Predicting the fundamental thermal niche of crop pests and diseases in a changing world: a case study on citrus greening.
Supplementary Information

Rachel A. Taylor[1,2,*], Sadie J. Ryan[3, 4, 5], Catherine A. Lippi[3,4], David G. Hall[6], Hossein A. Narouei-Khandan[4,7], Jason R. Rohr[1], and Leah R. Johnson[1,8]

[1]Department of Integrative Biology, University of South Florida, Tampa, FL, USA
[2]Department of Epidemiological Sciences, Animal and Plant Health Agency (APHA), Weybridge, UK
[3]Quantitative Disease Ecology and Conservation (QDEC) Lab, Department of Geography, University of Florida, Gainesville, FL 32601
[4]Emerging Pathogens Institute, University of Florida, Gainesville, FL 32610
[5]School of Life Sciences, University of KwaZulu, Natal, South Africa
[6]USDA-ARS, 2001 South Rock Road, Fort Pierce, FL 34945, USA
[7]Department of Plant Pathology, University of Florida, Gainesville, FL 32601
[8]Department of Statistics, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA 24061, USA
[*]rachel.taylor@apha.gov.uk

May 18, 2019

Appendix S1 Model for transmission of HLB

The model used is the same as in Taylor *et al.* (2016), which is a compartmental model of Huanglongbing (HLB) disease status with transmission between trees and psyllids. We outline the model equations here alongside explanation of the calculation of $S(T)$.

In our model, Equations (S.1)-(S.9), citrus trees are categorized as either Susceptible, $S(t)$, Asymptomatic, $A(t)$, or Infected, $I(t)$, in which Infected implies the disease is detectable by symptoms; we assume Asymptomatic and Infected trees transmit the pathogen with the same probability. Adult psyllids are Susceptible, $S_V(t)$; Exposed, $E_V(t)$; or Infected, $I_V(t)$; where Exposed indicates that the psyllids are infected but are not yet able to pass the disease on to another tree. Once infected, psyllids remain so for their entire lifespan. Successful transmission between psyllid and tree can only occur when psyllids feed off the phloem of the tree; the feeding rate is independent of grove size thus transmission is frequency-dependent. We assume well-mixing between trees and psyllids. A very small rate of natural death of susceptible and asymptomatic trees occurs and we include roguing of infected trees; together these trees are categorized as Removed, $R(t)$. However, removed trees are immediately replaced in the grove by susceptible trees, keeping the grove size constant. Thus, the removed category exists to keep track of how many trees have been removed and replaced over time; it does not represent alive trees in the grove. We assume that the grove has 100% susceptible trees initially, with psyllids feeding freely from the trees. At time 0, we introduce one infected tree.

$$\frac{dS}{dt} = -\frac{ab}{N} I_V S - rS + r(N - I) + r_1 I \tag{S.1}$$
$$\frac{dA}{dt} = \frac{ab}{N} I_V(t - \tau) S(t - \tau) e^{-r\tau} - \gamma A - rA \tag{S.2}$$
$$\frac{dI}{dt} = \gamma A - r_1 I \tag{S.3}$$
$$\frac{dR}{dt} = r(N - I) + r_1 I \tag{S.4}$$
$$\frac{dS_V}{dt} = \lambda F - \frac{ac}{N}(A + I) S_V - \mu S_V \tag{S.5}$$
$$\frac{dE_{V1}}{dt} = \frac{ac}{N}(A + I) S_V - 3\phi E_{V1} - \mu E_{V1} \tag{S.6}$$
$$\frac{dE_{V2}}{dt} = 3\phi E_{V1} - 3\phi E_{V2} - \mu E_{V2} \tag{S.7}$$
$$\frac{dE_{V3}}{dt} = 3\phi E_{V2} - 3\phi E_{V3} - \mu E_{V3} \tag{S.8}$$
$$\frac{dI_V}{dt} = 3\phi E_{V3} - \mu I_V. \tag{S.9}$$

