Updating estimates of *Plasmodium knowlesi* malaria risk in response to changing land use patterns across Southeast Asia

Ruarai J Tobin[1, †], Lucinda E Harrison[2], Meg K Tully[1], Inke N D Lubis[3], Rintis Noviyanti[4], Nicholas M Anstey[5], Giri S Rajahram[6], Matthew J Grigg[5], Jennifer A Flegg[2], David J Price[1,7], and Freya M Shearer[1,8, †]

[1]Infectious Disease Dynamics Unit, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia
[2]School of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia
[3]Department of Paediatrics, Faculty of Medicine, Universitas Sumatera Utara, Medan, Indonesia
[4]Eijkman Research Center for Molecular Biology, BRIN, Jakarta, Indonesia
[5]Menzies School of Health Research and Charles Darwin University, Darwin, Australia
[6]Infectious Diseases Society Kota Kinabalu Sabah, Menzies School of Health Research Clinical Research Unit, Hospital Queen Elizabeth II, and Clinical Research Centre, Queen Elizabeth Hospital, Ministry of Health, Kota Kinabalu, Malaysia
[7]Doherty Institute for Infection and Immunity, The Royal Melbourne Hospital and The University of Melbourne, Melbourne, Australia
[8]Infectious Disease Ecology Modelling Group, Telethon Kids Institute, Perth, Australia
[†]Corresponding authors: Ruarai J Tobin (ruarai.tobin@unimelb.edu.au) and Freya M Shearer (freya.shearer@unimelb.edu.au)

Abstract

Background

Plasmodium knowlesi is a zoonotic parasite that causes malaria in humans. The pathogen has a natural host reservoir in certain macaque species and is transmitted to humans via mosquitoes of the *Anopheles* Leucosphyrus Group. The risk of human *P. knowlesi* infection varies across Southeast Asia and is dependent upon environmental factors. Understanding this geographic variation in risk is important both for enabling appropriate diagnosis and treatment of the disease and for improving the planning and evaluation of malaria elimination. However, the data available on *P. knowlesi* occurrence are biased towards regions with greater surveillance and sampling effort. Predicting the spatial variation in risk of *P. knowlesi* malaria requires methods that can both incorporate environmental risk factors and account for spatial bias in detection.

Methods & Results

We extend and apply an environmental niche modelling framework as implemented by a previous mapping study of *P. knowlesi* transmission risk which included data up to 2015. We reviewed the literature from October 2015 through to March 2020 and identified 264 new records of *P. knowlesi*, with a total of 524 occurrences included in the current study following consolidation with the 2015 study. The modelling framework used in the 2015 study was extended, with changes including the addition of new covariates to capture the effect of deforestation and urbanisation on *P. knowlesi* transmission.

Discussion

Our map of *P. knowlesi* relative transmission suitability estimates that the risk posed by the pathogen is highest in Malaysia and Indonesia, with localised areas of high risk also predicted in the Greater Mekong Subregion, The Philippines and Northeast India. These results highlight areas of priority for *P. knowlesi* surveillance and prospective sampling to address the challenge the disease poses to malaria elimination planning.

Author Summary

Plasmodium knowlesi is a parasite that can cause malaria when it infects humans. Although most people do not experience severe illness from *Plasmodium knowlesi* infection, a small number will develop serious or even fatal disease. The parasite is found naturally in some monkeys throughout Southeast Asia, and spreads from these monkeys to humans through mosquitoes. Previous research predicted where the risk of being infected is highest according to what we know about the environment across Southeast Asia, such as if there are forests in an area or if the altitude is high. In this work, we extend this previous research with more up-to-date data on environmental conditions and infections to predict the risk of being infected with *Plasmodium knowlesi*. We show that the risk *Plasmodium knowlesi* poses to humans is high across much of Southeast Asia, and that the disease will continue to challenge national goals to eliminate malaria.

Introduction

Plasmodium knowlesi is a zoonotic pathogen of growing public health concern in Southeast Asia. The pathogen has a reservoir in *Macaca fascicularis* and the closely related *Macaca nemestrina* and *Macaca leonina* macaques, and is transmitted between macaques and from macaques to humans via mosquito vectors of the *Anopheles* Leucosphyrus Group [1, 2]. Although demonstrated experimentally [3], evidence of direct human-to-human transmission of *P. knowlesi* occurring in nature is limited [4, 1, 5, 6, 7]. Infection by *P. knowlesi* most often causes mild to moderate illness in humans [8]. However, a range of outcomes are possible, with both asymptomatic infection [9, 10, 11] and severe disease being reported. Studies of patients presenting to health care facilities in Malaysia reported severe disease in around 6-9% of patients [12, 13].

For humans, the likelihood of contracting a *P. knowlesi* infection has been found to be dependent upon a range of risk factors, with case-control and seroprevalence studies demonstrating associations between environmental variables and the occurrence of infection. A seroprevalence survey performed in Malaysia and the Philippines found that prior infection with *P. knowlesi* was associated with the proximity of forested areas to an individual's home and the clearing of forest near their home [14]. A similar study performed in northern Sabah, Malaysia, found associations between prior infection and an individual reporting that they have had activity in forested areas or that they have had contact with macaques [15]. A population-based case-control study performed within Sabah, Malaysia, found an association between current *P. knowlesi* infection and an individual reporting either that they had recently cleared vegetation or that their home was in proximity to long grass [11].

The spatial epidemiology of *P. knowlesi* malaria has historically been poorly understood. This is partially due to widespread misdiagnosis. By clinical presentation, the symptoms of *P. knowlesi* infection can be easily misattributed to other major human species of malaria such as *P. vivax* or *P. falciparum* [16]. Under microscopic examination, the parasite appears almost identical to *P. malariae* [17] and the early ring stages of *P. falciparum* [18]. One review of historical microscopy diagnoses demonstrated that across 375 studies, 57% of *P. knowlesi* infections were misdiagnosed [17]. In addition to misdiagnosis, the understood spatial distribution of *P. knowlesi* malaria has been biased by differences in surveillance effort. Within peer-reviewed literature, reported *P. knowlesi* infections are most common in Malaysia, which is likely reflective of both high burden and a substantial surveillance effort in the country [19]. Indigenous cases of *P. knowlesi* malaria have also been detected in Brunei, Cambodia, Indonesia, Laos, Myanmar, the Philippines, Thailand, and Vietnam, but these have historically been the result of small-scale prospective sampling efforts and individual case reports. One study reported identifying *P. knowlesi* in India within the Andaman and Nicobar Islands [20].

The elimination of malaria in at least 20 countries by 2025 is listed as a key milestone of the World Health Organisation's *2016–2030 Global technical strategy for malaria* [21]. *P. knowlesi* presents a challenge to these efforts, since interventions that are effective against the human malaria species such as indoor residual spraying will be less effective against *P. knowlesi* due to the pathogen's persistence in wildlife reservoirs. Furthermore, cross-reactivity of antibodies between *P. knowlesi* and the closely genetically related *P. vivax* may provide protection against *P. knowlesi* infection [22], implying that the elimination of *P. vivax* in a region could lead to reduced immunity and subsequently an increase in the number of *P. knowlesi* infections [19].

The incidence of *P. knowlesi* in humans appears to be increasing within Southeast Asia; in Malaysia, the number of recorded human *P. knowlesi* infections doubled over the period from 2015 to 2018 [23]. A similar trend is visible in the rising number of case reports within Indonesia [24]. Though these trends may simply reflect improvements in surveillance [23], it has been suggested that deforestation in the region may be leading to a real increase in the number of human *P. knowlesi* infections [25, 24, 26, 27]. A primary driver of deforestation in the region is the development of oil palm or timber plantations, which produce an environment that is believed to be of enhanced risk for *P. knowlesi* infection, with plantation labourers being required to live and work in proximity to recently disturbed forests that may contain *P. knowlesi* reservoirs and vectors.

Sustained transmission of vector-borne zoonoses can only occur at the nidus where pathogen, host and vector are present in sufficient abundance [28]. For each of these, certain constraints limit their distribution, for example: a pathogen may be unable to survive at certain temperatures; a host may be displaced by human activity; and a vector may be unable to reproduce without access to standing water. The field of geospatial information systems (GIS) provides a large amount of data on such environmental and anthropological factors [29]. Environmental niche modelling utilises this geospatial data to identify relationships between the presence of a pathogen, host or vector and the environments in which they have been observed, allowing for prediction of the suitability for transmission of a vector-borne zoonosis such as *P. knowlesi* across a geographic area of interest [30].

In 2015, Shearer and colleagues applied a niche modelling approach to produce the first predictive map of *P. knowlesi* malaria risk across Southeast Asia [31]. This map provided an initial evidence base for identifying areas where disease surveillance and epidemiological investigations would be most informative to improve understanding of *P. knowlesi* malaria risk. Since the publication of the 2015 occurrence database and risk map, the volume of *P. knowlesi* data has increased across Southeast Asia, with this including detections of the pathogen in new locations. As new data accrues, it is important to update risk predictions to ensure that the most up-to-date evidence is available to public health researchers, practitioners, and policymakers. Furthermore, since 2015, studies providing evidence of the importance of deforestation in the risk of *P. knowlesi* malaria have been published, and novel datasets characterising spatial and temporal variation in land use patterns have become available.

In this study, we present updates to the *P. knowlesi* infection database and risk map produced in 2015 [31]. We perform a comprehensive review of the literature from October 2015 through to March 2020 to produce a consolidated database of *P. knowlesi* infection occurrences across Southeast Asia. By combining this occurrence dataset with data on a range of environmental covariates using a niche modelling framework, we produce updated predictions of relative suitability for *P. knowlesi* transmission to humans at fine-scale across Southeast Asia. We compare the outputs of our model to those from the 2015 model.

Methods

Infection Data

The infection occurrence database is a listing of reported locations of *P. knowlesi* infections in either humans, macaques or mosquitoes. The infection occurrences used in the 2015 analysis were extracted from literature published up to October 2015. In order to identify new occurrences, we searched the 'Web of Science' database on March 2nd 2020, using the keywords "knowlesi" *or* "monkey malaria" and filtered for results published after October 2015 (Figure 1A). Following the exclusion of laboratory studies, we extracted infection occurrence records from publications which utilised validated *P. knowlesi*-specific diagnostics (i.e. semi-nested PCR or a combination of microscopy and molecular techniques, as in the 2015 review [31]). The data collection protocol used was the same as in the 2015 analysis and further detail can be found therein [31]. We combined the collected infection occurrences produced by the current study ($n = 264$, Figure 1) with those identified in the 2015 analysis ($n = 260$).

