suitable, implying a benefit in terms of the cost per case identified,
321 compared to a survey employing random selection of survey communities.

322 The validation metrics we calculated demonstrate the ability of the models to predict BU and *M.*
323 *ulcerans* occurrence with high accuracy. However, these measures do not indicate the models'
324 generalisability to areas beyond the range of known locations. Validation against external datasets
325 would be required to assess this quality and is a target for future analysis.

326 The scale of analysis (grid cells at 5km x 5km) may have limited our ability to quantify the effect of
327 predictors varying over small geographical scales. For example, the relatively small contribution of
328 distance to waterways on suitability for BU was a surprising result, given that proximity to and
329 contact with rivers have been identified as risk factors for BU disease [13, 18]. However, the
330 variation we were able to capture in this predictor was limited, since most of the land area in
331 tropical and sub-tropical Africa is within 10km of a river or stream, and much is within 5km. A
332 previous analysis of land use and landcover and BU presence at large spatial scales in Benin found no
333 association of BU at community level with proximity to water bodies [55]. The scale of analysis may
334 also have limited our ability to capture fine scale variation in environmental suitability for BU. Our
335 models predicted large contiguous areas of suitability in some areas with suitable bioclimatic
336 conditions and within close proximity to stable night lights and deforested areas. Such areas may be
337 suitable in reality, but exhibit an uneven distribution of disease due to factors not included in our
338 models.

339 We have identified areas of high suitability for BU and *M. ulcerans* within known endemic-areas, and
340 in areas not currently recognised as endemic, but with evidence of possible undiagnosed or
341 misdiagnosed BU. The population at highest risk of BU is within areas where BU and *M. ulcerans*
342 niches overlap, comprising almost 30 million people in 2020. The focal nature of BU distribution, the
343 recognised scale of under-detection, and the impact of late diagnosis on disease severity strongly
344 suggest a targeted approach to active case finding as a means to control this disease. The fine-scale,
345 evidence-based predictions presented here could provide a tool to target such efforts, which will

346 help to increase the proportion of cases linked to treatment, and contribute further research to

347 establish the burden and distribution of this devastating disease.

348

349

## Acknowledgements

The AIM Initiative was the sole funder of this work. We would like to recognise the contribution of all health workers, researchers and data managers who recorded cases of Buruli ulcer which were compiled into the global database of infections underpinning this study.

## Author contributions

RP and JC acquired the funding that enabled this project. JC conceptualised the study. JC and HS together developed the investigation and methodology- including the software (code for the analysis), and produced the visualisations presented. HS curated the data analysed, some of which was originally collected by other authors: ET, RP, MF, IM and JT. HS undertook the formal analysis and prepared the original draft. All authors critically reviewed the draft.

## Supporting Information

**S1 Text**: Weighting of occurrence points

**S2 Text:** Selection of background and pseudoabsence points

**S3 Text:** Environmental variables used in modelling

**S1 Figure**: Distribution and weights of confirmed BU occurrence locations, background and pseudoabsence points

**S2 Figure**: Distribution and weights of environmental *M. ulcerans* DNA occurrence locations, background and pseudoabsence points

**S3 Figure**: Predicted environmental suitability for the occurrence of BU disease in humans and associated error of prediction, including all clinically diagnosed cases of BU

**S4 Text:** Spatial structure of occurrence predictors

**S5 Text**: Individual model performance evaluation statistics

**S6 Text**: Variable importance plots for random forest models

**S7 Text**: Marginal effect plots for random forest models

**S1 Table**: Total area predicted suitable and population living in suitable areas for Buruli ulcer, *M. ulcerans*, and both, by country in African continent.

## References

1. Walsh DS, Portaels F, Meyers WM. Buruli ulcer (Mycobacterium ulcerans infection). Transactions of the Royal Society of Tropical Medicine and Hygiene. 2008;102(10):969-78. Epub 2008/07/29. doi: 10.1016/j.trstmh.2008.06.006. PubMed PMID: 18657836.

2. Roltgen K, Pluschke G. Epidemiology and disease burden of Buruli ulcer: a review. Research and Reports in Tropical Medicine. 2015;6:59-73. doi: 10.2147/rrtm.s62026. PubMed PMID: WOS:000219035500008.

3. Amoakoh HB, Aikins M. Household cost of out-patient treatment of Buruli ulcer in Ghana: a case study of Obom in Ga South Municipality. BMC health services research. 2013;13:507. Epub 2013/12/10. doi: 10.1186/1472-6963-13-507. PubMed PMID: 24313975; PubMed Central PMCID: PMCPmc4029088.

4. Chukwu JN, Meka AO, Nwafor CC, Oshi DC, Madichie NO, Ekeke N, et al. Financial burden of health care for Buruli ulcer patients in Nigeria: the patients' perspective. International health. 2017;9(1):36-43. Epub 2016/12/18. doi: 10.1093/inthealth/ihw056. PubMed PMID: 27986841.

5. Duker AA, Portaels F, Hale M. Pathways of Mycobacterium ulcerans infection: a review. Environment international. 2006;32(4):567-73. Epub 2006/02/24. doi: 10.1016/j.envint.2006.01.002. PubMed PMID: 16492390.

6. Rodhain F. [Buruli ulcer: hypothetical modes of transmission of Mycobacterium ulcerans]. Bulletin de l'Academie nationale de medecine. 2012;196(3):685-90; discussion 90-1. Epub 2013/03/12. PubMed PMID: 23472356.

7. Organisation WH. Buruli ulcer (Mycobacterium ulcerans infection) 2019 [27/08/2019]. Available from: https://www.who.int/news-room/fact-sheets/detail/buruli-ulcer-(mycobacterium-ulcerans-infection).

