nature microbiology

SUPPLEMENTARY INFORMATION

ARTICLE NUMBER: 15008 | DOI: 10.1038/NMICROBIOL.2015.8

Predicted global distribution of *Burkholderia pseudomallei* and burden of melioidosis

Direk Limmathurotsakul[1,2,3], Nick Golding[1], David AB Dance[4,5], Jane P Messina[6], David M Pigott[1], Catherine L Moyes[1], Dionne B Rolim[7], Eric Bertherat[8], Nicholas PJ Day[2,5], Sharon J Peacock[2,9,10], Simon I Hay[1,11,12]

[1] *Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, United Kingdom.*

[2] *Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, 10400, Thailand.*

[3] *Department of Tropical Hygiene, Faculty of Tropical Medicine, Mahidol University, Bangkok, 10400, Thailand.*

[4] *Lao-Oxford-Mahosot Hospital-Wellcome Trust Research Unit, Vientiane, Lao PDR.*

[5] *Centre for Tropical Medicine and Global Health, University of Oxford, Old Road Campus, Oxford, OX3 7FZ, United Kingdom.*

[6] *Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, United Kingdom*

[7] *Universidade de Fortaleza, Fortaleza, 60811-905, Brazil.*

[8] *Department of Pandemic and Epidemic Diseases, World Health Organization, Geneva, 27, Switzerland.*

[9] *Department of Medicine, University of Cambridge, Addenbrooke's Hospital, Cambridge, CB2 0QQ, United Kingdom.*

[10] *London School of Hygiene and Tropical Medicine, London, WC1E 7HT, United Kingdom.*

[11] *Institute of Health Metrics and Evaluation, University of Washington, Seattle, WA 98121, USA.*

[12] *Fogarty International Center, National Institutes of Health, Bethesda, MD 20892-2220, USA.*

DOI: 10.1038/NMICROBIOL.2015.8

Supplementary Methods

Development of melioidosis evidence consensus database on national level. Evidence for the presence of melioidosis was obtained from three types of source: health reporting organizations, peer-reviewed articles and case reports (see Supplementary Information Figure 1, Panel b). We then used a weighted scoring system to quantify evidence consensus (see Supplementary Information Figure 3).

For health organization evidence, we used the data from two health reporting organizations (Centers for Disease Control and Prevention (CDC) and the Global Infectious Disease and Epidemiological Network (GIDEON)). The World Health Organization (WHO) does not provide data at country level. A consensus of presence or absence (++ or --) scored six or zero, respectively, while a lack of consensus (+- or -+) scored three. This gave a maximum score for this category of six.

For supporting evidence of positive occurrences, each publication and case report about melioidosis cases or presence of *B. pseudomallei* was scored independently for contemporariness and diagnostic accuracy. For contemporariness, the year of occurrence was used for scoring as follows: between 2005-2014=3, 1996-2005=2, pre 1996=1. This score was then added to a score for accuracy, whereby excellent accuracy and a score of four was characterised by a report of more than ten indigenous culture-confirmed cases at a single location, or repeated culture-positives for *B. pseudomallei* from the environment at a single location at least a year apart. High accuracy and a score of three was characterised by a report of indigenous culture-confirmed cases or presence of *B. pseudomallei* with a definite bacterial identification by either genotyping, PCR identification, latex agglutination, animal virulence test or arabinose test (for environmental *B. pseudomallei*)[3,33]. Medium accuracy and a score of two was characterised by a report of indigenous culture-confirmed case or presence of *B. pseudomallei* without a definite bacterial identification for *B. pseudomallei*. Low accuracy and a score of one was characterised by a report of melioidosis cases diagnosed by clinical presentation or serological evidence, or a report of an exported case. To provide a single score per country, the total of the highest-scoring ten publications per country was calculated, and then divided by ten. To represent the diversity of evidence, evidence was categorized into three types; human cases, animal cases and environmental detection. If a

single type of evidence was present in a country, a score of one was added. If two types were present, a score of two was added. If all types were present, a score of three was added. This resulted in a maximum available score of ten for this category.

For supporting evidence of absence, the evidence was graded using total annual healthcare expenditure (HE) per capita at average US Dollar exchange for the year 2013[34]. Higher HE has been linked to better overall public health infrastructure, which includes high-quality diagnostic resources[34]. Therefore, the lower the HE, the less certain we can be that an absence of data accurately reflects an absence of cases. All overseas territories were assumed to have the same HE as their parent nations. The following criteria were used. For countries with no evidence of positive occurrences, having HE < $100 gave a score of -1, $100 ≤ HE < $300 gave a score of -3, $300 ≤ HE < $1000 gave a score of -6, and HE ≥ $1000 gave a score of -9. The maximum score for this category was -9.

We derived an overall country evidence score by adding the scores for all three evidence categories, dividing by the maximum possible score (16 when score was higher than or equal to zero, and nine when score was less than zero) and then multiplying by 100. Evidence consensus was then categorized into nine interval categories from 100% to -100%, defined as complete (±90% to ±100%), good (±61% to ±90%), moderate (±31% to ±60%), poor (±11 to ±30%) or indeterminate (-10 to 10%).

Explanatory covariates. We assembled gridded (5 km x 5 km pixel) global data for a suite of explanatory covariates of soil characteristics, climatic conditions and other covariates. These were chosen based on factors known or hypothesised to contribute to presence of *B. pseudomallei* in the soil. Covariates included (i) soil characteristics from Harmonized World Soil Database (HWSD) (http://www.iiasa.ac.at)[21], (ii) precipitation and land surface temperature variables from the WorldClim database (www.worldclim.org), and (iii) a vegetation/moisture index from the Advanced Very High Resolution Radiometer (AVHRR) database of the National Oceanic and Atmospheric Administration (www.noaa.gov)[35,36]. HWSD and WorldClim are freely available sets of global soil and climate data at a

DOI: 10.1038/NMICROBIOL.2015.8

1 km x1 km spatial resolution, respectively. All grids were resampled to the same 5 km x 5 km grid to ensure uniformity of land/water boundaries and spatial resolution.

Predicting environmental suitability for *B. pseudomallei*. We used a boosted regression tree (BRT) approach[31] to establish a multivariable empirical relationship between the distribution of occurrence records of *B. pseudomallei* and the environmental conditions at each location, as has been previously applied to mapping dengue[8], avian influenza[9], Ebola[10] and the Leishmaniases[37] (see Supplementary Information Figure 1, Panel c). The BRT method has been shown to fit complicated response functions efficiently, while guarding against overfitting and is therefore widely used for vector and disease distribution mapping. The BRT approach combines regression trees with gradient boosting, whereby an initial regression tree is fitted and iteratively improved upon in a forward stagewise manner (boosting) by minimising the variation in the response not explained by the model at each iteration[31].

Like other niche mapping approaches, the BRT models require not only presence data but also absence data, defining areas of disease absence and potentially unsuitable environmental conditions at unsampled locations. Since data on absence of disease are rarely reported, pseudo-absence data that are representative of areas where the disease is absent are used instead. In order to account for spatial bias in reporting of occurrence records in presence-only distribution modelling, it is preferable to use pseudo-absence data which is subject to the same spatial reporting bias[38]. As melioidosis cases are most likely to be correctly diagnosed in large health facilities we used the locations of 10,567 global major cities from Aneki World Cities site[39,40] as pseudo-absence records.