N is the total number of trees in the grove, which is kept constant, and $V = S_V + E_{V1} + E_{V2} + E_{V3} + I_V$ is the total number of psyllids. We split the Exposed stage into three compartments to more accurately represent the length of the extrinsic incubation period. Following Lloyd (2001), using n compartments, in which the rate of leaving each compartment is $n\phi$, produces a Gamma distribution for overall psyllid progression to the infectious class with a mean rate of ϕ. The more compartments used leads to a Gamma distribution with lower variance around the mean. This is a useful alternative to fixed time delays, which can be problematic when parameters are temperature-dependent. We measure time in years so all rates are per year. a is the feeding rate of the psyllid on the trees, b is the probability that a susceptible tree becomes infected from contact with an infected psyllid, and c is the probability that a susceptible psyllid becomes infected from contact with an infected tree. Hence bc is the vector competence. We impose a time delay τ on trees moving from susceptible to asymptomatic state to represent the length of the incubation period when a tree is infected but not yet infectious. This time delay is long (approximately 6 months (Gottwald, 2010)), hence we use a fixed time delay of length τ to represent this more accurately than using a simple exponentially distributed exposure period (Kuang, 1993). r is the natural death rate of susceptible and asymptomatic trees. Trees that are exposed may not survive the exposure period due to natural death, thus we include a discount term $e^{-r\tau}$ to correctly model how many trees move from susceptible to asymptomatic. Asymptomatic trees develop symptoms and move to the infected class with rate γ. The rate of removal of

infectious trees by roguing is r_1. We assume all removed and dead trees are replanted with susceptible trees, hence the addition of these trees in equation (S.1).

Adult psyllids have a fixed birth rate λ, which includes the development of eggs and nymphs. We include a term F to represent flush seasons, which is when trees produce new leaves. Development of psyllids is very closely connected to availability of flush, as eggs are laid on flush and nymphs remain on the same flush for their development period. Thus, the birth rate λ is defined as the number of adult psyllids produced on a single flush, and F determines how many flush are in the grove over the year. The birth rate per flush patch, λ, can be expressed as the product of the number of eggs laid over a lifetime $\frac{F_E}{\mu}$ (where F_E is the number of eggs laid per female per year) and the probability the eggs survive to adulthood p_{EA}, over the average duration of the immature stages (Mordecai *et al.*, 2013). Hence,

$$\lambda = \frac{F_E p_{EA} D_P}{\mu} \tag{S.10}$$

where D_P is the vector development rate, i.e. 1/time for a vector to develop from egg to adult. The death rate of psyllids is μ. ϕ is the development rate of the bacteria within the psyllid, determining the length of their extrinsic incubation period.

The size of the vector population can be modeled as a function of the demographic parameters. The probability of having V vectors at time t tends to a Poisson distribution with mean $\frac{\lambda F}{\mu}$ (Parham & Michael, 2010).

Mathematical models of disease systems often use R_0, the basic reproductive number, as a measure of disease prevalence. It is a measure of how many secondary hosts will become infected if one initial host is infected in a naïve population. The equation for R_0 for Equations (S.1)-(S.9) is calculated using next-generation matrices as put forward in Diekmann & Heesterbeek (2000); Diekmann *et al.* (2009). This leads to the following equation:

$$R_0 = \left(\frac{F_E p_{EA} D_P a^2 b c F}{N \mu^3} \left(\frac{3\phi}{3\phi + \mu} \right)^3 e^{-r\tau} \left(\frac{1}{\gamma + r} + \frac{\gamma}{(\gamma + r) r_1} \right) \right)^{1/2}. \tag{S.11}$$

An interpretation of the terms involved in R_0 is given in the main text.

The focus of this paper is on the temperatures that promote and prohibit the spread of HLB. Therefore we consider only relative values of R_0 as temperature changes across regions. We assume that all other parameters related to the trees, such as average lifespan of a citrus tree or the length of time until symptoms appear, will be the same regardless of whether the tree is in Florida or Spain or anywhere else. Therefore, we focus on four parameters which we expect to vary with temperatures and for which there exists data across a sufficient

temperature range: fecundity (F_E); probability of egg to adult survival (p_{EA}); mortality rate (μ); and the development rate of psyllids from eggs into adults (D_P).

Appendix S2 Details of Bayesian Fitting

In this paper we use a Bayesian statistical approach to fitting the thermal responses of psyllid traits to data. Taking this approach allows a straightforward way to quantify the uncertainty in the data and also incorporate additional prior information into the analysis if needed. Further, this approach allows us to combine the traits and the uncertainty in fits to propagate the uncertainty into uncertainty in $S(T)$ overall, and partition the sources of uncertainty in $S(T)$ (Johnson *et al.*, 2015).