8. Phillips RO, Robert J, Abass KM, Thompson W, Sarfo FS, Wilson T, et al. Rifampicin and clarithromycin (extended release) versus rifampicin and streptomycin for limited Buruli ulcer lesions: a randomised, open-label, non-inferiority phase 3 trial. The Lancet. 2020.

9. Simpson H, Deribe K, Tabah EN, Peters A, Maman I, Frimpong M, et al. Mapping the global distribution of Buruli ulcer: a systematic review with evidence consensus. Lancet Glob Health.

2019;7(7):e912-e22. Epub 2019/06/16. doi: 10.1016/S2214-109X(19)30171-8. PubMed PMID: 31200890; PubMed Central PMCID: PMCPMC6614043.

10. Kanga JM. Aspects épidémiologiques de l'ulcère de Buruli en Côte d'Ivoire : résultats d'une enquête nationale. Bull Soc Pathol Exot. 2001;94:46-51.

11. Amofah G, Bonsu F, Tetteh C, Okrah J, Asamoa K, Asiedu K, et al. Buruli ulcer in Ghana: results of a national case search. Emerging infectious diseases. 2002;8(2):167-70. Epub 2002/03/19. doi: 10.3201/eid0802.010119. PubMed PMID: 11897068; PubMed Central PMCID: PMCPmc2732443.

12. Noeske J, Kuaban C, Rondini S, Sorlin P, Ciaffi L, Mbuagbaw J, et al. Buruli ulcer disease in Cameroon rediscovered. The American journal of tropical medicine and hygiene. 2004;70(5):520-6. Epub 2004/05/25. PubMed PMID: 15155984.

13. Garchitorena A, Roche B, Kamgang R, Ossomba J, Babonneau J, Landier J, et al. Mycobacterium ulcerans ecological dynamics and its association with freshwater ecosystems and aquatic communities: results from a 12-month environmental survey in Cameroon. PLoS neglected tropical diseases. 2014;8(5):e2879. doi: 10.1371/journal.pntd.0002879. PubMed PMID: 24831924; PubMed Central PMCID: PMCPMC4022459.

14. Williamson HR, Benbow ME, Nguyen KD, Beachboard DC, Kimbirauskas RK, McIntosh MD, et al. Distribution of Mycobacterium ulcerans in buruli ulcer endemic and non-endemic aquatic sites in Ghana. PLoS neglected tropical diseases. 2008;2(3):e205. Epub 2008/03/28. doi: 10.1371/journal.pntd.0000205. PubMed PMID: 18365034; PubMed Central PMCID: PMCPmc2268743.

15. Campbell LP, Finley AO, Benbow ME, Gronseth J, Small P, Johnson RC, et al. Spatial Analysis of Anthropogenic Landscape Disturbance and Buruli Ulcer Disease in Benin. PLoS neglected tropical diseases. 2015;9(10):e0004123. Epub 2015/10/17. doi: 10.1371/journal.pntd.0004123. PubMed PMID: 26474482; PubMed Central PMCID: PMCPmc4608567.

16. Wu UI, Holland SM. Host susceptibility to non-tuberculous mycobacterial infections. The Lancet Infectious diseases. 2015;15(8):968-80. Epub 2015/06/08. doi: 10.1016/s1473-3099(15)00089-4. PubMed PMID: 26049967.

17. Landier J, Gaudart J, Carolan K, Lo Seen D, Guegan JF, Eyangoh S, et al. Spatio-temporal patterns and landscape-associated risk of Buruli ulcer in Akonolinga, Cameroon. PLoS neglected

435 tropical diseases. 2014;8(9):e3123. Epub 2014/09/05. doi: 10.1371/journal.pntd.0003123. PubMed
436 PMID: 25188464; PubMed Central PMCID: PMCPmc4154661.

437 18. Merritt RW, Walker ED, Small PL, Wallace JR, Johnson PD, Benbow ME, et al. Ecology and
438 transmission of Buruli ulcer disease: a systematic review. PLoS neglected tropical diseases.
439 2010;4(12):e911. Epub 2010/12/24. doi: 10.1371/journal.pntd.0000911. PubMed PMID: 21179505;
440 PubMed Central PMCID: PMCPmc3001905.

441 19. Portaels F, Meyers WM, Ablordey A, Castro AG, Chemlal K, de Rijk P, et al. First cultivation
442 and characterization of Mycobacterium ulcerans from the environment. PLoS neglected tropical
443 diseases. 2008;2(3):e178. Epub 2008/03/28. doi: 10.1371/journal.pntd.0000178. PubMed PMID:
444 18365032; PubMed Central PMCID: PMCPMC2268003.

445 20. Aboagye SY, Danso E, Ampah KA, Nakobu Z, Asare P, Otchere ID, et al. Isolation of
446 Nontuberculous Mycobacteria from the Environment of Ghanian Communities Where Buruli Ulcer Is
447 Endemic. Applied and environmental microbiology. 2016;82(14):4320-9. Epub 2016/05/22. doi:
448 10.1128/aem.01002-16. PubMed PMID: 27208141; PubMed Central PMCID: PMCPMC4959205.

449 21. Marsollier L, Aubry J, Milan G, Brodin P. Aquatic insects and transmission of Mycobacterium
450 ulcerans. M S-Med Sci. 2007;23(6-7):572-5. doi: 10.1051/medsci/20072367572. PubMed PMID:
451 WOS:000248183300005.

452 22. Willson SJ, Kaufman MG, Merritt RW, Williamson HR, Malakauskas DM, Benbow ME. Fish
453 and amphibians as potential reservoirs of Mycobacterium ulcerans, the causative agent of Buruli
454 ulcer disease. Infection ecology & epidemiology. 2013;3. Epub 2013/02/27. doi:
455 10.3402/iee.v3i0.19946. PubMed PMID: 23440849; PubMed Central PMCID: PMCPmc3580280.