To account for and quantify uncertainty in the model we used a bootstrap resampling procedure to randomly select data sets of both positive occurrences and pseudo-negative locations. To take into account spatial uncertainty in location of exposure to *B. pseudomallei* of melioidosis cases, for each of these bootstraps we randomly sampled the locations of positive melioidosis cases from an area with a

radius of 20 km from the reported locations. To incorporate national-level strength of evidence for the presence or absence of melioidosis, we also selected pseudo-absence records based on national-level evidence of melioidosis. For each bootstrap, we randomly sampled pseudo-absence locations from the candidate set of pseudo-absences (locations of major cities) with sampling probability inversely proportional to the evidence consensus score so that cities in regions with the highest level of evidence for presence had a sampling probability of 0 and those with highest evidence for absence had a probability of 1. Then we combined positive and negative records and randomly selected the bootstrap data sets using a bootstrap resampling procedure. We carried out this bootstrap resampling procedure 100 times, fitting the model to each resampled data set to create an ensemble map of all 100 sub-models. From this ensemble, we estimated the median prediction and their 95% credible intervals for all pixels and summary statistics (see Supplementary Information Figure 9).

Model validation. Model accuracy was assessed by calculating the mean cross-validated area under the curve (AUC) statistic for each submodel[32]. This procedure split each data set into ten subsets, each containing approximately equal numbers of presence and background points. The ability of a model to predict the distribution in each withheld subset when trained on the other 90% of records was assessed by AUC and the mean value taken across all ten folds. Additionally, to prevent inflation of the accuracy statistics due to spatial sorting bias, these statistics were estimated using a pairwise distance sampling procedure[32]. Consequently, the AUC statistics presented here are lower than would be returned by standard procedures but give a more realistic quantification of the model's ability to extrapolate predictions to new regions[41]. Finally, the overall mean and 95% confidence interval of AUC from all 100 sub-models were estimated.

Estimation of populations at risk. People living in each 5 km x 5 km pixel with a predicted environmental suitability for *B. pseudomallei* greater than the value of the fifth percentile of the positive occurrence records were considered at risk of acquiring melioidosis in each submodel. The population

DOI: 10.1038/NMICROBIOL.2015.8

density map was derived from Global Rural-Urban Mapping Project (GRUMP) 2010 (http://sedac.ciesin.org/data/collection/grump-v1) and levelled to match national and global population projections for 2015 from UN World Population Prospects (http://esa.un.org/wpp/). World region was categorized according to the World Bank[24]. The median population at risk was estimated from all 100 sub-models.

Estimation of incidence of melioidosis. Published literature reporting annual incidence rates of melioidosis from the literature search described above were used to make spatial prediction of incidence rates (see Supplementary Information Figure 1, Panel d). Inclusion criteria were restricted to the literature reporting annual incidence rate (per 100,000 population) of culture-confirmed melioidosis or equivalent in a defined area. The reports were abstracted, standardised and geo-positioned. In total, 16 reports of annual incidence rates of melioidosis at different locations were used (see Supplementary Information Figure 6). We then used a negative binomial model to estimate the incidence of melioidosis cases based on the *B. pseudomallei* suitability and the prevalence of diabetes and aboriginal population. This model was fitted using Bayesian inference by integrated nested Laplace approximation[42] in R[43], with uninformative priors on regression coefficients and a gamma (1, 1) prior on the overdispersion parameter of the model. This model was fitted once for each bootstrap of the suitability model, updating the suitability covariate each time. The fitted negative binomial model was then applied to the 2015 human population surface, the *B. pseudomallei* suitability, the prevalence of diabetic[25] and the prevalence of indigenous Australians[44] to provide a mapped estimate of incidence. To account for and quantify uncertainty in the model, we randomly selected 25 sets of parameter values from the model posterior for each bootstrap of the suitability model. This resulted in 2,500 random realisation maps which were used to estimate incidence of melioidosis cases nationally, regionally and globally with 95% credible intervals.

Estimation of mortality of melioidosis. Published literature reporting case fatality rates (CFRs) of melioidosis from the literature search described above were used to make spatial prediction of CFRs.

Inclusion criteria were restricted to the literature reporting CFRs (%) of culture-confirmed melioidosis in a defined area. The reports were abstracted, standardised and geo-positioned. In total, 19 reports of CFRs of 3,641 melioidosis patients at different locations were used (see Supplementary Information Figure 6). We evaluated the association of reported CFRs with national-level HE per capita, log10 transformed HE per capita, the national-level under-5 mortality rate (U5MR)[24] and log10 transformed U5MR by constructing multivariable logistic regression models and assessing their goodness of fit to the data. Reported HE and U5MR were expected to be a proxy for the capacity of medical services in the countries and associated with the CFRs of melioidosis. The optimal logistic regression model contained only log10 transformed U5MR. This model was then used to predict the number of deaths due to melioidosis by applying the model to national-level log10 transformed U5MR and multiplying the predicted CFRs by the gridded predictions of melioidosis incidence. As with the incidence model, we randomly sampled 2,500 sets of regression coefficients according to the density of the likelihood of the logistic regression model, applying each prediction to a different realisation of the incidence model.

Development of priority country list. Priority countries included countries where melioidosis is known to be endemic but underreported, and countries where melioidosis is probably endemic but is never reported. Countries where melioidosis is known to be endemic were defined as countries with a national level of evidence from “complete presence” to “poor presence” and having the lower limit of 95% credible interval of predicted incidence of human melioidosis as more than or equal to one case per year. Countries where melioidosis is probably endemic were defined as countries with a national level of evidence from “indeterminate” to “good absence” and having the lower limit of 95% credible interval of predicted incidence of human melioidosis as more than or equal to one case per year. Countries with a level of evidence equal to “complete absence” were considered not endemic for melioidosis, and the level of risk of *B. pseudomallei* establishment was evaluated in the countries within this category.

DOI: 10.1038/NMICROBIOL.2015.8

Additional References

31 Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *J Anim Ecol* **77**, 802-813 (2008).

32 Hijmans, R. J. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* **93**, 679-688 (2012).

33 Limmathurotsakul, D. et al. Systematic review and consensus guidelines for environmental sampling of *Burkholderia pseudomallei*. *PLoS NTD* **7**, e2105 (2013).

34 World Health Organization (WHO). World Health Statistics 2013. http://www.who.int/gho/publications/world_health_statistics/2013/en/

35 Hay, S. I. An overview of remote sensing and geodesy for epidemiology and public health application. *Adv Parasitol* **47**, 1-35 (2000).

36 Hay, S. I., Tatem, A. J., Graham, A. J., Goetz, S. J. & Rogers, D. J. Global environmental data for mapping infectious disease distribution. *Adv Parasitol* **62**, 37-77 (2006).

37 Pigott, D. M. *et al.* Global distribution maps of the leishmaniases. *eLife* **3** (2014).

38 Phillips, S. J. *et al.* Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol Appl* **19**, 181-197 (2009).

39 Aneki World Cities site. World City Locations Data https://github.com/bahar/WorldCityLocations (2013).

40 Aneki. World Cities. http://www.aneki.com/cities.html (2013).

41 Wenger, S. J. & Olden, J. D. in Methods in Ecology and Evolution, **3**(2), 260-267 (2012).

42 Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B (Stat Methodol).* **71,** 319–392 (2009).

43 Rue, H. *et al.* INLA: Functions which allow to perform full Bayesian analysis of latent Gaussian models using Integrated Nested Laplace Approximation. R package version 0.0. http://www.r-inla.org/ (2015).

44 Australian Bureau of Statistics. Estimates of Aboriginal and Torres Strait Islanders Australians, June 2011. http://www.abs.gov.au/ausstats/abs@.nsf/mf/3238.0.55.001 (2013).

45 R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Australia http://www.R-project.org/ (2013).

46 Hanna, J. N., Humphreys, J. L., Brookes, D. L., Messina, T. & Raulli, A. Melioidosis in north Queensland, 2000-2009. *Commun Dis Intell Q Rep* **34**, 444-447 (2010).

47 Currie, B. J., Ward, L. & Cheng, A. C. The epidemiology and clinical spectrum of melioidosis: 540 cases from the 20 year Darwin prospective study. *PLoS NTD* **4**, e900 (2010).