Each location in the infection occurrence database could either take the form of a point or a polygon record. We created point records where the likely exposure site was reported with enough precision that it could be assigned to a 5×5 km grid cell. Where this level of precision was not available, we created *polygonal* records, assigning the likely exposure site to a region bounded by a polygon (Figure 1B). We created these polygons as either administrative level 1 (the first subdivision below national, e.g. state or province) or administrative level 2 (the second subdivision below national, e.g. district or regency) then disaggregated these polygons onto to the 5×5 km grid for model fitting, prediction and evaluation.

Prior to model fitting and evaluation, we excluded nine records which spanned an area greater than 1,000 grid cells (approximately 25,000 km^2). These records were unlikely to affect results given that each had substantial overlap with other more precise spatial records.

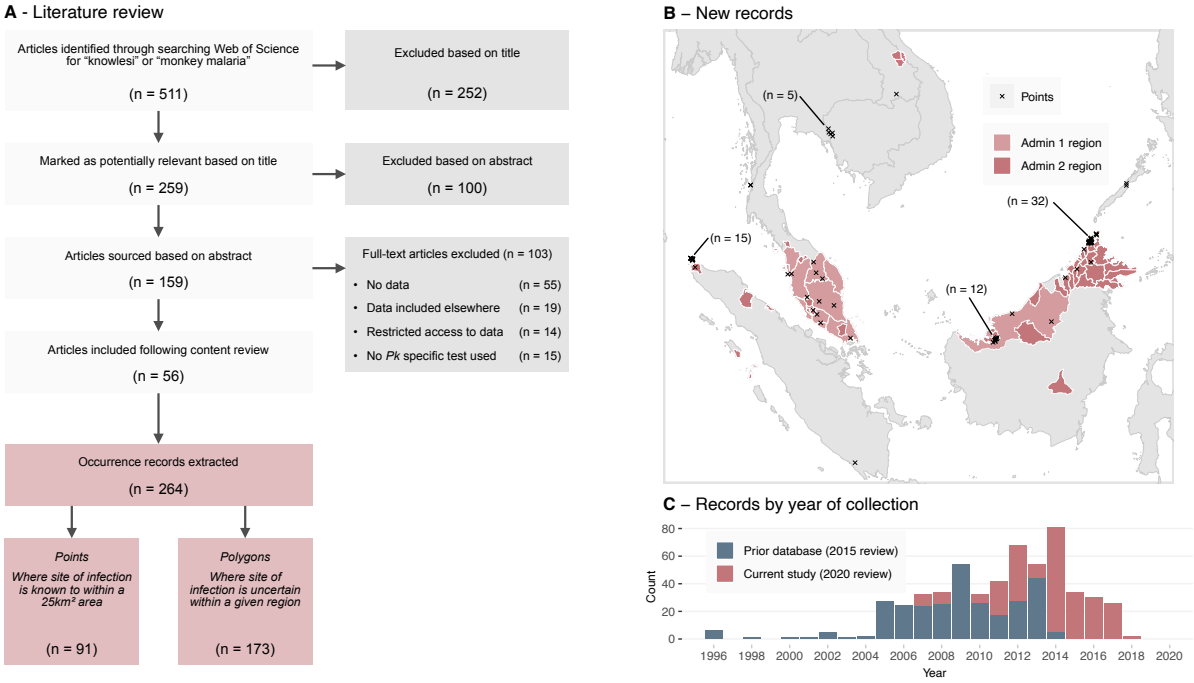

Figure 1: **A**: Study and sample selection process for the 2020 infection occurrence database. Records were produced via a literature review which was performed on March 2nd 2020, filtering for publications released after October 2015. **B**: Newly extracted point and polygon occurrence records across Southeast Asia by spatial type. Admin 1 regions are the first subdivision below national, e.g. state or province. Admin 2 regions are the second subdivision below national, e.g. district or regency. **C**: The number of occurrence samples in each occurrence database by the year the sample was collected.

Covariate Data

The infection risk model incorporated 20 environmental covariates (Table 1), each a 5 × 5 km gridded raster covering Southeast Asia. Of these 20 covariates, we treated 14 as time-varying with an annual resolution, allowing the model to associate each infection occurrence records with covariate values corresponding to the year the infection was recorded, capturing the variation of risk factors over time. Data for these annually-varying covariates were available for each year from 2001 to 2019, extending upon the coverage of the 2015 model (which covered 2001 to 2015). We assigned five samples which were collected before the year 2001 covariate values for the year 2001.

While tasseled-cap values (transformed Landsat imagery which can help differentiate areas of vegetation and urbanisation) and human population density were included as synoptic (static) variables in the 2015 model, in this work we incorporated them as temporally-varying covariates. The 2015 model incorporated an urban accessibility metric which defined the travel time to the nearest city of 50,000 people or more by land- or water-based travel in the year 2000 [32]. Here, we instead used the healthcare accessibility surface — a modelled measure of travel time to the nearest healthcare facility produced by the Malaria Atlas Project which used data up to mid-2019 [33] — as a measure of urban accessibility.

We replaced the intact and disturbed forest coverage layers used in the 2015 model with covariates that better captured the temporal and spatial dynamics of forest change in Southeast Asia. The forest coverage data sets used in the 2015 model were derived from the Intact Forest Landscapes project, which utilised a strict, manually assessed criteria for defining intact versus disturbed forest [34, 31, 35]. However, the temporal resolution of this dataset is low, with data only available for four distinct years (2000, 2013, 2016 and 2020). We chose instead to utilise data provided through the Global Forest Change project, which provides annual data on tree coverage over the last 20 years on forest presence at the resolution of 1 arc-second (roughly 30 m) [36].

We aggregated the Global Forest Change dataset up to the 5 × 5 km grid over the Southeast Asia study region through the calculation of both a tree coverage and a tree loss metric. We defined tree loss to be the proportion of forest area lost within each 5 × 5 km cell for each study year. Similarly, we defined tree coverage as the proportion of land where forest coverage was present at the beginning of the Global Forest Change data period and where no subsequent loss was recorded up until each study year. As the Global Forest Change project has not calculated forest gain past the year 2012, we were not able to include any possible increase in forest coverage.

Name	Description	Temporal resolution
Host species distribution		
Macaca fascicularis suitability	Modelled suitability of inhabitation by macaques of species *M. fascicularis* [35].	Synoptic
Macaca nemestrina suitability	Modelled suitability of inhabitation by macaques of species *M. nemestrina* [35].	Synoptic
Anopheles Leucosphyrus Group suitability	Modelled suitability of inhabitation by mosquitoes of the *Anopheles* Leucosphyrus Group [35].	Synoptic
Environmental		
SRTM elevation	Mean elevation [37]	Synoptic
Tasseled cap wetness s.d.	Tasseled-cap transformed MODIS data [38, 39]. **Now treated as temporally-varying**.	Annual
Tasseled cap wetness mean	" " "	Annual
Tasseled cap brightness s.d.	" " "	Annual
Plasmodium falciparum temperature suitability	Modelled temperature suitability index for *P. falciparum* transmission used as proxy for suitability of *P. knowlesi* [40].	Synoptic
Forest loss	Proportion of land where forest coverage has been lost in a given year [36]. **Replaced the disturbed forest dataset.**	Annual
Forest coverage	Proportion of land with forest coverage present in a given year [36]. **Replaced the intact forest dataset.**	Annual
Sociodemographic		
Healthcare accessibility	Modelled duration travel time to the nearest healthcare facility [33]. **Replaced the urban accessibility dataset.**	Synoptic
WorldPop human population	Mean human population density [41, 42]. **Now treated as temporally-varying**.	Annual
MODIS/IGBP landcover		
Open shrublands	Proportion of land with given land classification [43].	Annual
Woody savannas	" " "	Annual
Savannas	" " "	Annual
Grasslands	" " "	Annual
Permanent wetlands	" " "	Annual
Croplands	" " "	Annual
Cropland/natural vegetation mosaic	" " "	Annual
Urban and built up	" " "	Annual

Table 1: The set of raster covariate datasets used in model fitting and prediction. Differences in raster datasets between this work and those used in the 2015 *P. knowlesi* risk model appear in bold. STRM: Shuttle Radar Topography Mission, MODIS: Moderate Resolution Imaging Spectroradiometer, IGBP: International Geosphere-Biosphere Programme.

Model Fitting

We utilised a bootstrapped boosted regression tree modelling framework to characterise relationships between a region's environment and the occurrence of *P. knowlesi* transmission. Regression trees produce an approximation of some latent function (e.g. the probability of a *P. knowlesi* infection occurring) by recursively splitting across potential predictor variables (e.g. environmental covariates). The points at which these splits occur and the value assigned across each split region are selected such that the error between the regression tree and the observations is minimised [44]. Boosted regression trees extend upon the regression tree framework by producing a large number of trees and combining them in an ensemble (a process known as boosting) such that they better approximate the latent function [45]. Boosted regression trees are able to fit complex nonlinear responses including high-dimensional interactions between explanatory variables due to their hierarchical tree structure and have been shown to exhibit high predictive accuracy [46]. Finally, bootstrapping of the boosted regression tree process can be performed, allowing for uncertainty in the output to be estimated [47].

When applied to presence-absence data (such as from a systematic survey), niche models generally use a binomial likelihood to represent the probability of a species being present at a given location. Where most of the data available for modelling are presence-only, as is the case for *P. knowlesi* malaria, it is common practice in niche modelling to supplement occurrence records with "background" points to represent areas where the species or disease has not been reported [46]. A variety of approaches have been employed to select background points, including sampling to ensure that their spatial distribution emulates the sampling bias in the presence records [48].

Most *P. knowlesi* occurrences to date are recorded in Malaysia, Brunei and Singapore, with all three of these countries having eliminated the human malaria species (*e.g.*, *P. vivax* and *P. falciparum*), such that that *P. knowlesi* is routinely considered a potential cause of malaria cases. Outside of these countries, surveillance for *P. knowlesi* is limited and infection records are sparse. As per the 2015 study, the goal of our niche modelling analysis is to predict broadly into the under-sampled regions outside of Malaysia, Brunei, and Singapore, using a model fit to data from within these three countries (i.e. the model training region, Figure 2A) where we can account for reporting bias through the selection of background points. Data from outside of these three countries formed the evaluation dataset (i.e. the model evaluation region, Figure 2B), which we used to assess the model's predictive ability outside of the training region.