456 23. Eric Benbow M, Kimbirauskas R, McIntosh MD, Williamson H, Quaye C, Boakye D, et al.
457 Aquatic macroinvertebrate assemblages of Ghana, West Africa: understanding the ecology of a
458 neglected tropical disease. EcoHealth. 2014;11(2):168-83. Epub 2013/12/07. doi: 10.1007/s10393-
459 013-0886-7. PubMed PMID: 24306551.

460 24. Carson C, Lavender CJ, Handasyde KA, O'Brien CR, Hewitt N, Johnson PD, et al. Potential
461 wildlife sentinels for monitoring the endemic spread of human buruli ulcer in South-East australia.
462 PLoS neglected tropical diseases. 2014;8(1):e2668. Epub 2014/02/06. doi:
463 10.1371/journal.pntd.0002668. PubMed PMID: 24498452; PubMed Central PMCID:
464 PMCPmc3907424.

25. McIntosh M, Williamson H, Benbow ME, Kimbirauskas R, Quaye C, Boakye D, et al. Associations between Mycobacterium ulcerans and aquatic plant communities of West Africa: implications for Buruli ulcer disease. EcoHealth. 2014;11(2):184-96. Epub 2014/01/21. doi: 10.1007/s10393-013-0898-3. PubMed PMID: 24442959.

26. Marion E, Chauty A, Yeramian E, Babonneau J, Kempf M, Marsollier L. A case of guilt by association: Water bug bite incriminated in M. ulcerans infection. International journal of mycobacteriology. 2014;3(2):158-61. Epub 2014/06/01. doi: 10.1016/j.ijmyco.2014.01.004. PubMed PMID: 26786340.

27. Williamson HR, Benbow ME, Campbell LP, Johnson CR, Sopoh G, Barogui Y, et al. Detection of Mycobacterium ulcerans in the environment predicts prevalence of Buruli ulcer in Benin. PLoS neglected tropical diseases. 2012;6(1):e1506. Epub 2012/02/04. doi: 10.1371/journal.pntd.0001506. PubMed PMID: 22303498; PubMed Central PMCID: PMCPmc3269429.

28. Johnson PDR. Buruli Ulcer in Australia. In: Pluschke G, Röltgen K, editors. Buruli Ulcer: Mycobacterium Ulcerans Disease. Cham: Springer International Publishing; 2019. p. 61-76.

29. Wu J, Tschakert P, Klutse E, Ferring D, Ricciardi V, Hausermann H, et al. Buruli Ulcer Disease and Its Association with Land Cover in Southwestern Ghana. PLoS neglected tropical diseases. 2015;9(6):e0003840. Epub 2015/06/20. doi: 10.1371/journal.pntd.0003840. PubMed PMID: 26091265; PubMed Central PMCID: PMCPmc4474842.

30. N'Krumah R T, Kone B, Tiembre I, Cisse G, Pluschke G, Tanner M, et al. Socio-Environmental Factors Associated with the Risk of Contracting Buruli Ulcer in Tiassale, South Cote d'Ivoire: A Case-Control Study. PLoS neglected tropical diseases. 2016;10(1):e0004327. Epub 2016/01/09. doi: 10.1371/journal.pntd.0004327. PubMed PMID: 26745723; PubMed Central PMCID: PMCPmc4712845.

31. Marston B, Diallo MO, Horsburgh CR, Diomande I. Emergence of Buruli ulcer Disease in the Daloa region of Côte d'Ivoire. The American journal of tropical medicine and hygiene. 1995;52:291-24.

32. Raghunathan PL, Whitney EA, Asamoa K, Stienstra Y, Taylor TH, Jr., Amofah GK, et al. Risk factors for Buruli ulcer disease (Mycobacterium ulcerans Infection): results from a case-control study in Ghana. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2005;40(10):1445-53. Epub 2005/04/22. doi: 10.1086/429623. PubMed PMID: 15844067.

33. Aiga H, Amano T, Cairncross S, Adomako J, Nanas OK, Coleman S. Assessing water-related risk factors for Buruli ulcer: a case-control study in Ghana. The American journal of tropical medicine and hygiene. 2004;71(4):387-92. Epub 2004/11/02. PubMed PMID: 15516631.

34. Simpson H, Deribe, K, Tabah, EN, Maman, I, Frimpong, M, Ampadu, E, Saunderson, P, Pullan, RL and Cano, J. Occurrences of Buruli ulcer infections in humans and Mycobacterium ulcerans in animal and environmental samples. 2018.

35. Fyfe JA, Lavender CJ, Johnson PD, Globan M, Sievers A, Azuolas J, et al. Development and application of two multiplex real-time PCR assays for the detection of Mycobacterium ulcerans in clinical and environmental samples. Applied and environmental microbiology. 2007;73(15):4733-40. Epub 2007/05/29. doi: 10.1128/aem.02971-06. PubMed PMID: 17526786; PubMed Central PMCID: PMCPMC1951036.

36. Lavender CJ, Stinear TP, Johnson PD, Azuolas J, Benbow ME, Wallace JR, et al. Evaluation of VNTR typing for the identification of Mycobacterium ulcerans in environmental samples from Victoria, Australia. FEMS microbiology letters. 2008;287(2):250-5. Epub 2008/08/30. doi: 10.1111/j.1574-6968.2008.01328.x. PubMed PMID: 18754785.

37. Stragier P, Ablordey A, Meyers WM, Portaels F. Genotyping Mycobacterium ulcerans and Mycobacterium marinum by using mycobacterial interspersed repetitive units. J Bacteriol. 2005;187(5):1639-47. doi: 10.1128/jb.187.5.1639-1647.2005. PubMed PMID: WOS:000227191600011.