48 Parameswaran, U., Baird, R. W., Ward, L. M. & Currie, B. J. Melioidosis at Royal Darwin Hospital in the big 2009-2010 wet season: comparison with the preceding 20 years. *Med J Aust* **196**, 345-348 (2012).

49 Abdul Kadir, K. A., Sattayavani, M. & Pande K. Melioidosis: antibiogram of cases in Brunei Darussalam. *Brunei Int Med J* **10**, 19-24 (2014).

50 Vlieghe, E. et al. Melioidosis, phnom penh, Cambodia. *Emerg Infect Dis* **17**, 1289-1292 (2011).

51 Dai, D., Chen, Y. S., Chen, P. S. & Chen, Y. L. Case cluster shifting and contaminant source as determinants of melioidosis in Taiwan. *Trop Med Int Health* **17**, 1005-1013 (2012).

52 Chou, D. W., Chung, K. M., Chen, C. H. & Cheung, B. M. Bacteremic melioidosis in southern Taiwan: clinical characteristics and outcome. *J Formos Med Assoc* **106**, 1013-1022 (2007).

53 Quan, H. B., Li, T. Y., Gao, Y. Y. & Chen, D. X. Clinical features of diabetes mellitus cases complicated by *Burkholderia pseudomallei* septicemia. *Genet Mol Res* **13**, 3108-3116 (2014).

54 Jesudason, M. V., Anbarasu, A. & John, T. J. Septicaemic melioidosis in a tertiary care hospital in south India. *Indian J Med Res* **117**, 119-121 (2003).

55 Mohd Roslani, A. D., Tay, S. T., Puthucheary, S. D., Rukumani, D. V. & Sam, I. C. Short report: Predictors of severe disease in melioidosis patients in Kuala Lumpur, Malaysia. *Am J Trop Med Hyg* **91**, 1176-1178 (2014).

DOI: 10.1038/NMICROBIOL.2015.8

56 Deris, Z. Z., Hasan, H. & Siti Suraiya, M. N. Clinical characteristics and outcomes of bacteraemic melioidosis in a teaching hospital in a northeastern state of Malaysia: a five-year review. *J Infect Dev Ctries* **4**, 430-435 (2010).

57 Hassan, M. R. et al. Incidence, risk factors and clinical epidemiology of melioidosis: a complex socio-ecological emerging infectious disease in the Alor Setar region of Kedah, Malaysia. *BMC Infect Dis* **10**, 302 (2010).

58 How, S. H. et al. Pahang melioidosis registry. Med J Malaysia **64**, 27-30 (2009).

59 Le Hello, S. et al. Melioidosis in New Caledonia. *Emerg Infect Dis* **11**, 1607-1609 (2005).

60 Warner, J. M., Pelowa, D. B., Currie, B. J. & Hirst, R. G. Melioidosis in a rural community of Western Province, Papua New Guinea. *Trans R Soc Trop Med Hyg* **101**, 809-813 (2007).

61 Lo, T. J., Ang, L. W., James, L. & Goh, K. T. Melioidosis in a tropical city state, Singapore. *Emerg Infect Dis* **15**, 1645-1647 (2009).

62 Bhengsri, S. et al. Incidence of bacteremic melioidosis in eastern and northeastern Thailand. *Am J Trop Med Hyg* **85**, 117-120 (2011).

63 Phuong, D. M. et al. Clinical and microbiological features of melioidosis in northern Vietnam. *Trans R Soc Trop Med Hyg* **102** Suppl 1, S30-36 (2008).

Country name	Predicted incidence	Predicted mortality
East Asia & Pacific		
Indonesia *	20038 (7859 - 52812)	10224 (3944 - 27524)
Vietnam *	10430 (4097 - 27480)	4703 (1827 - 12631)
Philippines *	9116 (4819 - 18999)	4510 (2369 - 9739)
Thailand *	7572 (3396 - 17685)	2838 (1259 - 6678)
China *	7174 (3099 - 15752)	2614 (1148 - 5828)
Myanmar *	6247 (2513 - 15400)	3687 (1449 - 9299)
Cambodia *	2083 (850 - 5451)	1149 (464 - 3042)
Malaysia *	1752 (718 - 4581)	519 (209 - 1352)
Lao PDR *	420 (172 - 1072)	260 (104 - 650)
Singapore	276 (66 - 925)	47 (11 - 174)
Australia	149 (56 - 416)	35 (13 - 98)
Papua New Guinea *	129 (43 - 337)	79 (26 - 214)
Hong Kong SAR, China *	67 (23 - 288)	12 (4 - 50)
Brunei Darussalam	29 (12 - 71)	9 (4 - 21)
Timor-Leste *	10 (3 - 35)	6 (2 - 22)
Fiji *	4 (1 - 14)	2 (0 - 6)
South Asia		
India *	52506 (22335 - 124652)	31425 (13404 - 75601)
Bangladesh *	16931 (7814 - 37794)	9454 (4325 - 21621)
Sri Lanka *	1881 (705 - 4488)	619 (230 - 1501)
Nepal †	914 (317 - 2354)	502 (174 - 1353)
Pakistan *	442 (95 - 1718)	260 (58 - 1059)
Bhutan †	13 (5 - 42)	8 (3 - 24)
Sub-Saharan Africa		
Nigeria *	13481 (4839 - 38348)	8324 (2959 - 23933)
Guinea †	1372 (472 - 3810)	842 (288 - 2425)
Cote d'Ivoire *	1144 (414 - 3368)	704 (252 - 2073)
Benin †	919 (348 - 2580)	565 (213 - 1623)
Madagascar *	880 (326 - 2464)	532 (196 - 1528)
Burkina Faso *	627 (196 - 2102)	382 (122 - 1321)
Sierra Leone *	600 (212 - 1715)	371 (128 - 1097)
Mali †	580 (190 - 1912)	356 (118 - 1205)
Cameroon †	540 (169 - 1699)	327 (102 - 1077)
Liberia †	445 (148 - 1288)	276 (90 - 814)
Chad *	401 (114 - 1432)	245 (71 - 903)
Ghana †	389 (111 - 1446)	237 (67 - 891)
Niger *	368 (78 - 1371)	220 (48 - 858)
Tanzania †	307 (74 - 991)	183 (44 - 611)
Congo, Rep. †	262 (98 - 716)	162 (61 - 457)
Ethiopia †	261 (61 - 885)	155 (38 - 569)
Mozambique †	238 (65 - 775)	142 (40 - 492)
Congo, Dem. Rep. †	222 (53 - 772)	133 (34 - 485)
Malawi *	221 (69 - 636)	133 (43 - 406)
Togo †	157 (43 - 500)	96 (27 - 313)
Central African Republic †	142 (45 - 422)	86 (28 - 265)
Zambia †	112 (30 - 374)	67 (19 - 240)
Guinea-Bissau †	100 (28 - 345)	61 (17 - 220)
Kenya *	100 (27 - 327)	60 (16 - 205)
Somalia †	71 (13 - 254)	43 (8 - 161)
Sudan †	62 (8 - 247)	36 (5 - 153)
Senegal †	60 (13 - 247)	36 (8 - 160)
Gabon *	45 (16 - 127)	28 (10 - 77)
South Sudan †	39 (8 - 131)	22 (5 - 81)
Uganda †	30 (5 - 131)	17 (3 - 84)
Angola †	29 (6 - 116)	17 (4 - 74)
South Africa *	28 (6 - 103)	15 (4 - 62)
Mauritania †	28 (6 - 110)	17 (4 - 69)
Eritrea †	27 (6 - 101)	15 (3 - 62)
Gambia, The *	8 (1 - 33)	4 (1 - 22)
Zimbabwe †	7 (2 - 28)	4 (1 - 17)
Equatorial Guinea †	6 (2 - 17)	4 (1 - 11)
Mauritius *	5 (1 - 18)	2 (0 - 7)