To produce background points for the human and mosquito records, as in the 2015 model [31], we sampled points across the training region, with this sampling weighted by human population density [41] under the assumption that more populous areas would have a greater probability of reporting human cases and that the locations of mosquito infection studies were selected based on the presence of human *P. knowlesi* cases. To produce macaque background records, we sampled points from a survey of macaques and other mammals [35], as we expected this survey to have similar sampling bias to that of macaque *P. knowlesi* infection records. This approach is not biased by the under-ascertainment of *P. knowlesi* infections that arise due to asymptomatic/submicroscopic or spontaneously resolving disease [9, 10, 11], given such effects would be expected to be uniform geographically.

As in the 2015 model [31], the geographic distribution of *Macaca leonina* — a putative host species of *P. knowlesi* which was only classified as a species distinct from *Macaca nemestrina* in 2001 [49] — has not been included as an explanatory covariate in model fitting as the species is not found in the model training region.

To produce each bootstrap we performed sampling with replacement across each of the combined occurrence polygons, occurrence point records and background points, using occurrence records present in the training region of Malaysia, Brunei and Singapore. We constrained this sampling so that at least 10 presence and 10 background points were present within each bootstrap. We then degraded the occurrence polygon records sampled to points via spatially uniform sampling of a singular point across the set of points bounded by each polygon (Figure 2A). For each bootstrap, we assigned weights to sampled points such that the sum of weights for presence points was equal to the sum of weights for the background points, and environmental values were assigned to each point from the set of covariate rasters corresponding to the spatial location and year the sample was recorded. We produced a covariate for host species, indicating if the sample was collected from a human, a mosquito or a macaque. We repeated this process to produce 500 bootstrapped datasets.

For each bootstrapped dataset, we fit boosted regression trees using the `gbm3` and `seegSDM` packages. Hyperparameters for model fitting were unchanged from the defaults provided by `seegSDM` version 0.1-9 (initial trees = 10, learning rate/shrinkage = 0.005, tree complexity = 4, maximum trees = 10,000). We produced predictions across each of the 500 bootstrapped models, with summary statistics including mean, variance, and interquartile range calculated for each 5×5 km grid cell across Southeast Asia (Figure S1). As in the 2015 model, we restricted predictions to areas within the range of macaque and mosquito species known to be required for zoonotic transmission of *P. knowlesi* (i.e. the overlap in range maps of at least one reservoir and one vector species), using predicted species extent maps previously reported [35]. This includes areas where such populations may not yet be present, such as Sulawesi, where *M. fascicularis* and *M. nemestrina* macaques are currently kept as pets and there is the potential for a feral population to establish [35].

We produced a multivariate environment similarity surface (MESS) map (Figure 2C), indicating geographic areas where the value of at least one environmental covariate was outside the range of values present in the training data (i.e. the model is extrapolating) or vice-versa [50].

Prediction results for each bootstrapped model, rasters of summary statistics, the code used to produce results, and the updated occurrence database have been made available at `osf.io/k5bsa` (DOI 10.17605/OSF.IO/K5BSA).

Model Evaluation

We evaluated the model's predictive performance by calculating the area under the curve (AUC) metric across both the training and evaluation datasets. For the training dataset, we estimated a 10-fold cross-validated AUC throughout the tree count optimisation process, and reported the training AUC for each bootstrap as that of the optimal model selected. Across the evaluation dataset, we calculated AUC across each bootstrapped model, with pairwise distance selection of samples performed to avoid spatial sorting bias [51].

We calculated covariate relative influence scores for each bootstrapped model, representing the number of times a variable is selected for regression tree splitting, weighted by the squared improvement to the model as a result of each split and averaged over all trees [52]. We summarised these scores across the models as means and 95% confidence intervals, with mean values also being used to rank the relative covariate importance. We further calculated accumulated local effect (ALE) scores to describe the average effect of a covariate on the prediction value across the range of each covariate. The ALE score achieves this by identifying how the model prediction changes in response to small changes in the covariate of interest while all other covariates are kept constant, allowing for the effects of covariates to be identified even when the covariates may be highly correlated [53].

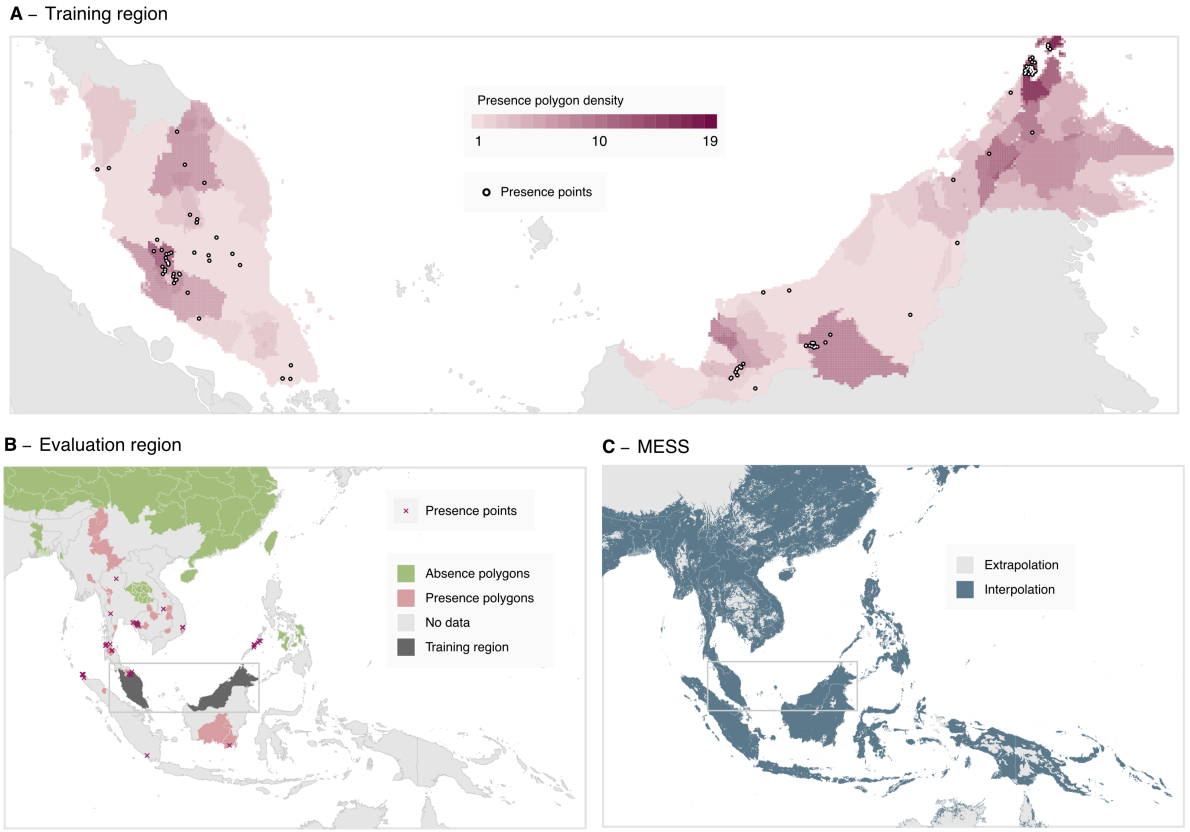

Figure 2: **A**: The data-set of occurrence points and polygons used for fitting the boosted regression tree model across the model training region of Malaysia, Brunei, and Singapore. Presence polygons are displayed as the number of polygons covering each given pixel, with this density being proportional to the probability distribution of points sampled from the polygons for each bootstrap. **B**: The presence and absence records used in the model evaluation process, across the evaluation region of Southeast Asia excluding Malaysia, Brunei and Singapore. **C**: Multivariate environmental similarity surface (MESS) for the model, where areas shaded in light grey indicate that at least one covariate value at that point is outside the range of values within the training data (extrapolation).

Results

Infection data

The literature review of articles including data on *P. knowlesi* infection occurrences published between October 2015 and March 2020 returned 511 candidate articles. Following a review of titles and abstracts, 159 articles were deemed likely to contain data for extraction, and 56 articles were identified as meeting the final criteria (Figure 1A). From these 56 articles, 264 occurrences of *P. knowlesi* were extracted, with 91 (34%) being assigned a point record type and 173 (66%) being assigned a polygon record type. Of the 264 extracted records, 241 (91%) were infections identified in humans, with only 14 in macaques and nine in mosquitoes (Table S1). The number of records by year of sample collection was greatest in 2014 with 80 records across 14 publications (Figure 1C).

A majority of records added to the 2015 database were collected in Malaysia ($n = 201$, 76%). Within Malaysia, the spatial distribution of records was highly heterogeneous (Figure 1, Table S1), with 127 polygon records assigned to the region of Sabah in contrast to the three records identified in the capital region of Kuala Lumpur. Malaysia was also the location of eight of the nine observed infected mosquitoes in the dataset, consistent with the greater sampling effort within the country [19].

Our literature search reveals that more infection occurrences from Indonesia have been reported since 2015, comprising 17% ($n = 45$) of the new presence records (where the prior 2015 literature search identified only five infections within the country). These records are the result of a small number of high-quality surveys and case reports from Aceh [54] and North Sumatra [55]. The literature review dataset contains three records from Laos, where the first confirmed human *P. knowlesi* infection was reported in 2016 [56].

In combination with the 260 occurrences used in the 2015 analysis, the total number of infection occurrences used in model fitting and evaluation was 524. Of these, 396 were within the training region of Malaysia, Brunei and Singapore, with the remaining 128 located elsewhere in Southeast Asia (Table S1).