38. Elith* J, H. Graham* C, P. Anderson R, Dudík M, Ferrier S, Guisan A, et al. Novel methods improve prediction of species' distributions from occurrence data. Ecography. 2006;29(2):129-51.

39. Hijmans RJaE, J. Species distribution modeling with R. 2017.

40. Senay SD, Worner SP, Ikeda T. Novel three-step pseudo-absence selection technique for improved species distribution modelling. PloS one. 2013;8(8):e71218.

41. Simpson H, Deribe K, Tabah EN, Peters A, Maman I, Frimpong M, et al. Mapping the global distribution of Buruli ulcer: a systematic review with evidence consensus. The Lancet Global Health. 2019;7(7):e912-e22.

522 42. Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. Selecting pseudo-absences for species
523 distribution models: how, where and how many? Methods in ecology and evolution. 2012;3(2):327-
524 38.

525 43. Stoffel V, Barthelme B, Chague F. [Tropical ecopathology: up hill and down dale Buruli ulcer].
526 Sante publique (Vandoeuvre-les-Nancy, France). 2005;17(2):191-7. Epub 2005/07/09. PubMed
527 PMID: 16001561.

528 44. Wilfried Thuiller DG, Robin Engler and Frank Breiner. biomod2: Ensemble Platform for
529 Species Distribution Modeling. R package version 3.3-15/r728. 2017.

530 45. Liu C, White M, Newell G. Measuring and comparing the accuracy of species distribution
531 models with presence–absence data. Ecography. 2011;34(2):232-43.

532 46. School of Geography and Environmental Science UoS. WorldPop Africa Continental
533 Population Datasets (2000- 2020). WorldPop 2018.

534 47. Yotsu RR, Suzuki K, Simmonds RE, Bedimo R, Ablordey A, Yeboah-Manu D, et al. Buruli Ulcer:
535 a Review of the Current Knowledge. Current Tropical Medicine Reports. 2018:1-10.

536 48. Roltgen K, Qi W, Ruf MT, Mensah-Quainoo E, Pidot SJ, Seemann T, et al. Single nucleotide
537 polymorphism typing of Mycobacterium ulcerans reveals focal transmission of buruli ulcer in a highly
538 endemic region of Ghana. PLoS neglected tropical diseases. 2010;4(7):e751. Epub 2010/07/24. doi:
539 10.1371/journal.pntd.0000751. PubMed PMID: 20652033; PubMed Central PMCID:
540 PMCPmc2907412.

541 49. III IdSC. Buruli ulcer in Equatorial Guinea [presentation]. In press 2006.

542 50. Ham A. AFIP Mycobacteriology Chief continues to attack Buruli ulcer in Third World
543 countries. the AFIP Letter. 2000;158(3):5.

544 51. Agbo I. The gendered impact of Buruli ulcer on the household production of health: why
545 decentralization favours women. Tropical Medicine & International Health. 2015;20:298-. PubMed
546 PMID: WOS:000360758801329.

547 52. Vincent QB, Ardant MF, Adeye A, Goundote A, Saint-Andre JP, Cottin J, et al. Clinical
548 epidemiology of laboratory-confirmed Buruli ulcer in Benin: a cohort study. The Lancet Global
549 health. 2014;2(7):e422-30. Epub 2014/09/05

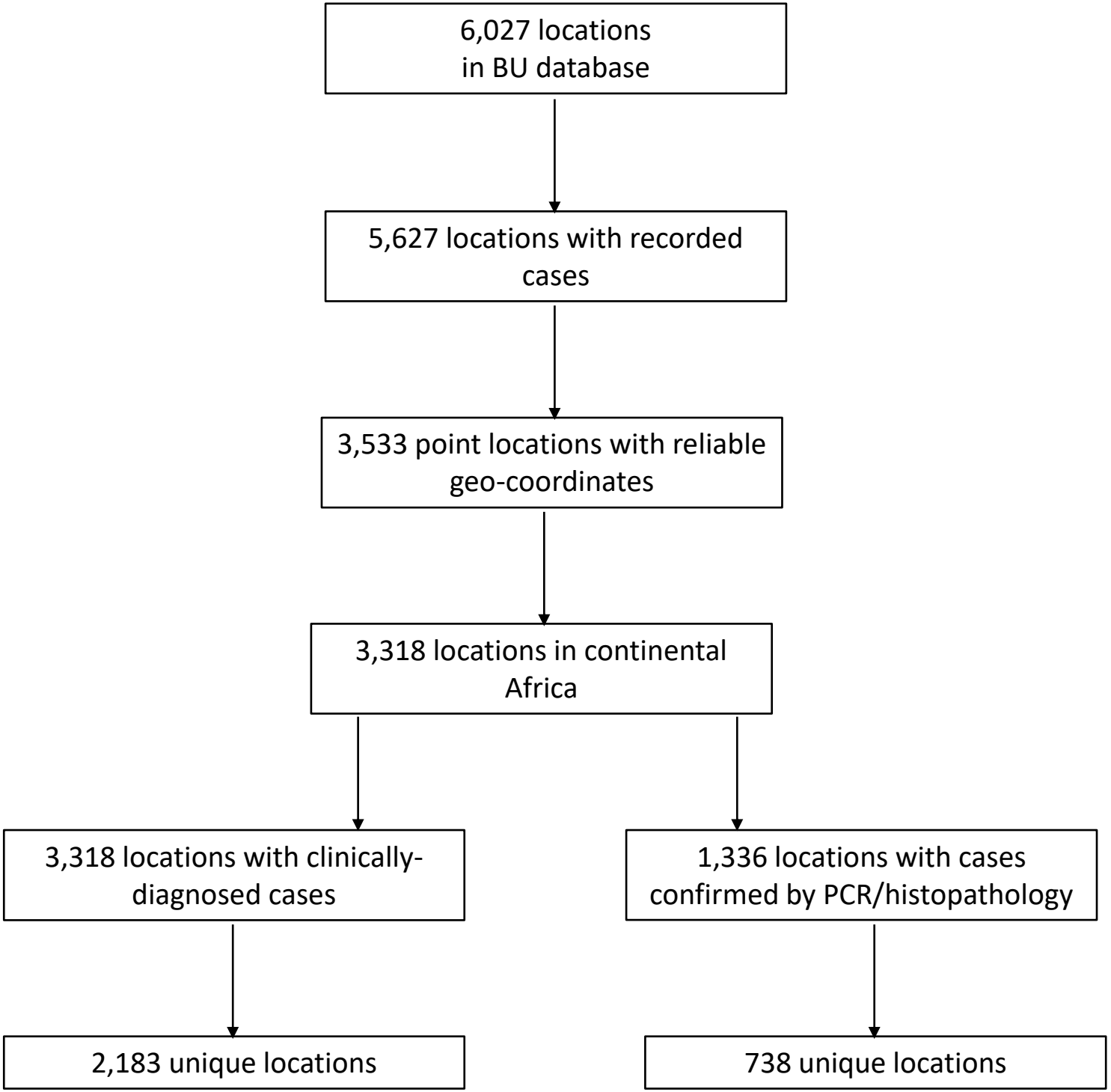
2014/08/12. doi: 10.1371/journal.pntd.0003123