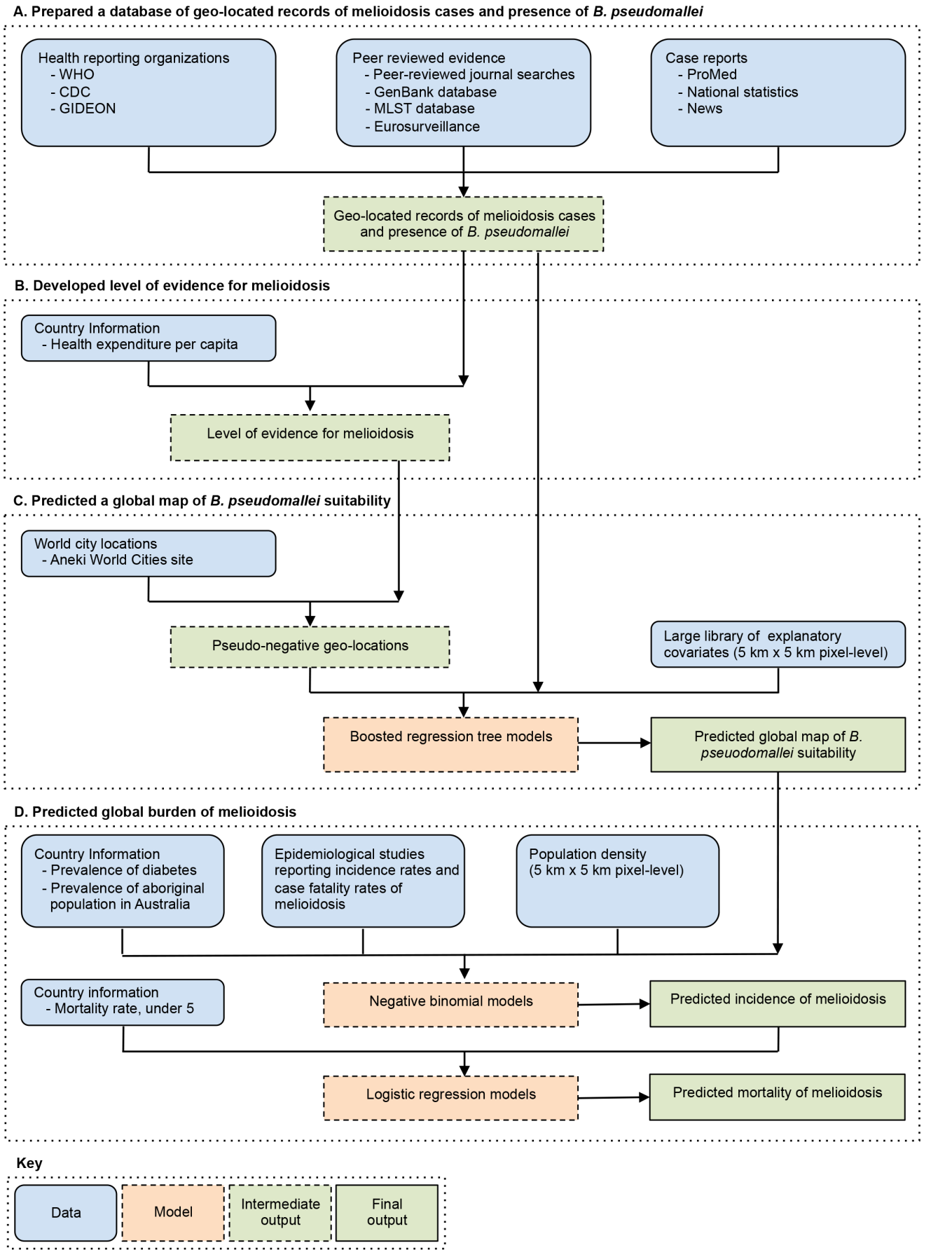
Country name	Predicted incidence	Predicted mortality
Latin America & Caribbean		
Brazil *	872 (273 - 2905)	339 (105 - 1127)
Mexico *	550 (158 - 1712)	216 (62 - 671)
Colombia *	157 (43 - 496)	64 (17 - 206)
El Salvador *	114 (37 - 295)	45 (15 - 119)
Venezuela, RB *	103 (31 - 311)	40 (12 - 121)
Honduras *	86 (22 - 264)	38 (10 - 127)
Panama *	67 (23 - 179)	28 (9 - 75)
Guatemala *	66 (20 - 197)	33 (10 - 100)
Nicaragua †	65 (18 - 196)	30 (8 - 91)
Peru *	39 (10 - 128)	16 (4 - 55)
Haiti *	24 (5 - 86)	14 (3 - 56)
Cuba †	20 (4 - 89)	5 (1 - 23)
Argentina *	18 (3 - 75)	7 (1 - 29)
Costa Rica *	16 (5 - 49)	5 (2 - 16)
Suriname †	14 (4 - 39)	6 (2 - 17)
Paraguay †	13 (2 - 59)	5 (1 - 27)
Bolivia †	13 (3 - 49)	7 (2 - 28)
Guyana *	12 (3 - 36)	6 (2 - 18)
Middle East & North Africa		
Yemen, Rep. †	99 (29 - 302)	59 (17 - 182)
Saudi Arabia *	52 (13 - 197)	16 (4 - 61)
Iraq †	21 (4 - 127)	11 (2 - 67)
Iran, Islamic Rep. *	15 (3 - 73)	6 (1 - 30)
Oman †	6 (2 - 19)	2 (1 - 6)

* Endemic but under reported

† Predicted to be endemic but never reported

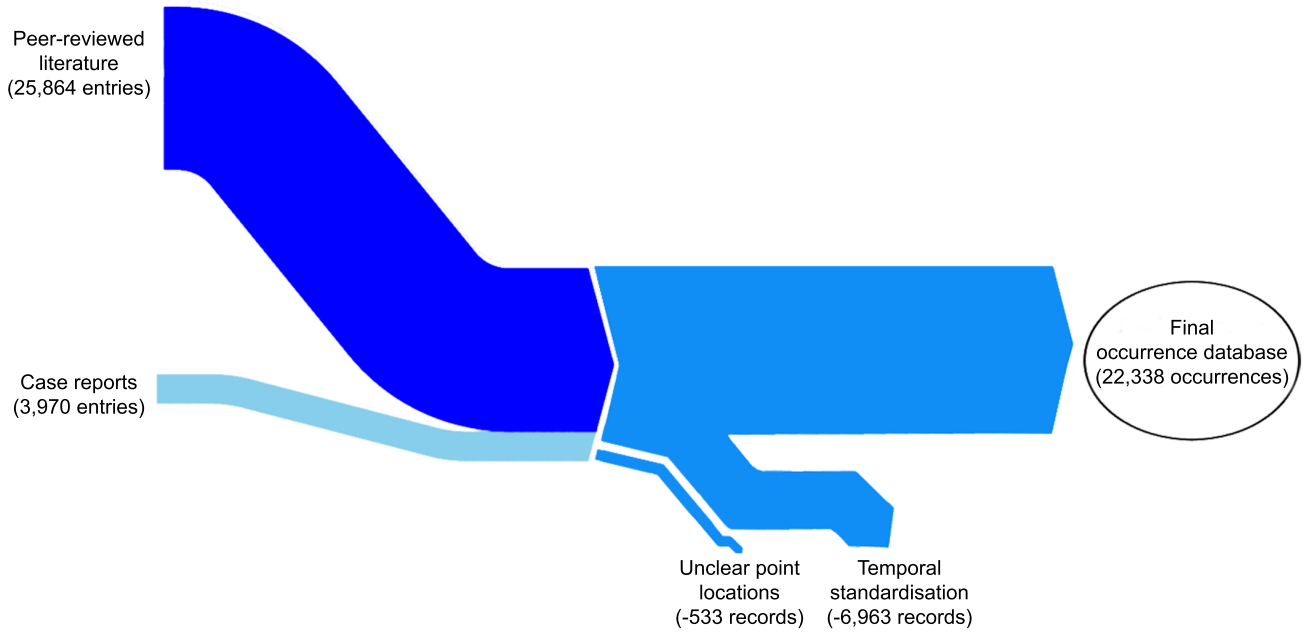
Supplementary Information Table 1 | Predicted incidence and mortality of melioidosis in 2015, by countries