Transmission Suitability Model Output

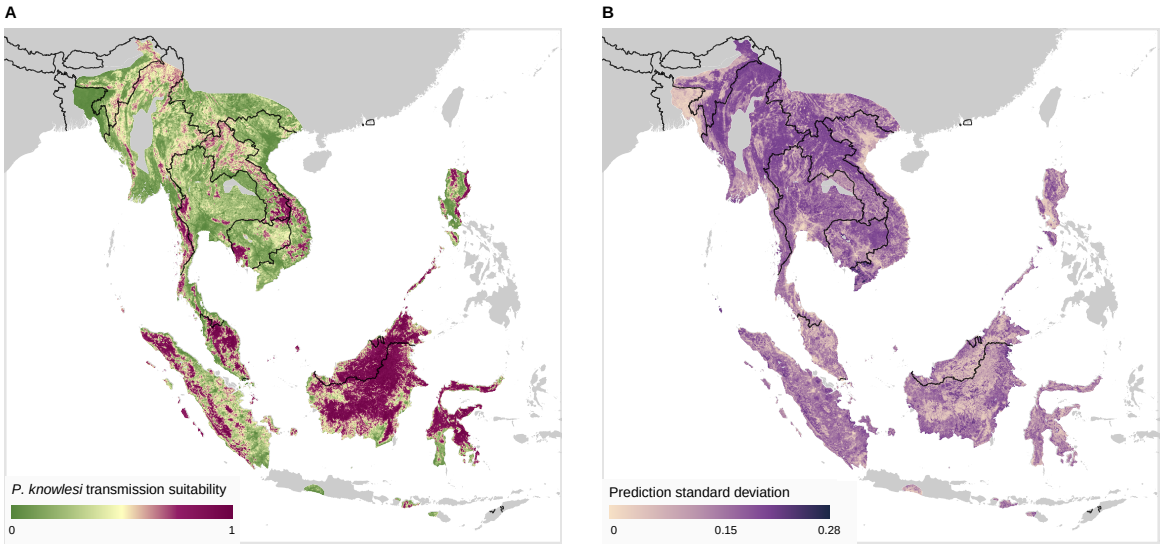

Figure 3: **A:** Modelled transmission suitability mean over Southeast Asia across the 500 bootstraps. Results are displayed only where an area is within the range of both a vector and reservoir species necessary for transmission (see Methods), regions outside of this range (displayed as grey) are considered to be very low risk for *P. knowlesi* transmission. Transmission suitability is a relative measure of the risk of *P. knowlesi* transmission from known reservoir species (via vector species) to humans. **B:** Standard deviation of the predicted transmission suitability across the 500 bootstraps.

The mean and standard deviation of predicted *P. knowlesi* transmission suitability across at-risk areas of Southeast Asia is presented in Figure 3. Further summary statistics of transmission suitability are presented in Figure S1.

The map of *P. knowlesi* transmission suitability (Figure 3A) shows highly heterogeneous levels of predicted risk across Southeast Asia. On the island of Borneo, all areas other than lower-lying coastal regions are expected to have a relatively high risk of *P. knowlesi* transmission. Other more sparsely distributed areas of relatively high risk are predicted in Indonesia within the provinces of Sulawesi, Sumatra and West Nusa Tenggara. Peninsula Malaysia is predicted to have inland areas of high transmission risk. Thailand, Laos, Cambodia, Vietnam, Myanmar and the Philippine island of Luzon have smaller, localised areas of high predicted risk, with greater uncertainty in these predictions (Figure 3B) as a result of environmental differences to the model training region of Malaysia, Brunei and Singapore.

Within the training region, a mean area under the curve (AUC) of 0.81 was produced across the 500 bootstrapped models with a standard error of 0.001. For the evaluation region, the mean AUC was found to be 0.75 with a standard error of 0.003. These values indicate a high degree of predictive performance.

Examining predictions within the evaluation region of the model (Southeast Asia excluding Malaysia, Brunei and Singapore), we may qualitatively assess the model's predictive performance. Regions with both a high modelled transmission suitability and previously identified occurrence samples of *P. knowlesi* — indicative of good model sensitivity — include the Aceh province of Sumatra island in Indonesia, the Koh Kong province in southern Cambodia and the Mimaropa region of the Philippines (Figure S2A). We also see that there are a substantial number of regions where the model predicts high transmission suitability where *P. knowlesi* occurrence has not previously been identified as of the 2020 literature review (i.e. omission errors [57]). These areas include much of northern Sulawesi and the province of West Nusa Tenggara in Indonesia (Figure S2A). Such predictions may be suggestive of a lack of surveillance in these regions, or that an environment is conducive to transmission but currently lacking widespread occurrence of a necessary vector or host species (e.g. in Sulawesi, there are no native *M. nemestrina* or *M. fascicularis* macaques).

The covariate of human population density was found to have the highest ranked relative influence for the majority (496/500, 99.2%) of the bootstrapped models, closely followed by that of healthcare accessibility. Mean and 95% confidence intervals of relative influence scores across bootstraps are presented in Figure S5. The new covariates of tree coverage and forest loss were found to be highly influential; out of 21 covariates (20 environmental covariates and the species covariate), the median rank for tree coverage was 5 (95% confidence interval: 2-11), and for forest loss was 10 (95% confidence interval: 6-15). Plots of the accumulated local effects (ALE) describing the influence of each continuous covariate across the covariate's range are presented in Figure S6.

Discussion

In this study, we utilised an environmental niche modelling approach to predict the relative suitability for *P. knowlesi* transmission to humans across Southeast Asia. We extended a previous analysis that incorporated data up to 2015 [31] by adding infection and environmental data up to 2020, and improving the utilisation of data on land use patterns. Through a review of literature published between October 2015 and March 2020, we identified 264 published occurrences of *P. knowlesi*. This resulted in a total of 524 records being utilised in model fitting and evaluation for the current study. As changes in *P. knowlesi* transmission risk may be expected where substantial amounts of deforestation have occurred [25, 24, 26, 27], we now capture this in the model by deriving annual forest loss and coverage datasets. We predict that the distribution of *P. knowlesi* risk is highly heterogeneous across Southeast Asia, with the largest areas of predicted risk in Malaysia and Indonesia, and smaller, localised regions of high risk predicted in the Greater Mekong Subregion, The Philippines and Northeast India.

Our analysis can help to guide the prioritisation of locations for future sampling and surveillance for P. knowlesi malaria by highlighting areas of high predicted risk that may have been under-sampled. Since the publication of the 2015 analysis, there has been no change to the World Health Organization's malaria elimination status of any country believed to be at risk for indigenous *P. knowlesi* transmission [58]. However, within the Greater Mekong Subregion of Cambodia, Myanmar, Thailand, Laos and Vietnam, substantial declines have been observed in the total number of reported malaria cases as of 2021 [59]. The 2015 analysis noted that Laos, Myanmar, Thailand and Vietnam were likely high-value sites for future sampling efforts [31] and our literature search revealed only a small number of additional *P. knowlesi* occurrences in these countries as of 2020 (Figure 1, Table S1). Our current analysis predicts localised areas of moderate-to-high relative transmission risk in this region (Figure 3), suggesting an ongoing need for surveillance of *P. knowlesi* malaria.

Indonesia has a stated goal to eliminate malaria by 2030 [60, 61], and may be on track given that a majority of administrative regions have declared elimination [62]. However, the presence of *P. knowlesi* across the country presents a serious challenge to these efforts. In March 2022, the WHO Malaria Policy Advisory Group (MPAG) concluded that certification of malaria elimination status should only occur where the risk of *P. knowlesi* was 'negligible', i.e. below some low threshold of annual incidence [63, 58] – a requirement that has already prevented Malaysia from receiving elimination certification [64]. Given this requirement, continued surveillance and mitigation of *P. knowlesi* throughout at-risk regions of Indonesia will be important. Between the period of 2015 and 2020, a small number of studies have identified substantial numbers of *P. knowlesi* infections within Indonesia [65, 66, 67], particularly within northern Sumatra [54, 68, 69], a region identified as a valuable target for surveillance effort in the 2015 model [31]. Despite this, the Indonesian region of Kalimantan on the island of Borneo still has a relative scarcity of occurrence data given its high predicted transmission suitability and the number of *P. knowlesi* cases reported in adjacent areas of Malaysia.

Our map of *P. knowlesi* transmission risk may also help to quantitatively guide site selection for public health surveillance or intervention. For example, surveillance sampling could be concentrated in regions where the model predicts a high transmission suitability but with a high variance, such that the understanding of the geographical distribution of the disease is maximised for the least effort and that the uncertainty in these regions could be reduced in future risk mapping outputs. If value was instead placed on maximising the probability of identifying cases of *P. knowlesi*, sampling could be concentrated where a high transmission suitability is accompanied by lower variance. Efficient deployment of sampling resources could be achieved by combining the modelling outputs with constraints; for example, sites where access would require a prohibitive amount of travel time could be excluded [70].

As niche modelling frameworks are correlative, the secondary results described in this work should be interpreted with care. The relative influence scores (Figure S5) and the accumulated local effect plots (Figure S6) may provide insight into risk factors for *P. knowlesi* transmission. However, these results do not provide evidence for causal relationships, which would instead be more appropriately identified through studies utilising a causal inference framework. For example, the covariate of healthcare accessibility, which ranks highly according to relative influence scores (Figure S5), could capture a direct causal effect on the risk of being diagnosed with *P. knowlesi* (e.g. likelihood that someone is identified as having a *P. knowlesi* infection increasing with access to healthcare) or may simply be confounded by a common variable (e.g. likelihood of acquiring a *P. knowlesi* infection increasing for those who work at plantations, confounded by such plantations occurring in areas of lower healthcare accessibility).

Our model predicts the relative suitability for *P. knowlesi* transmission, not the prevalence of infection nor the incidence of cases (which would require different input data that are not widely available for *P. knowlesi* malaria). While transmission suitability is a useful metric for prioritising locations for future P. knowlesi surveys, the absolute values are specific to the input data and model parameterisation, and we therefore cannot directly compare absolute values produced by the model presented here and those from the model developed in 2015. Although we expect that the transmission suitability prediction produced by either of the models should be qualitatively related to the underlying 'true' risk of *P. knowlesi* infection, little can be said of this relationship other than that it is expected to be monotonic under the assumption that the background data points are biased in the same manner as the presence data [48]. This means, for instance, that any differences between the models that could arise as a result of dilation in this relationship (such as the upwards dilation observed in Figures S4A and S4B) cannot be taken alone as indicating a change in underlying transmission suitability.

Noting the limitations in these comparisons, we find that the predictions in our work and the 2015 model broadly align, though with clear differences in the local spatial variation of the prediction surface (Figures S3, S4A). As an example, on the island of Borneo our predictions form a smooth region of high predicted risk, whereas in the 2015 model predictions over the same area varied substantially at a small spatial scale; this pattern is repeated similarly elsewhere across Southeast Asia [31]. In countries such as Laos, Myanmar and Vietnam, we predict overall a lower transmission suitability than those presented in the 2015 model, though within these countries we continue to predict small areas of high transmission risk. Comparing the overall distributions of predicted transmission suitability between the 2015 and 2020 models shows that our new predictions produce a more highly contrasting bimodal distribution of risk compared to that produced by the 2015 model [31] (Figure S4A).

The temperature suitability index covariate used in the model attempts to describe the effect of temperature on the basic reproduction number for some combination of malaria parasites and mosquito vectors. As data on the incubation periods for *P. knowlesi* under differing temperatures and mosquito hosts is currently unavailable, no suitability index for the species can currently be produced. In this work, we instead utilise a proxy in the form of a suitability index for *P. falciparum* [40]. Even if this proxy does not itself accurately capture mechanistic limits on *P. knowlesi* reproduction, it is not immediately obvious what bias this would introduce into the results, if any, as the boosted regression tree model may still infer suitability under some transformation of the index. There is clear value in further laboratory research on the reproduction of *P. knowlesi* under different temperatures that could inform a species-specific suitability index.