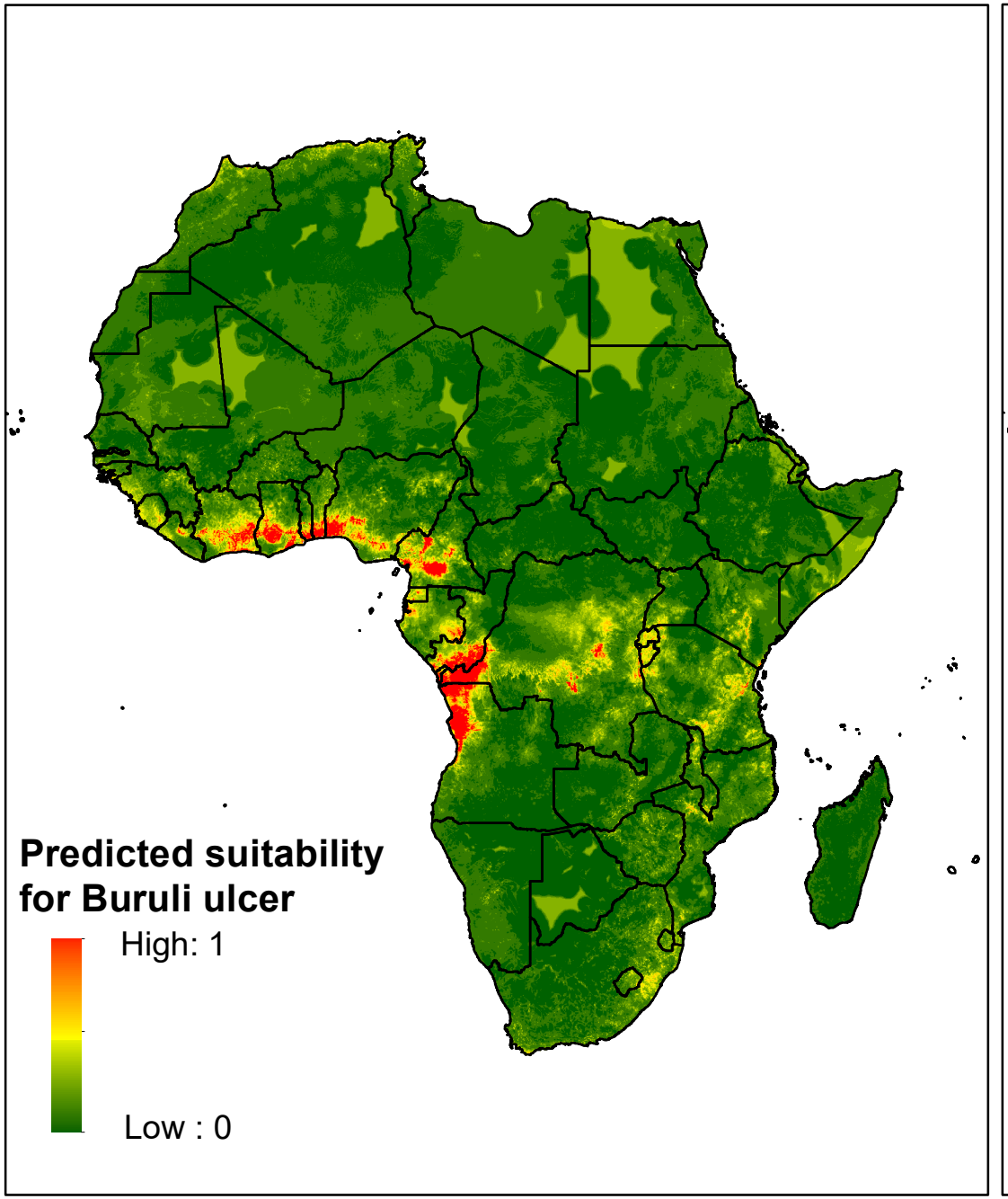
10.1016/s2214-109x(14)70223-2. PubMed PMID: 25103396.