DOI: 10.1038/NMICROBIOL.2015.8

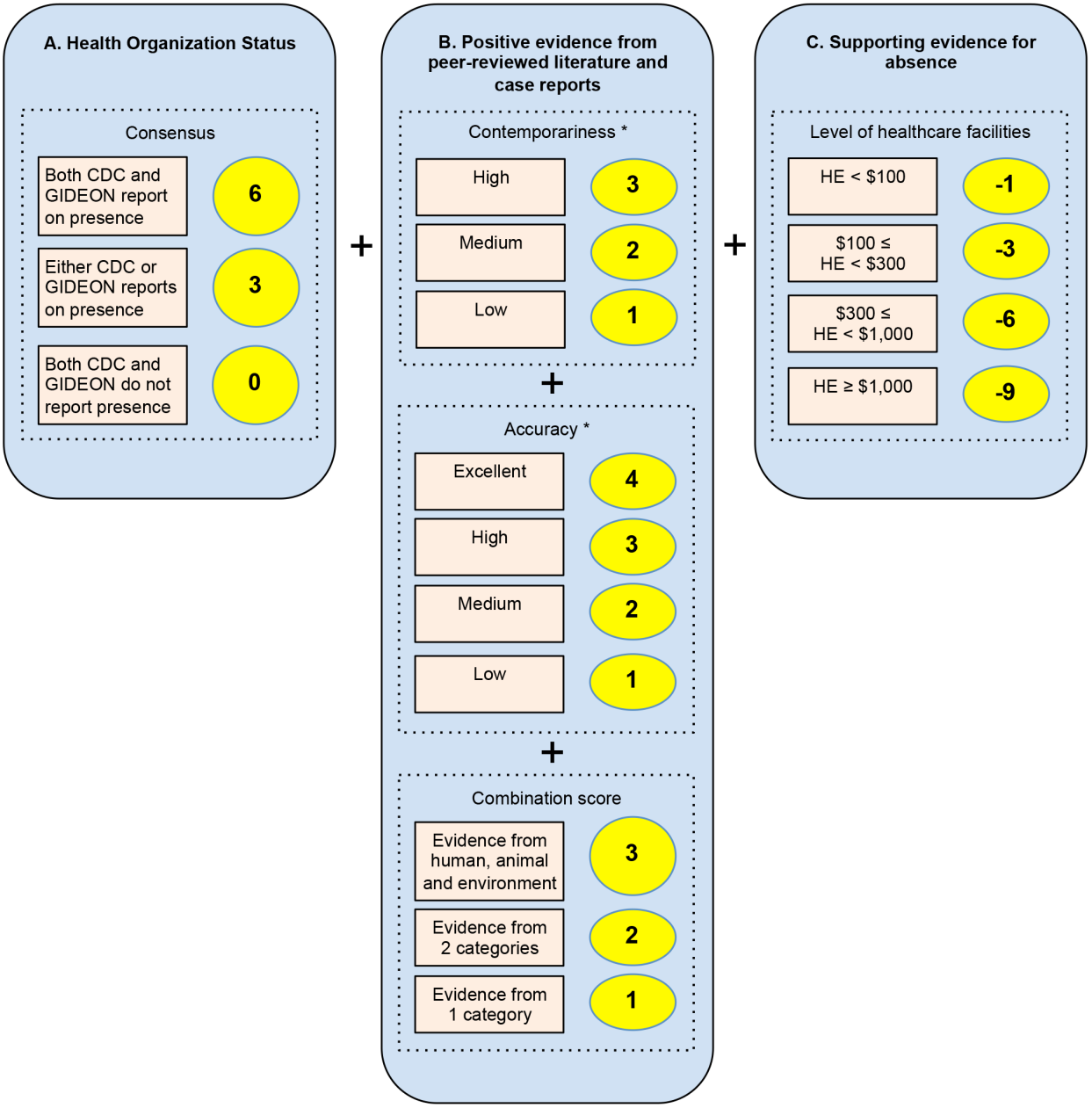


Supplementary Information Figure 1 | Schematic overview of main input data, model components, and output. Each component is summarised in the Methods

DOI: 10.1038/NMICROBIOL.2015.8



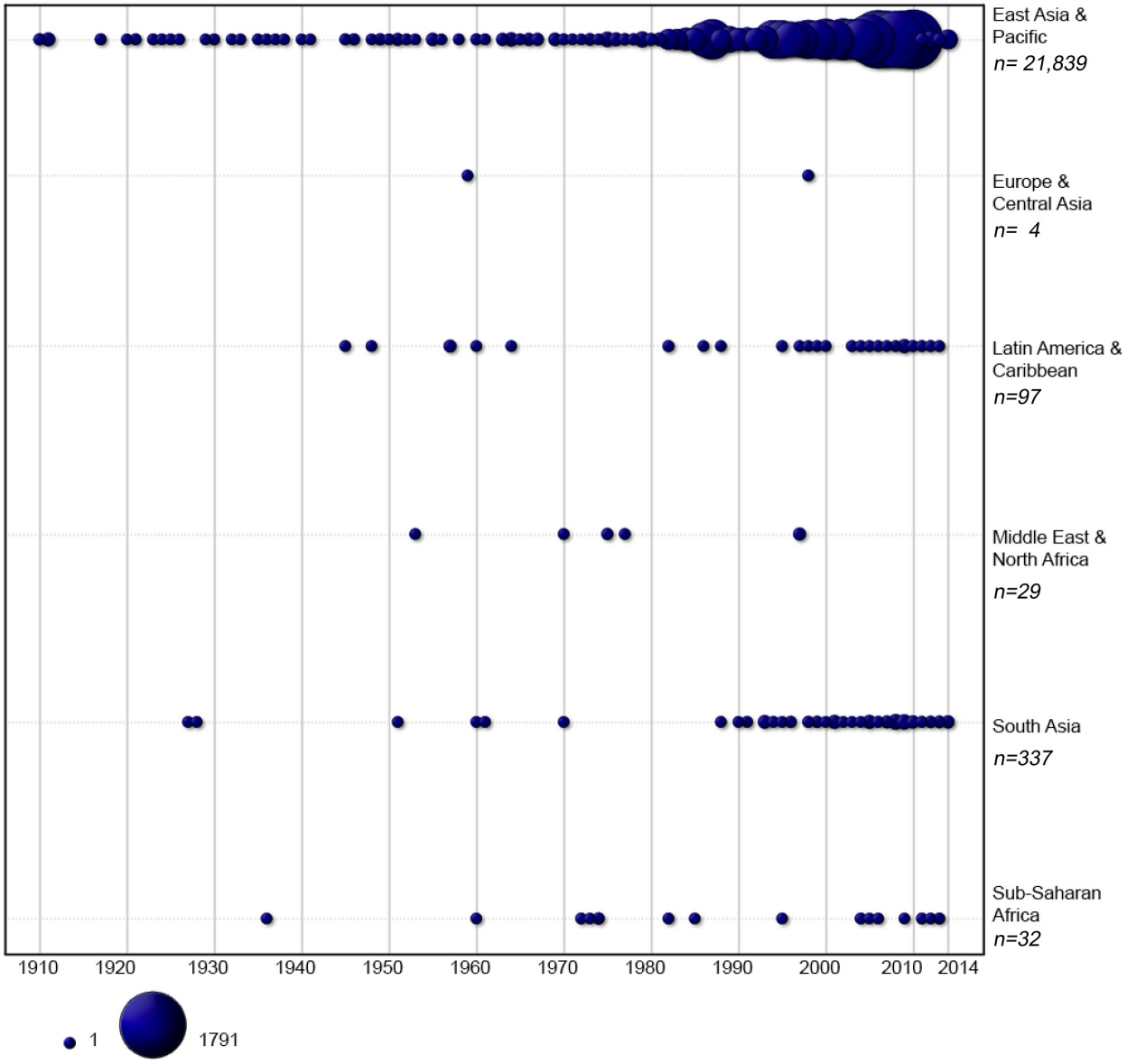
Supplementary Information Figure 2 | Occurrence data processing summary. The process began with the inputs and shows the proportion of data gained or lost through the stages of temporal standardization and quality control before reaching the final occurrence database.