It is believed that workers involved in the development and cultivation of oil palm plantations are at greater risk for developing *P. knowlesi* infection given their proximity to *P. knowlesi* vector and reservoir species [71]. However, we were unable to include this as a covariate in our model as there is currently no published dataset of palm oil plantations with complete coverage across the Southeast Asia region.

Annual data is not available for some of the covariates used in the model where the underlying phenomena may be expected to change over time; the covariates of reservoir/vector species distribution and temperature suitability are dependent upon variables such as climate or land cover, and the covariate of healthcare accessibility is dependent upon changes in transportation infrastructure and locations of healthcare sites. In lieu of available data on change in these covariates over time they are instead assumed to be constant. In effect, this means that the modelled species distributions as of 2014, temperature suitability index for *P. falciparum* as of 2010 and healthcare accessibility as of 2019 are all assumed constant over the years 2001 to 2019. Future modelling efforts could be improved by considering the change in these covariates over time.

Our map of *P. knowlesi* transmission suitability predict high *P. knowlesi* disease risk across broad areas of Southeast Asia, with large regions of high predicted *P. knowlesi* risk that have not yet been sampled for the pathogen. Our work demonstrates the importance of continued surveillance and prospective sampling of the pathogen, especially in regions where malaria elimination is currently being pursued.

Acknowledgements

This study was supported through funding provided by the Australian Centre for International Agricultural Research (ACIAR), as part of the 'Evaluating zoonotic malaria transmission and agricultural and forestry land use in Indonesia' (ZOOMAL) project (LS/2019/116, www.aciar.gov.au).

Further support for this project was provided by the National Health and Medical Research Council of Australia through its Centres of Research Excellence (ACREME, GNT1134989 www.nhmrc.gov.au).

FMS was supported by the National Health and Medical Research Council of Australia Investigator Grant Scheme (Emerging Leader Fellowship, 2021/GNT2010051 www.nhmrc.gov.au). JAF was supported by the Australian Research Council (ARC, FT210100034 and DP200100747 www.arc.gov.au). LEH was supported by a Melbourne Research Scholarship from the University of Melbourne (www.unimelb.edu.au). MJG was supported by the National Health and Medical Research Council of Australia Investigator Grant Scheme (Emerging Leader 2 Fellowship, 2023/GNT2017436 www.nhmrc.gov.au). GSR and MJG were supported by the National Institutes of Health, USA (R01AI160457-01 www.nih.gov). GSR was also supported by the Malaysian Ministry of Health (Grant Number BP00500/117/1002 www.moh.gov.my).

This research was supported by The University of Melbourne's Research Computing Services and the Petascale Campus Initiative (www.unimelb.edu.au).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

We thank Dr Timothy William for their support. We would like to also thank the Director General of Health Malaysia for the permission to publish this article.

Data Availability

Prediction results for each bootstrapped model, rasters of summary statistics, the code used to produce results, and the updated occurrence database have been made available at osf.io/k5bsa (DOI 10.17605/OSF.IO/K5BSA).

References

[1] I. Vythilingam, M. L. Wong, and W. S. Wan-Yussof. Current status of *Plasmodium knowlesi* vectors: a public health concern? *Parasitology*, 145(1):32–40, May 2016.

[2] William E. Collins and John W. Barnwell. *Plasmodium knowlesi*: Finally being recognized. *The Journal of Infectious Diseases*, 199(8):1107–1108, April 2009.

[3] William Chin, Edward Alpert, William E. Collins, Marvin H. Jeter, and Peter G. Contacos. Experimental Mosquito-Transmission of *Plasmodium Knowlesi* to Man and Monkey. *The American Journal of Tropical Medicine and Hygiene*, 17(3):355–358, May 1968.

[4] Pablo Ruiz Cuenca, Stephanie Key, Kim A. Lindblade, Indra Vythilingam, Chris Drakeley, and Kimberly Fornace. Is there evidence of sustained human-mosquito-human transmission of the zoonotic malaria *Plasmodium knowlesi*? a systematic literature review. *Malaria Journal*, 21(1), March 2022.

[5] Balbir Singh and Cyrus Daneshvar. Human Infections and Detection of *Plasmodium knowlesi*. *Clinical Microbiology Reviews*, 26(2):165–184, April 2013.

[6] George Robert Coatney. *The primate malarias*. US National Institute of Allergy and Infectious Diseases, 1971.

[7] Paddy M. Brock, Kimberley M. Fornace, Minnie Parmiter, Jonathon Cox, Chris J. Drakeley, Heather M. Ferguson, and Rowland Raymond Kao. *Plasmodium knowlesi* transmission: integrating quantitative approaches from epidemiology and ecology to understand malaria as a zoonosis. *Parasitology*, 143(4):389–400, January 2016.

[8] Bridget E. Barber, Matthew J. Grigg, Daniel J. Cooper, Donelly A. van Schalkwyk, Timothy William, Giri S. Rajahram, and Nicholas M. Anstey. Clinical management of *Plasmodium knowlesi* malaria. In *Current research on naturally transmitted Plasmodium knowlesi*, pages 45–76. Elsevier, 2021.

[9] Mallika Imwong, Wanassanan Madmanee, Kanokon Suwannasin, Chanon Kunasol, Thomas J Peto, Rupam Tripura, Lorenz von Seidlein, Chea Nguon, Chan Davoeung, Nicholas P J Day, Arjen M Dondorp, and Nicholas J White. Asymptomatic natural human infections with the simian malaria parasites *Plasmodium cynomolgi* and *Plasmodium knowlesi*. *The Journal of Infectious Diseases*, 219(5):695–702, October 2018.

[10] Kimberly M. Fornace, Nor Afizah Nuin, Martha Betson, Matthew J. Grigg, Timothy William, Nicholas M. Anstey, Tsin W. Yeo, Jonathan Cox, Lau Tiek Ying, and Chris J. Drakeley. Asymptomatic and submicroscopic carriage of *Plasmodium knowlesi* malaria in household and community members of clinical cases in Sabah, Malaysia. *Journal of Infectious Diseases*, 213(5):784–787, October 2015.

[11] Matthew J Grigg, Jonathan Cox, Timothy William, Jenarun Jelip, Kimberly M Fornace, Patrick M Brock, Lorenz von Seidlein, Bridget E Barber, Nicholas M Anstey, Tsin W Yeo, and Christopher J Drakeley. Individual-level factors associated with the risk of acquiring human *Plasmodium knowlesi* malaria in Malaysia: a case-control study. *The Lancet Planetary Health*, 1(3):e97–e104, June 2017.

[12] Cyrus Daneshvar, Timothy M. E. Davis, Janet Cox-Singh, Mohammad Zakri Rafa'ee, Siti Khatijah Zakaria, Paul C. S. Divis, and Balbir Singh. Clinical and laboratory features of human *Plasmodium knowlesi* infection. *Clinical Infectious Diseases*, 49(6):852–860, September 2009.

[13] Matthew J Grigg, Timothy William, Bridget E Barber, Giri S Rajahram, Jayaram Menon, Emma Schimann, Kim Piera, Christopher S Wilkes, Kaajal Patel, Arjun Chandna, Christopher J Drakeley, Tsin W Yeo, and Nicholas M Anstey. Age-related clinical spectrum of *Plasmodium knowlesi* malaria and predictors of severity. *Clinical Infectious Diseases*, 67(3):350–359, March 2018.

[14] Kimberly M. Fornace, Lou S. Herman, Tommy R. Abidin, Tock Hing Chua, Sylvia Daim, Pauline J. Lorenzo, Lynn Grignard, Nor Afizah Nuin, Lau Tiek Ying, Matthew J. Grigg, Timothy William, Fe Espino, Jonathan Cox, Kevin K. A. Tetteh, and Chris J. Drakeley. Exposure and infection to *Plasmodium knowlesi* in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines. *PLOS Neglected Tropical Diseases*, 12(6):e0006432, June 2018.

[15] Kimberly M Fornace, Paddy M Brock, Tommy R Abidin, Lynn Grignard, Lou S Herman, Tock H Chua, Sylvia Daim, Timothy William, Catriona L E B Patterson, Tom Hall, Matthew J Grigg, Nicholas M Anstey, Kevin K A Tetteh, Jonathan Cox, and Chris J Drakeley. Environmental risk factors and exposure to the zoonotic malaria parasite *Plasmodium knowlesi* across northern sabah, malaysia: a population-based cross-sectional survey. *The Lancet Planetary Health*, 3(4):e179–e186, April 2019.

[16] M Vadivelan, TK Dutta, et al. Recent advances in the management of *Plasmodium knowlesi* infection. *Trop Parasitol*, 4(1):31–4, 2014.

[17] Aongart Mahittikorn, Frederick Ramirez Masangkay, Kwuntida Uthaisar Kotepui, Giovanni De Jesus Milanez, and Manas Kotepui. Quantification of the misidentification of *Plasmodium knowlesi* as *Plasmodium malariae* by microscopy: an analysis of 1569 P. knowlesi cases. *Malaria Journal*, 20(1), April 2021.

[18] Kim-Sung Lee, Janet Cox-Singh, and Balbir Singh. Morphological features and differential counts of *Plasmodium knowlesi* parasites in naturally acquired human infections. *Malaria Journal*, 8(1), April 2009.

[19] Daniel J Cooper, Giri S Rajahram, Timothy William, Jenarun Jelip, Rashidah Mohammad, Joseph Benedict, Danshy A Alaza, Eva Malacova, Tsin W Yeo, Matthew J Grigg, Nicholas M Anstey, and Bridget E Barber. *Plasmodium knowlesi* malaria in Sabah, Malaysia, 2015–2017: Ongoing increase in incidence despite near-elimination of the human-only *Plasmodium* species. *Clinical Infectious Diseases*, 70(3):361–367, March 2019.

[20] Rupesh K Tyagi, Manoj K Das, Shiv S Singh, and Yagya D Sharma. Discordance in drug resistance-associated mutation patterns in marker genes of Plasmodium falciparum and Plasmodium knowlesi during coinfections. *J. Antimicrob. Chemother.*, 68(5):1081–1088, May 2013.

[21] World Health Organization. *Global technical strategy for malaria 2016-2030*. World Health Organization, 2015.