Most classical statistical analyses focus on the likelihood, $\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{Y})$ defined as how likely are the data to be obtained under a particular setting of the parameters, and interpreted as a function of the parameters. The Bayesian approach instead focuses on the posterior distribution of parameters given the observed data, $\Pr[\boldsymbol{\theta}|\boldsymbol{Y}]$, where $\boldsymbol{\theta}$ is a vector of model parameters and $\boldsymbol{Y}$ are the data. This posterior distribution is related to the likelihood of the data through Bayes Theorem:

$$\Pr[\boldsymbol{\theta}|\boldsymbol{Y}] \equiv \frac{\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{Y})\ \Pr[\boldsymbol{\theta}]}{\Pr[\boldsymbol{Y}]}$$

where $\Pr[\boldsymbol{Y}]$ is the probability of the data and $\Pr[\boldsymbol{\theta}]$ is the prior probability of parameters. In practice the denominator is not known, so numerical methods, such as Markov Chain Monte Carlo (MCMC), are used to approximate the posterior distribution. Further details on Bayesian approaches in ecological contexts can be found in Clark (2007).

Appendix S2.1 Specification of Likelihoods and Priors

Our methods follow closely those of Johnson *et al.* (2015). For each thermal performance trait, we choose a parametric unimodal functional response as the mean function. We then specify the error distribution around this mean appropriate to the data. More specifically, we use a truncated normal distribution for parameters p_{EA}, D_P and $1/\mu$, and a Poisson distribution for F_E (both Liu & Tsai data and Hall data). This error distribution together with the mean function defines the likelihood for the data. We then set priors that represent biologically reasonable limits on the values of parameters but otherwise are uninformative (See Table S2.1).

Model Parameter	Mean Function	Parameters	Prior
Prob Egg Adult Survival	Quadratic	T_0	dunif(0, 20)
P_{EA}		T_M	dunif(30,50)
		qn	dgamma(10,1)
		τ	dgamma(0.01,0.01)
Longevity	Quadratic	T_0	dunif(-10, 20)
$1/\mu$		T_M	dunif(30,50)
		qn	dgamma(10,1)
		τ	dgamma(50,5)
Development rate	Brière	T_0	dunif(0, 20)
D_P		T_M	dunif(30,50)
		c	dgamma(10,1000)
		τ	dgamma(1,0.01)
Fecundity	Brière	T_0	dunif(0, 25)
F_E		T_M	dunif(30,50)
(Liu & Tsai, 2000)		c	dgamma(1,10)
Fecundity	Quadratic	T_0	dunif(0, 25)
F_E		T_M	dunif(30,50)
(Hall *et al.*, 2011)		qn	dgamma(10,1)

Table S2.1: Prior distributions for each of the parameters for the best fitting of the responses for each of the thermal traits considered.

Appendix S2.2 MCMC implementation and assessment of convergence

All analyses were conducted in the R programming language (R Development Core Team, 2008). MCMC samplers were implemented using the `JAGS/rjags` packages (Plummer, 2003, 2013). Details of the general MCMC approach are available elsewhere (e.g. Clark, 2007). For each trait, we initialized 3 independent chains and collected 50,000 adaptive samples during the tuning/burn-in phase. We assessed convergence using these samples, and, if the chains had not converged, increased the number of samples until convergence was achieved. We then collected an additional 5000 samples after convergence for each of the 3 chains. This resulted in a total of 15000 samples for each trait. All of these samples were used to calculate the posterior distributions for each thermal trait. We then took a random sub-sample of the parameters for each trait to use in the calculation of the posterior distribution of R_0 and summaries of the posterior. The subsample was of size 3000, chosen to allow computational tractability while including enough samples to assure good representation of the posterior.