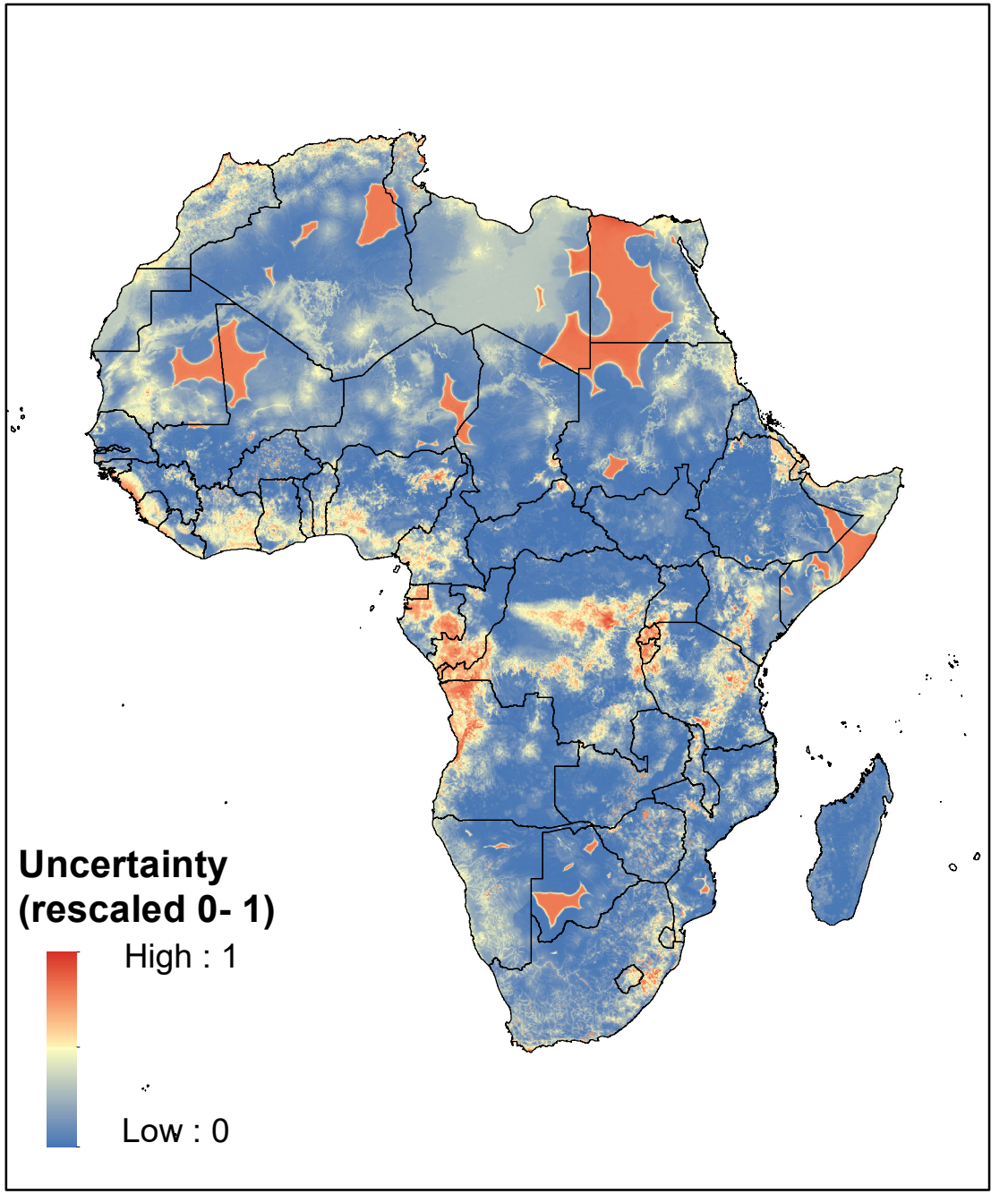
53. Clements AC, Reid HL, Kelly GC, Hay SI. Further shrinking the malaria map: how can geospatial science help to achieve malaria elimination? The Lancet Infectious Diseases. 2013;13(8):709-18. doi: 10.1016/S1473-3099(13)70140-3.