[22] Fauzi Muh, Namhyeok Kim, Myat Htut Nyunt, Egy Rahman Firdaus, Jin-Hee Han, Mohammad Rafiul Hoque, Seong-Kyun Lee, Ji-Hoon Park, Robert W. Moon, Yee Ling Lau, Osamu Kaneko, and Eun-Taek Han. Cross-species reactivity of antibodies against *Plasmodium vivax* blood-stage antigens to *Plasmodium knowlesi*. *PLOS Neglected Tropical Diseases*, 14(6):e0008323, June 2020.

[23] Abraham Zefong Chin, Marilyn Charlene Montini Maluda, Jenarun Jelip, Muhammad Saffree Bin Jeffree, Richard Culleton, and Kamruddin Ahmed. Malaria elimination in Malaysia and the rising threat of *Plasmodium knowlesi*. *Journal of Physiological Anthropology*, 39(1), November 2020.

[24] Pablo Ruiz Cuenca, Stephanie Key, Amaziasizamoria Jumail, Henry Surendra, Heather M. Ferguson, Chris J. Drakeley, and Kimberly Fornace. Epidemiology of the zoonotic malaria *Plasmodium knowlesi* in changing landscapes. In *Current research on naturally transmitted Plasmodium knowlesi*, pages 225–286. Elsevier, 2021.

[25] Patrick M. Brock, Kimberly M. Fornace, Matthew J. Grigg, Nicholas M. Anstey, Timothy William, Jon Cox, Chris J. Drakeley, Heather M. Ferguson, and Rowland R. Kao. Predictive analysis across spatial scales links zoonotic malaria to deforestation. *Proceedings of the Royal Society B: Biological Sciences*, 286(1894):20182351, January 2019.

[26] Gael Davidson, Tock H. Chua, Angus Cook, Peter Speldewinde, and Philip Weinstein. The role of ecological linkage mechanisms in *Plasmodium knowlesi* transmission and spread. *EcoHealth*, 16(4):594–610, January 2019.

[27] Danica J. Stark, Kimberly M. Fornace, Patrick M. Brock, Tommy Rowel Abidin, Lauren Gilhooly, Cyrlen Jalius, Benoit Goossens, Chris J. Drakeley, and Milena Salgado-Lynn. Long-tailed macaque response to deforestation in a *Plasmodium knowlesi*-endemic area. *EcoHealth*, 16(4):638–646, March 2019.

[28] William K. Reisen. Landscape epidemiology of vector-borne diseases. *Annual Review of Entomology*, 55(1):461–483, January 2010.

[29] Huan Yu, Xiangmeng Liu, Bo Kong, Ruopu Li, and Guangxing Wang. Landscape ecology development supported by geospatial technologies: A review. *Ecological Informatics*, 51:185–192, 2019.

[30] Robert P. Anderson, Enrique Martínez-Meyer, Miguel Nakamura, Miguel B. Araújo, A. Townsend Peterson, Jorge Soberón, and Richard G. Pearson. *Ecological Niches and Geographic Distributions (MPB-49)*. Princeton University Press, December 2011.

[31] Freya M Shearer, Zhi Huang, Daniel J Weiss, Antoinette Wiebe, Harry S Gibson, Katherine E Battle, David M Pigott, Oliver J Brady, Chaturong Putaporntip, Somchai Jongwutiwes, et al. Estimating geographical variation in the risk of zoonotic *Plasmodium knowlesi* infection in countries eliminating malaria. *PLOS Neglected Tropical Diseases*, 10(8):e0004915, 2016.

[32] D. J. Weiss, A. Nelson, H. S. Gibson, W. Temperley, S. Peedell, A. Lieber, M. Hancher, E. Poyart, S. Belchior, N. Fullman, B. Mappin, U. Dalrymple, J. Rozier, T. C. D. Lucas, R. E. Howes, L. S. Tusting, S. Y. Kang, E. Cameron, D. Bisanzio, K. E. Battle, S. Bhatt, and P. W. Gething. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553(7688):333–336, January 2018.

[33] D. J. Weiss, A. Nelson, C. A. Vargas-Ruiz, K. Gligorić, S. Bavadekar, E. Gabrilovich, A. Bertozzi-Villa, J. Rozier, H. S. Gibson, T. Shekel, C. Kamath, A. Lieber, K. Schulman, Y. Shao, V. Qarkaxhija, A. K. Nandi, S. H. Keddie, S. Rumisha, P. Amratia, R. Arambepola, E. G. Chestnutt, J. J. Millar, T. L. Symons, E. Cameron, K. E. Battle, S. Bhatt, and P. W. Gething. Global maps of travel time to healthcare facilities. *Nature Medicine*, 26(12):1835–1838, September 2020.

[34] Peter Potapov, Aleksey Yaroshenko, Svetlana Turubanova, Maxim Dubinin, Lars Laestadius, Christoph Thies, Dmitry Aksenov, Aleksey Egorov, Yelena Yesipova, Igor Glushkov, et al. Mapping the World's intact forest landscapes by remote sensing. *Ecology and Society*, 13(2), 2008.

[35] Catherine L Moyes, Freya M Shearer, Zhi Huang, Antoinette Wiebe, Harry S Gibson, Vincent Nijman, Jayasilan Mohd-Azlan, Jedediah F Brodie, Suchinda Malaivijitnond, Matthew Linkie, et al. Predicting the geographical distributions of the macaque hosts and mosquito vectors of *Plasmodium knowlesi* malaria in forested and non-forested areas. *Parasites & vectors*, 9(1):1–12, 2016.

[36] Matthew C Hansen, Peter V Potapov, Rebecca Moore, Matt Hancher, Svetlana A Turubanova, Alexandra Tyukavina, David Thau, SV Stehman, Scott J Goetz, Thomas R Loveland, et al. High-resolution global maps of 21st-century forest cover change. *science*, 342(6160):850–853, 2013.

[37] Tom G Farr, Paul A Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, Mimi Paller, Ernesto Rodriguez, Ladislav Roth, et al. The shuttle radar topography mission. *Reviews of geophysics*, 45(2), 2007.

[38] S. E. Lobser and Warren B. Cohen. MODIS tasselled cap: land cover characteristics expressed through transformed MODIS data. *International Journal of Remote Sensing*, 28:5079 – 5101, 2007.

[39] Daniel J Weiss, Peter M Atkinson, Samir Bhatt, Bonnie Mappin, Simon I Hay, and Peter W Gething. An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 98:106–118, 2014.

[40] Peter W Gething, Thomas P Van Boeckel, David L Smith, Carlos A Guerra, Anand P Patil, Robert W Snow, and Simon I Hay. Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and P. vivax. *Parasites & vectors*, 4(1):1–11, 2011.

[41] Andrea E. Gaughan, Forrest R. Stevens, Catherine Linard, Peng Jia, and Andrew J. Tatem. High resolution population distribution maps for Southeast Asia in 2010 and 2015. *PLOS ONE*, 8(2):e55882, February 2013.

[42] Andrea E Gaughan, Forrest R Stevens, Catherine Linard, Peng Jia, and Andrew J Tatem. High resolution population distribution maps for Southeast Asia in 2010 and 2015. *PLOS one*, 8(2):e55882, 2013.

[43] Mark A Friedl, Damien Sulla-Menashe, Bin Tan, Annemarie Schneider, Navin Ramankutty, Adam Sibley, and Xiaoman Huang. MODIS collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote sensing of Environment*, 114(1):168–182, 2010.

[44] Glenn De'ath and Katharina E. Fabricius. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, 81(11):3178–3192, November 2000.

[45] J. Elith, J. R. Leathwick, and T. Hastie. A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4):802–813, July 2008.

[46] Jane Elith, Catherine H. Graham, Robert P. Anderson, Miroslav Dudík, Simon Ferrier, Antoine Guisan, Robert J. Hijmans, Falk Huettmann, John R. Leathwick, Anthony Lehmann, Jin Li, Lucia G. Lohmann, Bette A. Loiselle, Glenn Manion, Craig Moritz, Miguel Nakamura, Yoshinori Nakazawa, Jacob McC. M. Overton, A. Townsend Peterson, Steven J. Phillips, Karen Richardson, Ricardo Scachetti-Pereira, Robert E. Schapire, Jorge Soberón, Stephen Williams, Mary S. Wisz, and Niklaus E. Zimmermann. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2):129–151, March 2006.

[47] JR Leathwick, J Elith, MP Francis, T Hastie, and P Taylor. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series*, 321:267–281, September 2006.

[48] Steven J. Phillips, Miroslav Dudík, Jane Elith, Catherine H. Graham, Anthony Lehmann, John Leathwick, and Simon Ferrier. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19(1):181–197, January 2009.

[49] Colin Groves. *Primate Taxonomy*. Smithsonian Series in Comparative Evolutionary Biology. Smithsonian Books, Washington, D.C., DC, May 2001.

[50] Jane Elith, Michael Kearney, and Steven Phillips. The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1(4):330–342, 2010.

[51] Robert J. Hijmans. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, 93(3):679–688, March 2012.

[52] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[53] Christoph Molnar. *Interpretable Machine Learning: A Guide For Making Black Box Models Explainable*. Independently published, 2020.

[54] Herdiana Herdiana, Chris Cotter, Farah N Coutrier, Iska Zarlinda, Brittany W Zelman, Yusrifar Kharisma Tirta, Bryan Greenhouse, Roly D Gosling, Peter Baker, Maxine Whittaker, et al. Malaria risk factor assessment using active and passive surveillance data from Aceh Besar, Indonesia, a low endemic, malaria elimination setting with *Plasmodium knowlesi*, *Plasmodium vivax*, and *Plasmodium falciparum*. *Malaria journal*, 15(1):1–15, 2016.

[55] Inke N. D. Lubis, Hendri Wijaya, Munar Lubis, Chairuddin P. Lubis, Paul C. S. Divis, Khalid B. Beshir, and Colin J. Sutherland. Contribution of *Plasmodium knowlesi* to multispecies human malaria infections in North Sumatera, Indonesia. *The Journal of Infectious Diseases*, 215(7):1148–1155, February 2017.

[56] Moritoshi Iwagami, Masami Nakatsu, Phonepadith Khattignavong, Pheovaly Soundala, Lavy Lorphachan, Sengdeuane Keomalaphet, Phonepadith Xangsayalath, Satoru Kawai, Bouasy Hongvanthong, Paul T Brey, et al. First case of human infection with *Plasmodium knowlesi* in Laos, 2018.

[57] A Townsend Peterson. *Mapping disease transmission risk: enriching models using biogeography and ecology*. JHU Press, 2014.

[58] World Health Organization. World malaria report 2022. 2022.

[59] World Health Organization. Accelerating malaria elimination in the greater mekong. 2022.