Supplementary Information Figure 3 | Overview of the evidence scoring system. Dashed lines surrounding individual parameters that were assessed and totaled in the scoring system. Evidence consensus was calculated as the proportion of the maximum possible score. For contemporariness, the year of occurrence between 2005-2014 was classified as high, 1996-2005 as medium, and pre 1996 as low. For accuracy, a report of indigenous culture-confirmed cases or presence of *B. pseudomallei* with a definite bacterial identification by either genotyping, PCR identification, latex agglutination or animal virulence test was classified as high. Medium accuracy was characterised by a report of an indigenous

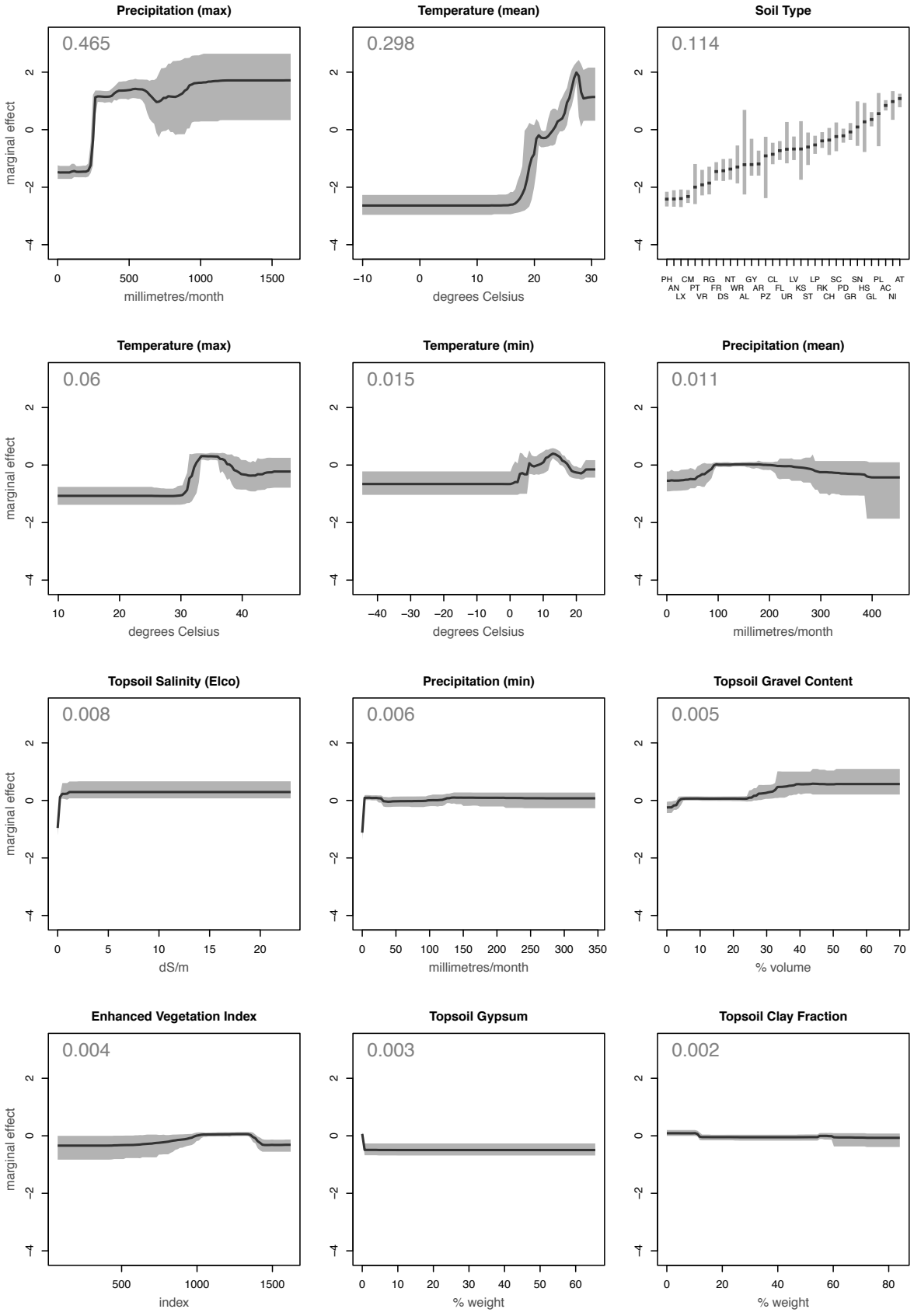
DOI: 10.1038/NMICROBIOL.2015.8

culture-confirmed case or presence of *B. pseudomallei* with an unclear bacterial identification. Low accuracy was characterised by a report of melioidosis cases diagnosed by clinical presentation or serological evidence, or a report of an exported case. HE = Total healthcare expenditure per capita at average US $ exchange rates. * Each individual piece of peer-reviewed evidence was scored for contemporariness and accuracy before taking the average of the whole set then adding to the combination score.