Appendix S3 Full posterior results for all thermal traits

Appendix S3.1 Probability of egg to adult survival, p_{EA}

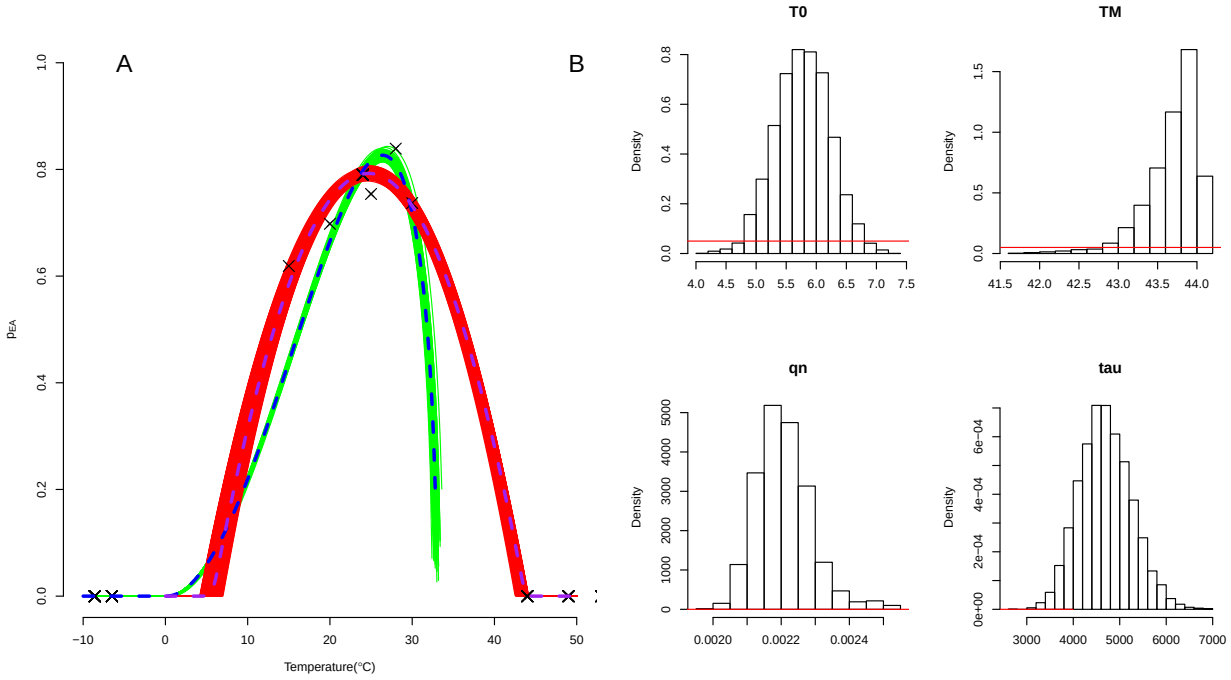

Figure S3.1: (A) The quadratic (red) and Brière (green) fits of p_{EA} data against temperature using vague priors. The quadratic is the better fit according to DIC. Data points are plotted in black and dashed lines indicate the mean trajectory. (B) Histograms of draws from the posterior distribution for each parameter of the quadratic fit. In B, the prior distribution for each parameter is plotted in red. The quadratic fit is determined by the equation $q_n(T - T0)(T - TM)$ using a normal distribution with precision τ.