[60] Departemen Kesehatan. Keputusan Menteri Kesehatan Republik Indonesia Nomor 293/MENKES/SK/IV/2009 28 April 2009 tentang Eliminasi Malaria di Indonesia, 2009.

[61] Ministry of Health Indonesia – Kementerian Kesehatan Republik Indonesia. Challenges Toward Malaria Elimination 2030. 2021.

[62] Vensya Sitohang, Elvieda Sariwati, Sri Budi Fajariyani, Dasom Hwang, Bayu Kurnia, Ratih Ketana Hapsari, Ferdinand Johannis Laihad, Maria Endang Sumiwi, Paul Pronyk, and William A Hawley. Malaria elimination in Indonesia: halfway there. *The Lancet Global Health*, 6(6):e604–e606, June 2018.

[63] World Health Organisation et al. WHO Malaria Policy Advisory Group (MPAG) meeting: meeting report, March. 2022.

[64] Kimberly M Fornace, Gabriel Zorello Laporta, Indra Vythilingham, Tock Hing Chua, Kamruddin Ahmed, Nantha K Jeyaprakasam, Ana Maria Ribeiro de Castro Duarte, Amirah Amir, Wei Kit Phang, Chris Drakeley, Maria Anice M Sallum, and Yee Ling Lau. Simian malaria: a narrative review on emergence, epidemiology and threat to global malaria elimination. *Lancet Infect. Dis.*, July 2023.

[65] I.N.D. Lubis, H. Wijaya, M. Lubis, C.P. Lubis, and C.J. Sutherland. Molecular identification of human *Plasmodium knowlesi* infections in North Sumatera, Indonesia. *International Journal of Infectious Diseases*, 45:182, April 2016.

[66] Agus A. Wibowo, Sitti R. Umniyati, Jontari Hutagalung, and Tanti Rahayu. Confirmation of *Anopheles balabacensis* as natural vector of malaria caused by *Plasmodium knowlesi* inhabits forested areas in *Kecamatan Balik Bukit*, pwestern lampung regency. 151:01028, 2020.

[67] Wuryantari Setiadi, Herawati Sudoyo, Hidayat Trimarsanto, Boy Adventus Sihite, Riahdo Juliarman Saragih, Rita Juliawaty, Suradi Wangsamuda, Puji Budi Setia Asih, and Din Syafruddin. A zoonotic human infection with simian malaria, *Plasmodium knowlesi*, in Central Kalimantan, Indonesia. *Malaria Journal*, 15(1), April 2016.

[68] Herdiana Herdiana, Irnawati Irnawati, Farah Novita Coutrier, Alfian Munthe, Mardiati Mardiati, Titik Yuniarti, Elvieda Sariwati, Maria Endang Sumiwi, Rintis Noviyanti, Paul Pronyk, et al. Two clusters of *Plasmodium knowlesi* cases in a malaria elimination area, Sabang Municipality, Aceh, Indonesia. *Malaria journal*, 17(1):1–10, 2018.

[69] Farah Coutrier, Chris Cotter, Yusrifar K Tirta, Alanna Schwartz, Iska Zarlinda, Herdiana H Basri, Jutta Marfurt, Nicholas Anstey, William A Hawley, Asik Surya, et al. Serial molecular identification to confirm the presence of *Plasmodium knowlesi* in Indonesia. In *American Journal of Tropical Medicine and Hygiene*, volume 93, pages 263–263, 2015.

[70] Joshua Longbottom, Ana Krause, Stephen J. Torr, and Michelle C. Stanton. Quantifying geographic accessibility to improve efficiency of entomological monitoring. *PLOS Neglected Tropical Diseases*, 14(3):e0008096, March 2020.

[71] Nurul Athirah Naserrudin, Rozita Hod, Mohammad Saffree Jeffree, Kamruddin Ahmed, Richard Culleton, and Mohd Rohaizat Hassan. The role of human behavior in *Plasmodium knowlesi* malaria infection: A systematic review. *International Journal of Environmental Research and Public Health*, 19(6):3675, March 2022.

Supplementary Information

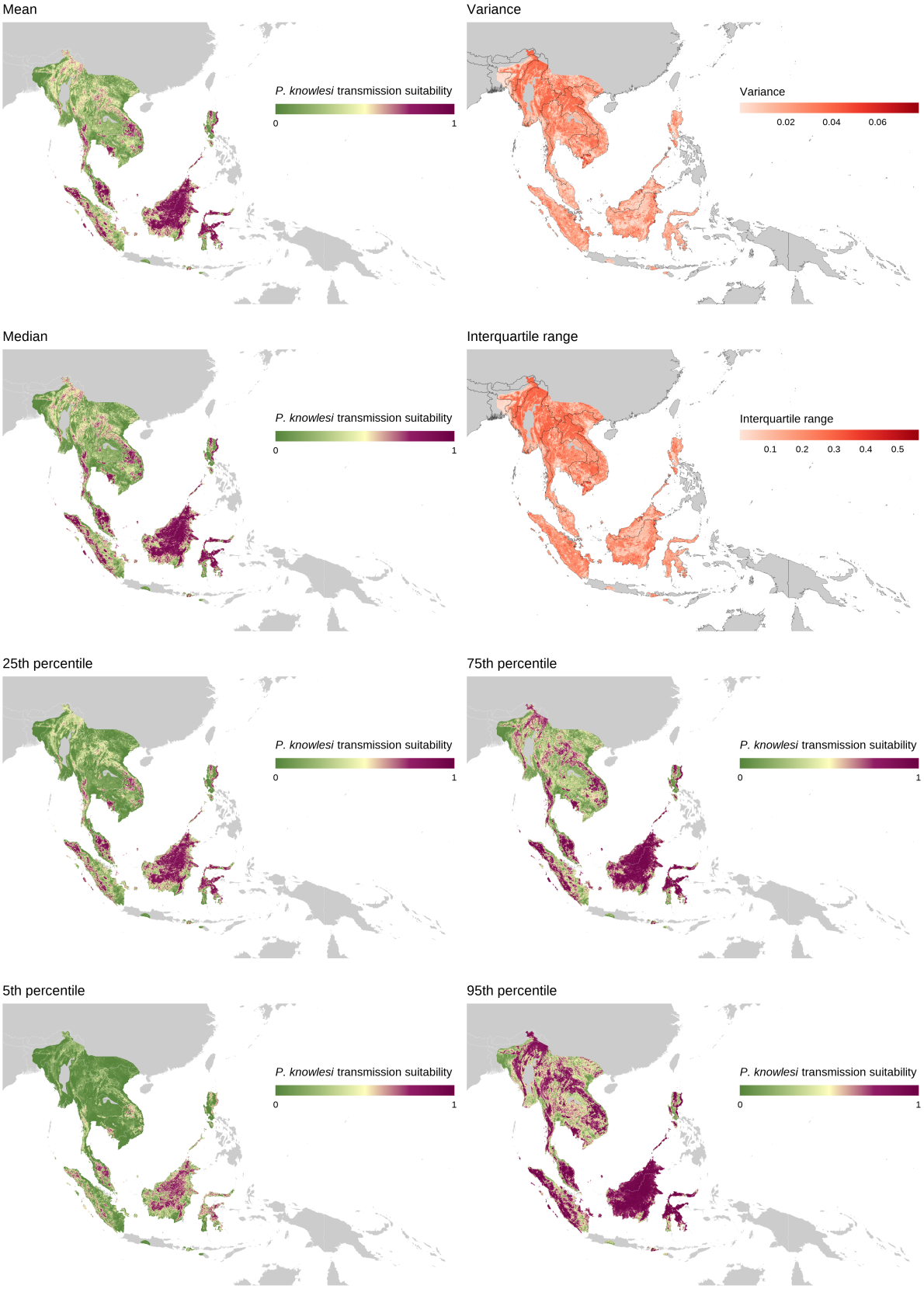

Figure S1: Summary statistics for modelled transmission suitability across Southeast Asia, calculated across the set of 500 bootstraps. Results are only displayed where an area is in the range of both a vector and reservoir species necessary for transmission (see Methods).

	Human	Macaque	Mosquito	Totals	
Country/Region	n	n	n	2020	(2015)
Brunei	0	0	0	0	(6)
Cambodia	5	0	0	5	(6)
Indonesia (Total)	44	1	0	45	(5)
Kalimantan Selatan	0	0	0	0	(3)
Kalimantan Tengah	1	0	0	1	(2)
Lampung	0	1	0	1	(0)
Nanggroe Aceh Darusalam	36	0	0	36	(0)
Sumatera Utara	7	0	0	7	(0)
Laos	3	0	0	3	(1)
Malaysia (Total)	181	8	12	201	(183)
Johor	3	0	0	3	(3)
Kedeh	2	0	0	2	(1)
Kelantan	5	0	0	5	(17)
Melaka	2	0	0	2	(1)
Negeri	2	0	2	4	(3)
Pahang	3	0	1	4	(11)
Perak	3	0	1	4	(1)
Pulau Pinang	1	0	0	1	(1)
Sabah	127	7	0	134	(60)
Sarawak	22	1	5	28	(52)
Selangor	7	0	3	10	(30)
Terengganu	1	0	0	1	(2)
W.P. Kuala Lumpur	3	0	0	3	(0)
W.P. Labuan	0	0	0	0	(1)
Myanmar	2	0	0	2	(3)
Philippines	1	0	2	3	(7)
Singapore	0	0	0	0	(6)
Thailand	4	0	0	4	(32)
Vietnam	1	0	0	1	(11)

Table S1: The number of human, macaque and mosquito samples in the occurrence database produced by the 2015 literature review, with samples in Indonesia and Malaysia shown stratified by region (province, state or territory). Total counts are shown for records from both the 2020 literature review and 2015 literature review.