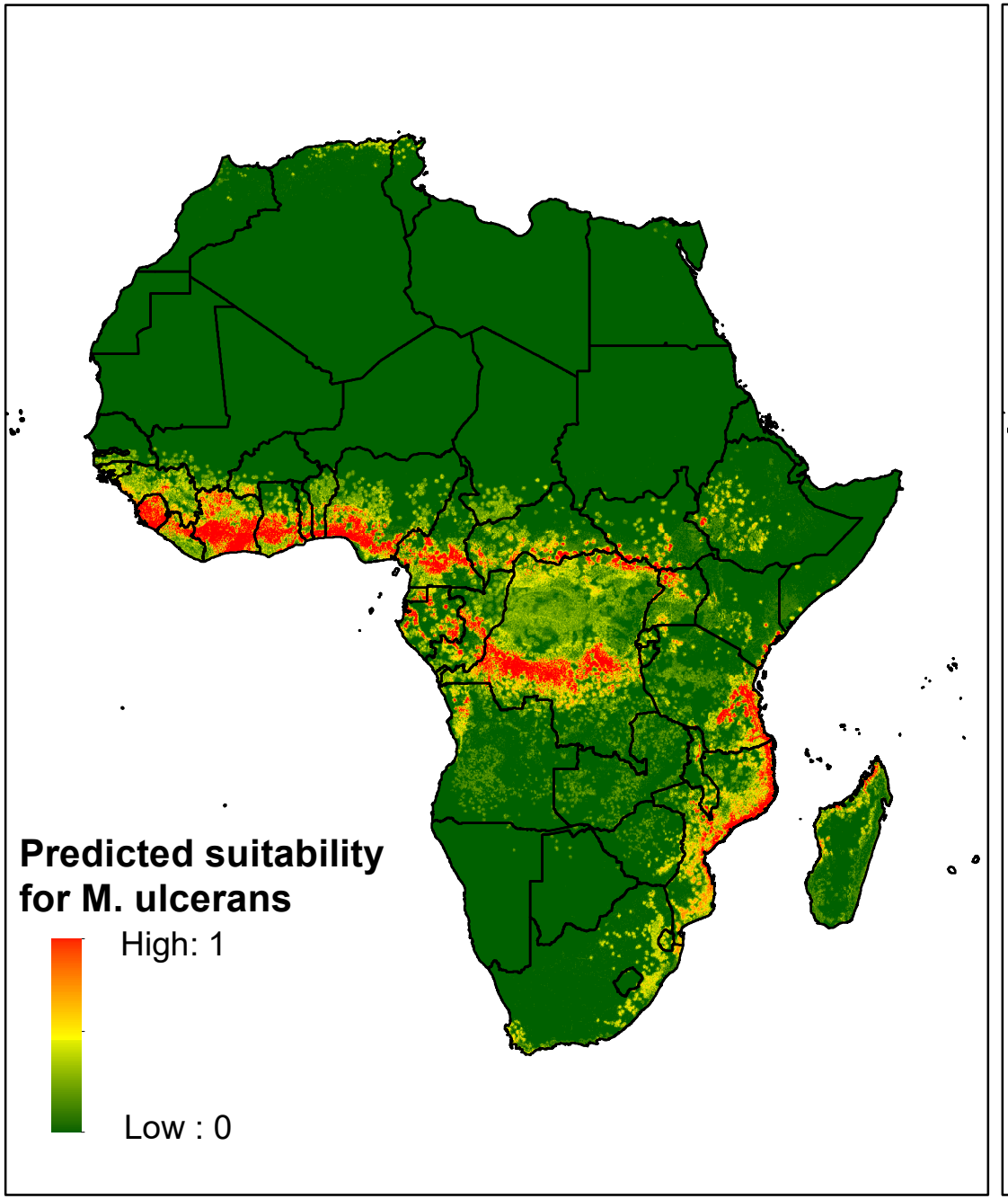
54. Deribe K, Beng AA, Cano J, Njouendo AJ, Fru-Cho J, Awah AR, et al. Mapping the geographical distribution of podoconiosis in Cameroon using parasitological, serological, and clinical evidence to exclude other causes of lymphedema. PLoS neglected tropical diseases. 2018;12(1).

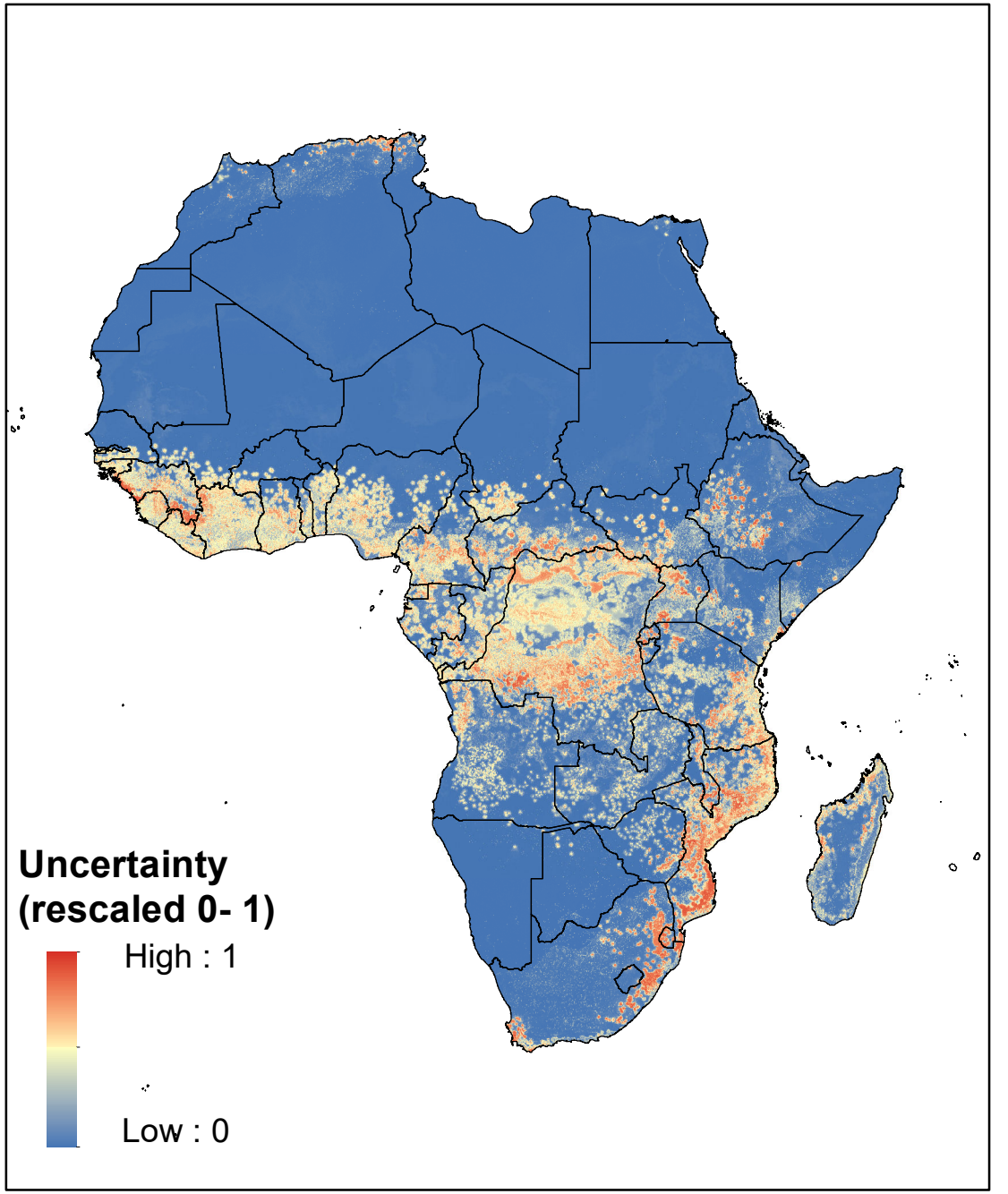
55. Wagner T, Benbow ME, Burns M, Johnson RC, Merritt RW, Qi J, et al. A Landscape-based model for predicting Mycobacterium ulcerans infection (Buruli Ulcer disease) presence in Benin, West Africa. EcoHealth. 2008;5(1):69-79. Epub 2008/07/24. doi: 10.1007/s10393-007-0148-7. PubMed PMID: 18648799.