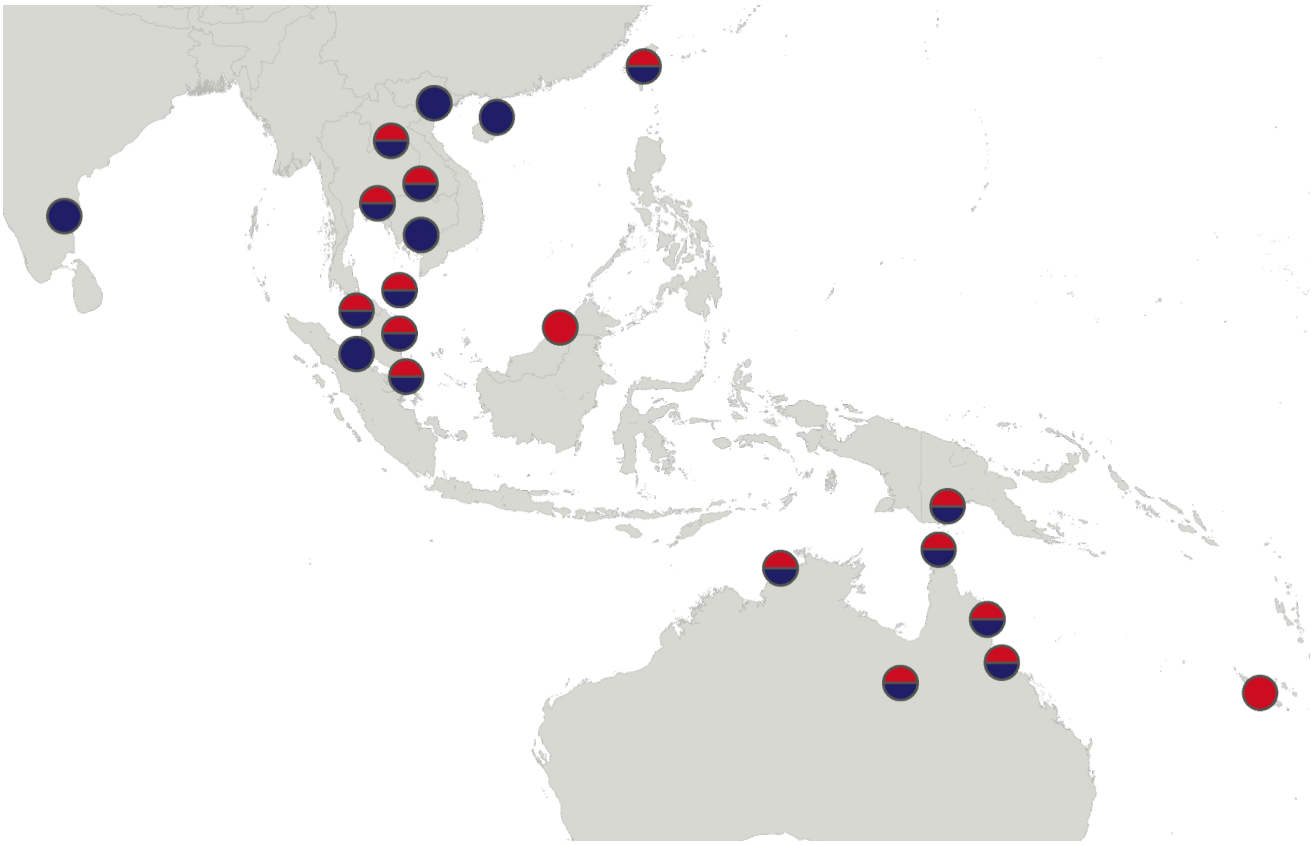
Supplementary Information Figure 4 | The numbers of occurrence locations per year separated by region.

DOI: 10.1038/NMICROBIOL.2015.8

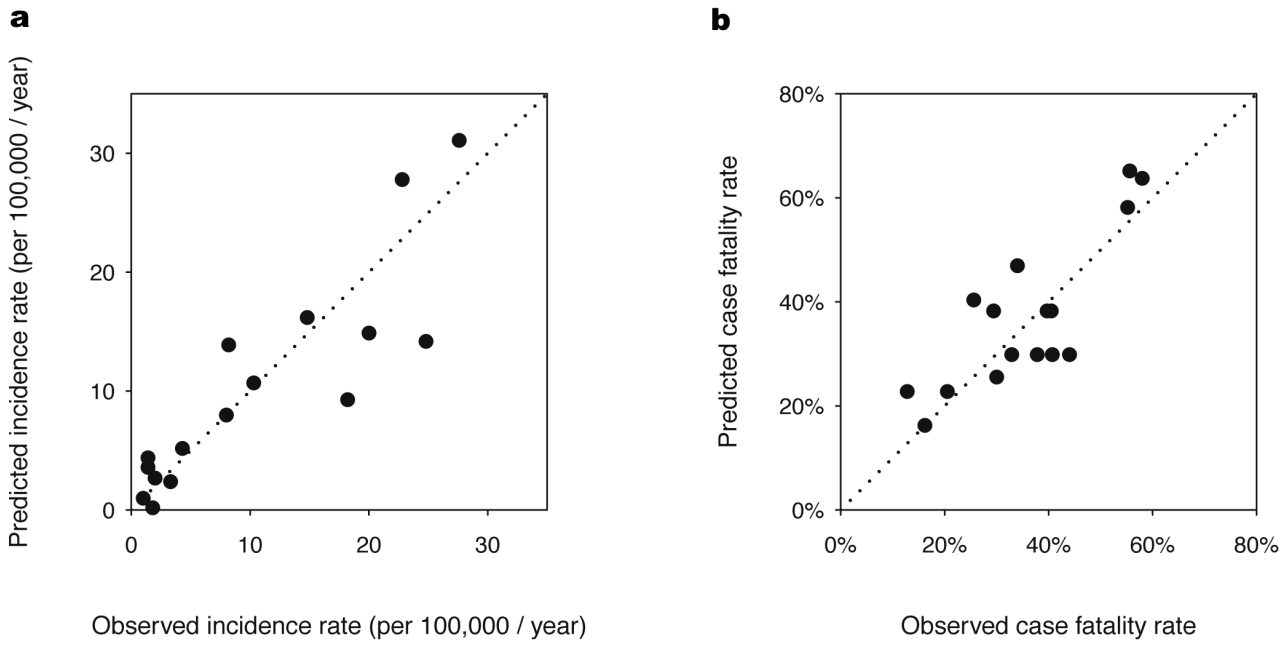


Supplementary Information Figure 5 | Partial dependence plots averaged over all 100 BRT ensembles. Black lines represent the mean partial dependence over all 100 BRT ensembles and grey envelopes the standard deviation from the mean. The y-axis is the untransformed logit response and the x-axis is the range of covariate values. The proportions at the upper left corners show the relative contributions of the covariates. Soil types include Acrisols (AC), Alisols (AL), Andosols (AN), Arenosols (AR), Anthrosols (AT), Chernozems (CH), Calcisols (CL), Cambisols (CM), Fluvisols (FR), Gleysols (GL), Greyzems (GR), Gypsisols (GY), Histosols (HS), Kastanozems (KS), Leptosols (LP), Luvisols (LV), Lixisols (LX), Nitisols (NT), Podzoluvisols (PD), Phaeozems (PH), Planosols (PL), Plinthosols (PT), Podzols (PZ), Regosols (RG), Solonchaks (SC), Solonetz (SN), Vertisols (VR), Rock Outcrop (RK), Sand Dunes (DS), Water Bodies (WR), Urban, mining, etc. (UR), Salt Flats (ST), No data (NI), Glaciers (GG) and Island (IS)[21].