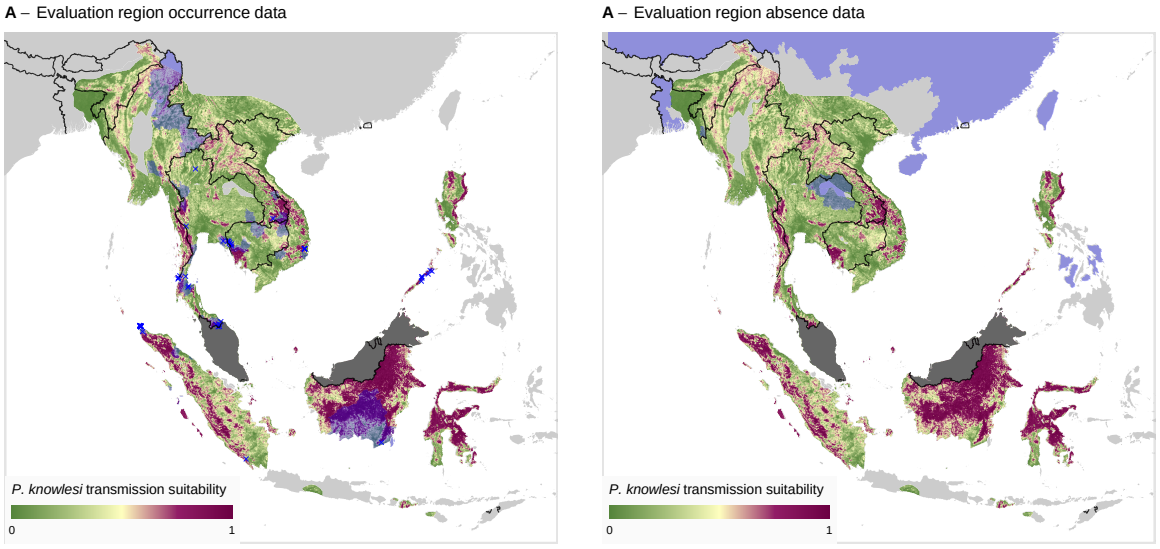

Figure S2: Predicted mean transmission suitability with overlay of infection occurrence data across the evaluation region. Results are only displayed where an area is in the range of both a vector and reservoir species necessary for transmission (see Methods). **A**: Predicted transmission suitability with infection occurrence polygons and points in blue. **B**: Predicted transmission suitability with infection absence polygons in blue.

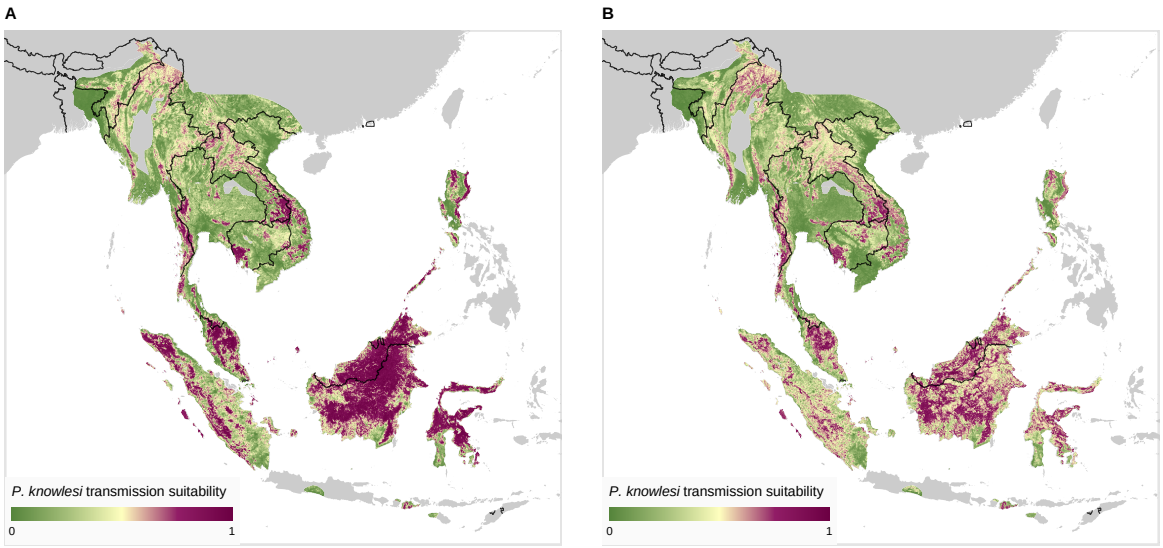

Figure S3: Comparison of the modelled mean transmission suitability value between the current work, with data up to 2020 (**A**), and the values presented in the 2015 model [31] (**B**). Note that the absolute value of predictions are not necessarily comparable given differences in model specification and training data. Results are only displayed where an area is in the range of both a vector and reservoir species necessary for transmission (see Methods).

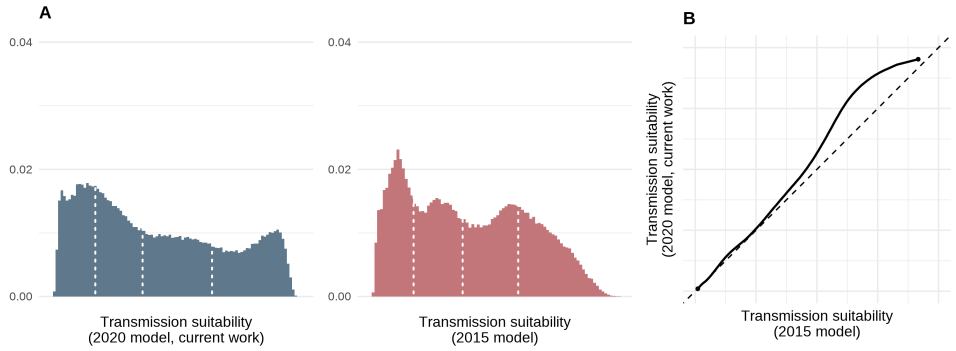

Figure S4: Histograms (**A**) and quantile-quantile plot (**B**) comparing the distributions of mean predicted transmission suitability for the 2015 and 2020 models of *P. knowlesi* transmission risk. Histograms are presented on a relative x-axis (ranging from minimal to maximal predicted mean risk), with quartiles of predicted risk displayed as dashed vertical lines.

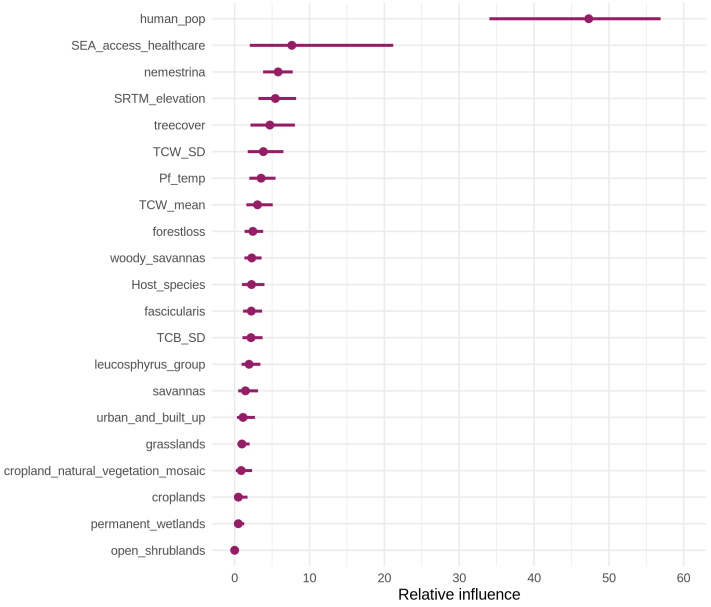

Figure S5: Relative influence scores for each covariate, calculated for each bootstrap, with points and lines representing median values and 95% confidence intervals respectively.

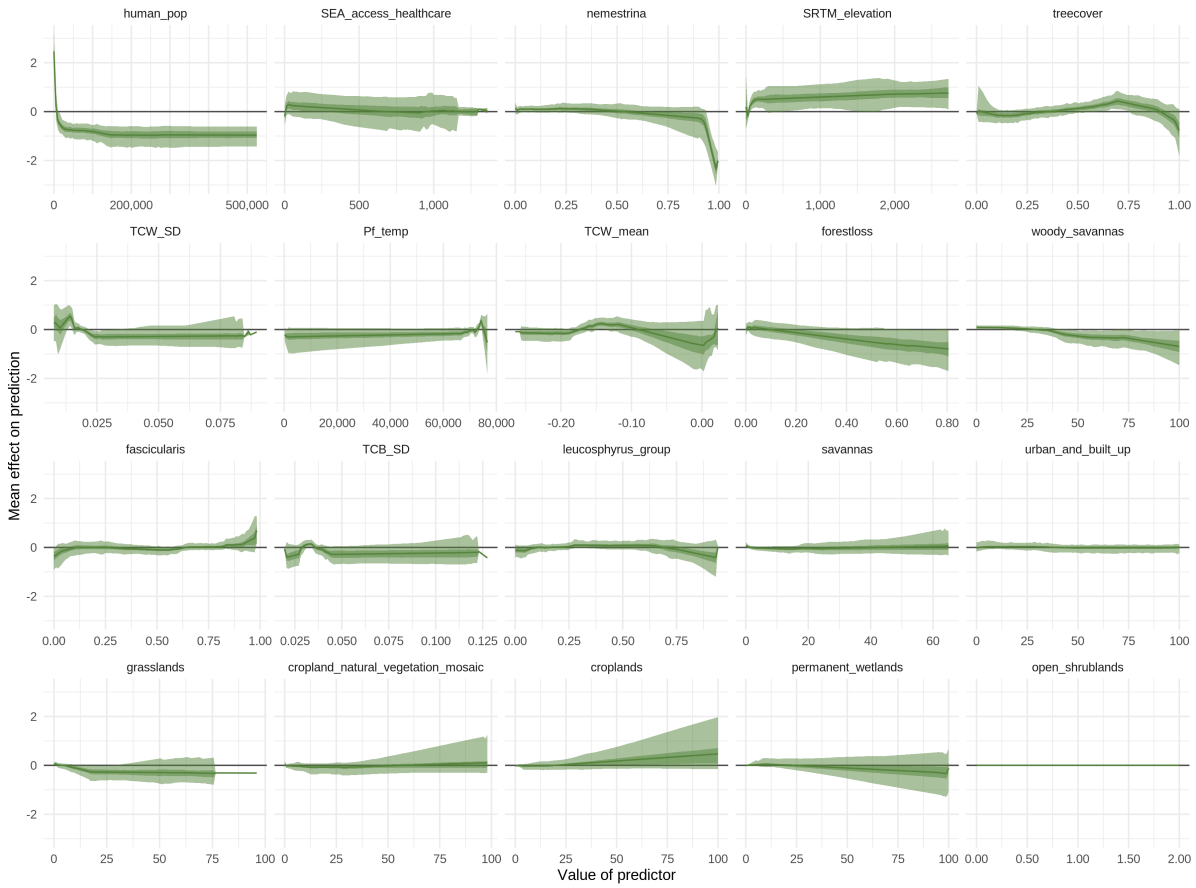

Figure S6: Accumulated local effect (ALE) plots for each covariate, indicating the mean effect of changing a covariate's value upon the prediction (on logistic scale) across the range of that covariate. The ALE values are calculated for each bootstrap, with the median value, 50% and 95% confidence intervals presented as lines, darker shaded regions and lighter shaded regions respectively.