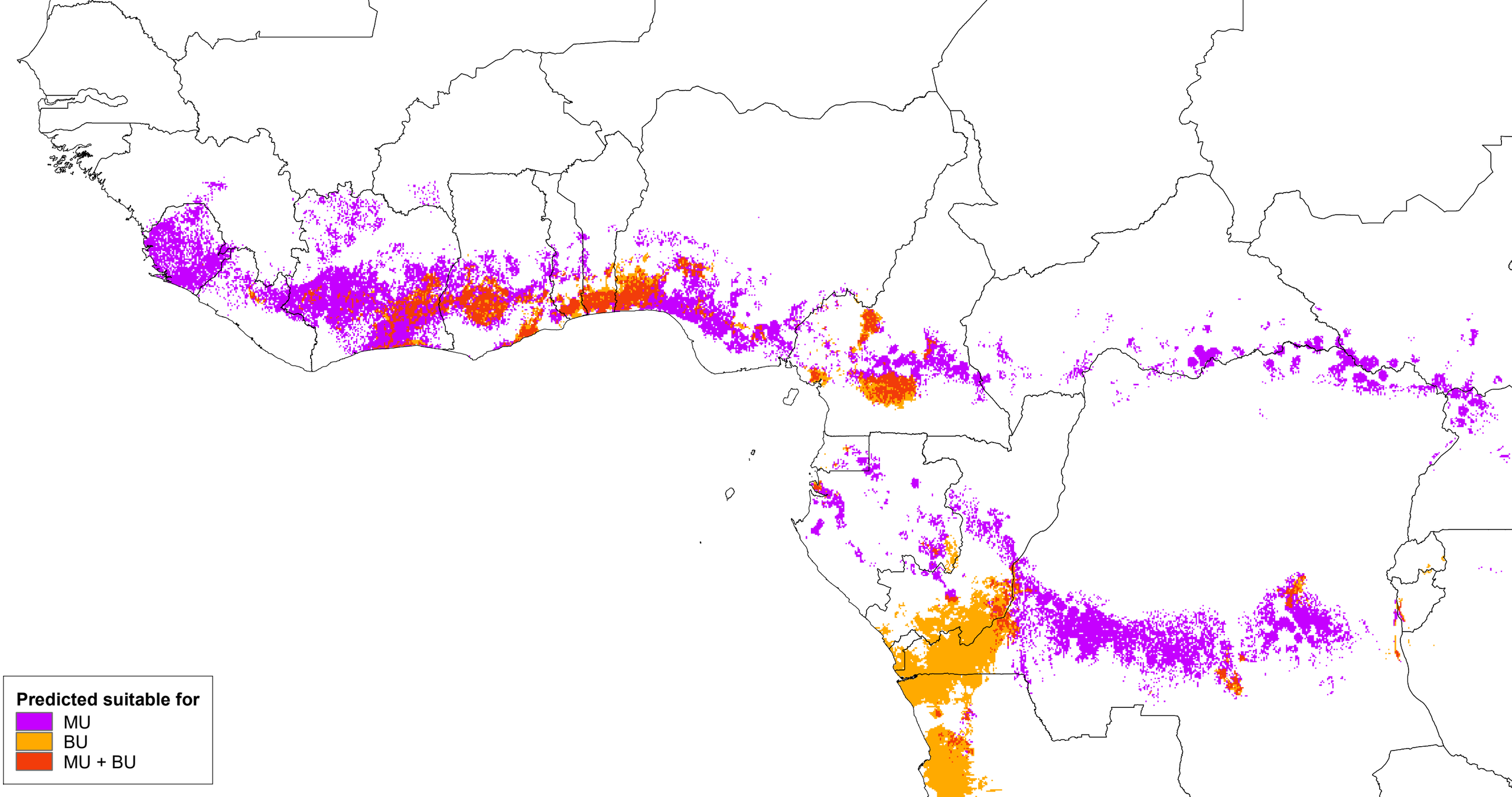

6,027 locations in BU database
5,627 locations with recorded cases
3,533 point locations with reliable geo-coordinates
3,318 locations in continental Africa
3,318 locations with clinically-diagnosed cases
1,336 locations with cases confirmed by PCR/histopathology
2,183 unique locations
738 unique locations


Predicted suitability
for Buruli ulcer
High: 1
Low : 0


Uncertainty
(rescaled 0- 1)
High : 1
Low : 0


Predicted suitability
for M. ulcerans
High: 1
Low : 0


Uncertainty
(rescaled 0- 1)
High : 1
Low : 0


Predicted suitable for
MU
BU
MU + BU