DOI: 10.1038/NMICROBIOL.2015.8

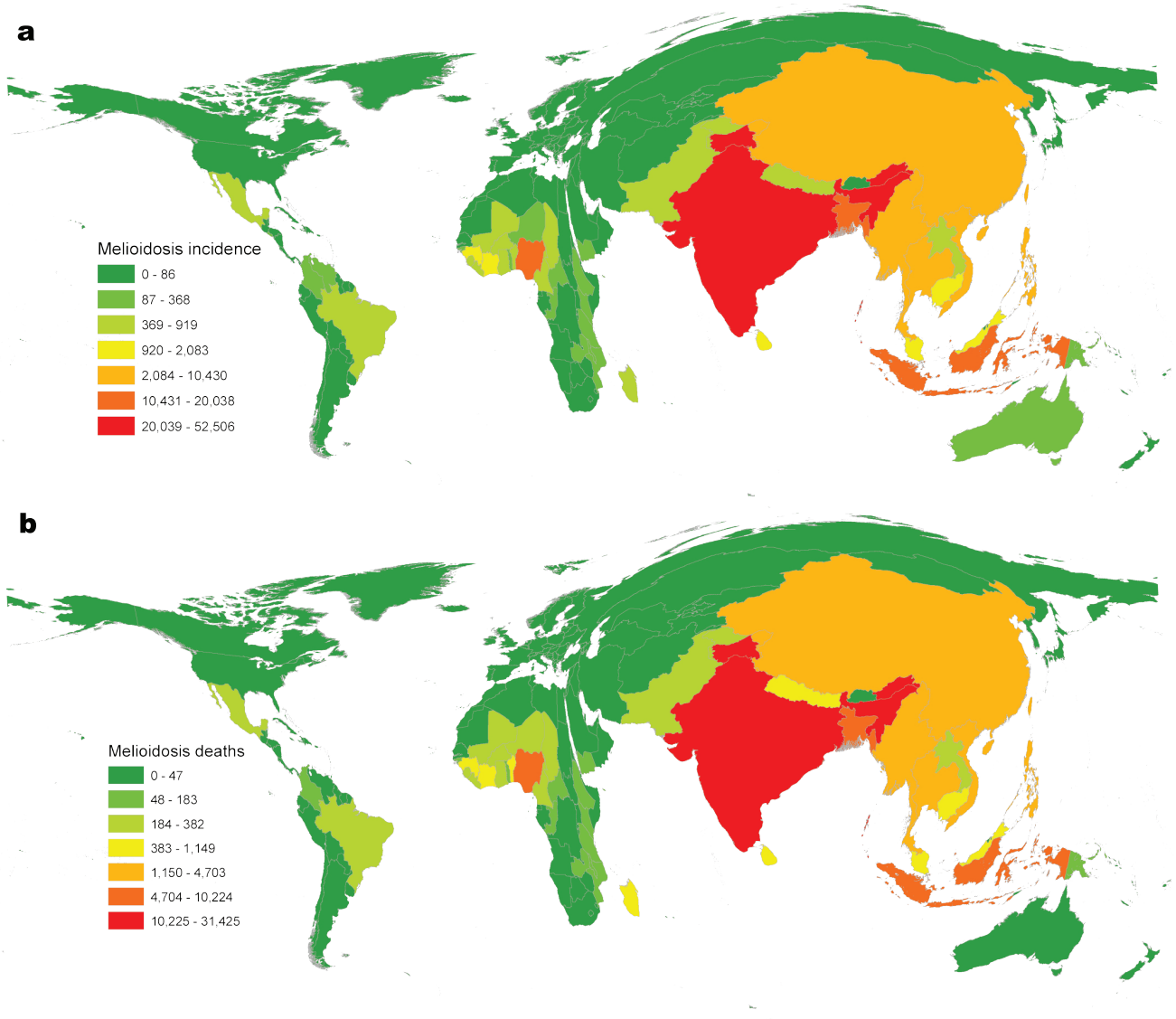


Supplementary Information Figure 6 | Locations of epidemiological studies of melioidosis. Red represents location of epidemiological studies with data for incidence rates, and blue represents location of epidemiological studies with data for case fatality rates at Cairns[46], Darwin[47,48], Mount Isa[46], Torres Strait[46], and Townsville[46] in Australia; Brunei Darussalam[49]; Phnom Penh in Cambodia[50]; Taiwan[51,52] and Hainan[53] in China; Vellore in India[54]; Kuala Lumpur[55], Kota Bharu[56], Kota Setar[57] and Kuantan[58] in Malaysia; New Caledonia[59]; Balimo in Papua New Guinea[60]; Singapore[13,61]; Nakhon Phanom[62], Sa Kaeo[62], and Ubon Ratchathani[12] in Thailand; and Hanoi in Vietnam[63].

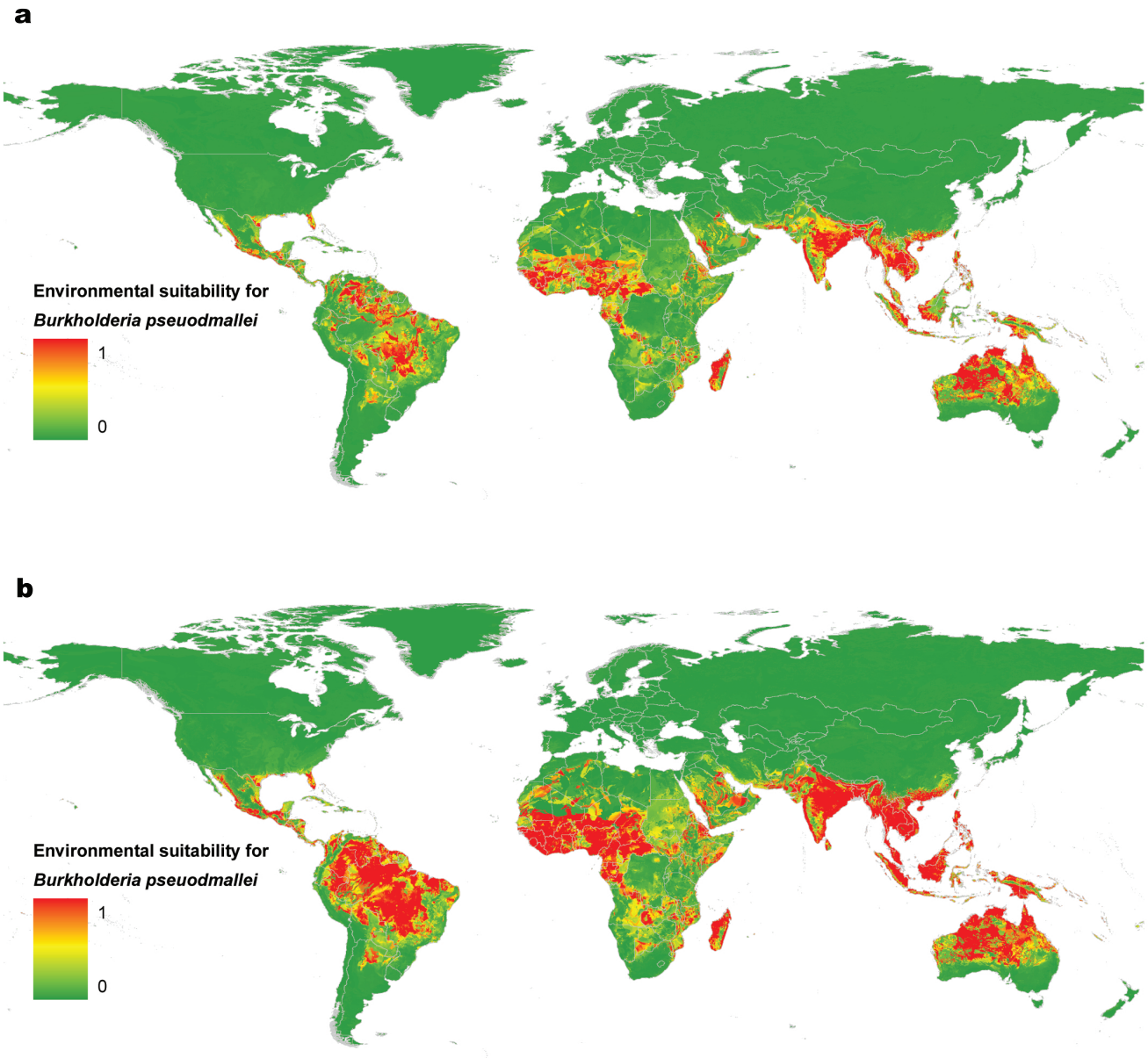


Supplementary Information Figure 7 | Comparison between observed and predicted parameters. a, comparison between observed and predicted incidence rates of melioidosis. The points represent epidemiological studies reporting incidence rates of melioidosis in different regions. The dashed line represents the linear fit line between observed and predicted incidence rates. **b,** comparison between observed and predicted case fatality rates of melioidosis. The points represent epidemiological studies reporting case fatality rates of melioidosis in different regions. The dashed line represents the linear fit line between observed and predicted parameters.

DOI: 10.1038/NMICROBIOL.2015.8



Supplementary Information Figure 8 | Global burden of melioidosis in 2015. a, shows cartograms of the incidence and **b,** shows cartograms of the mortalities as a proportion of national geographical area in 2015.



Supplementary Information Figure 9 | Uncertainty of predicted risk map of melioidosis. **a,** shows 5th and **b**, shows 95th percentiles of the *B. pseudomallei* suitability at 5 km x 5 km spatial resolution. Areas of high environmental suitability are shown in red and areas of low suitability in green.