Appendix S3.2 Development rate, D_P

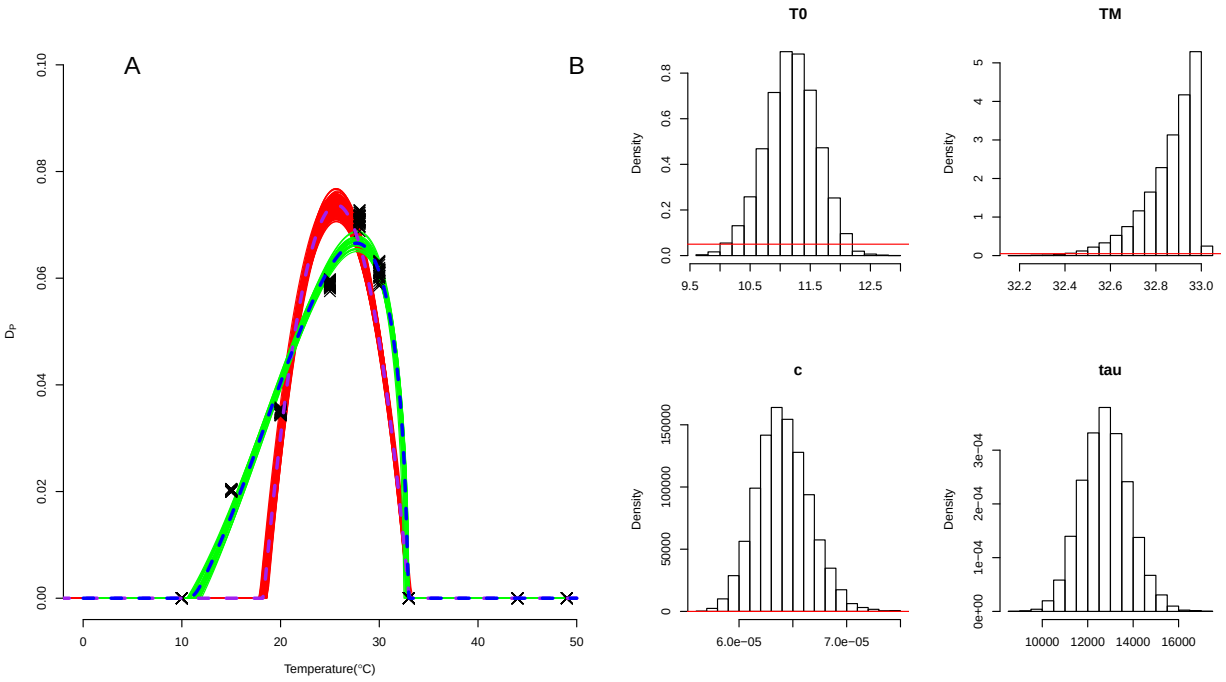

Figure S3.2: (A) The quadratic (red) and Brière (green) trajectories of D_P against temperature using vague priors. The Brière is the better fit according to DIC. Data points are plotted in black and dashed lines indicate the mean trajectory. (B) Histograms of draws from the posterior distribution for each parameter of the Brière fit. In B, the prior distribution for each parameter is plotted in red. The Brière fit is determined by the equation $cT(T-T0)(T-TM)^{1/2}$ using a normal distribution with precision τ.

Appendix S3.3 Adult psyllid longevity, $1/\mu$

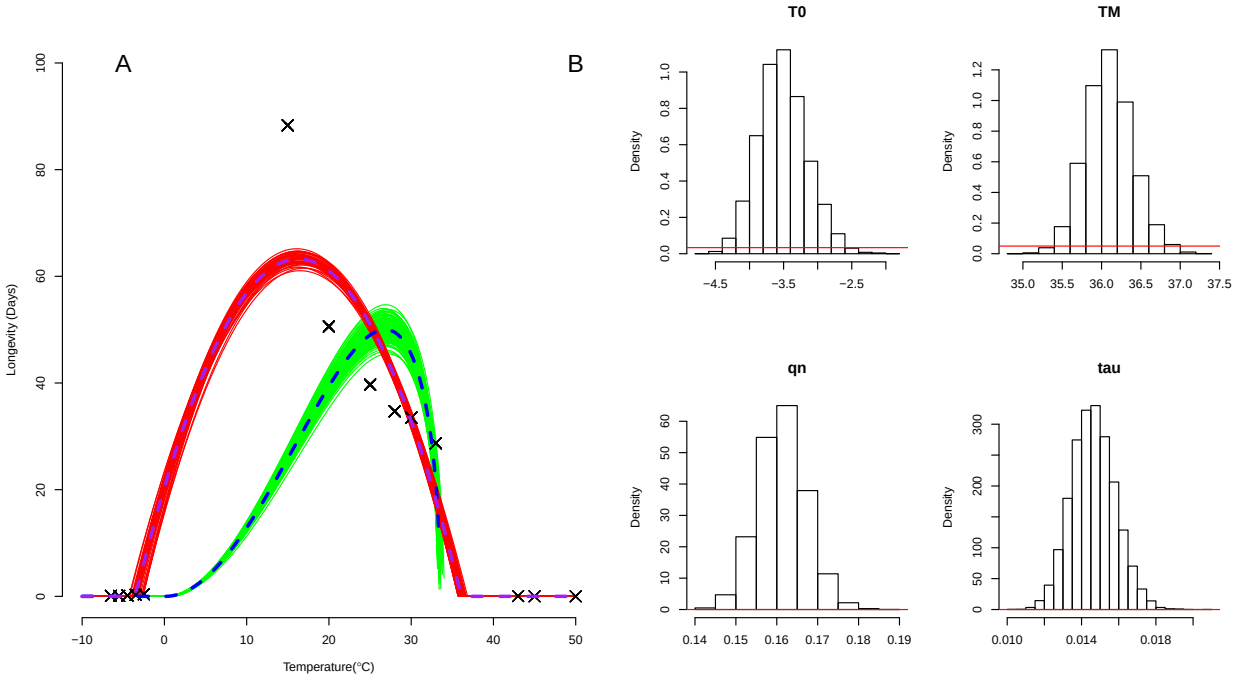

Figure S3.3: (A) The quadratic (red) and Brière (green) trajectories of $1/\mu$ against temperature using vague priors. The quadratic is the better fit according to DIC. Data points are plotted in black and dashed lines indicate the mean trajectory. (B) Histograms of draws from the posterior distribution for each parameter of the quadratic fit. The prior distribution for each parameter is plotted in red. The quadratic fit is determined by the equation $q_n(T - T0)(T - TM)$ using a normal distribution with precision τ.

Appendix S3.4 Fecundity, F_E, from Liu & Tsai (2000)

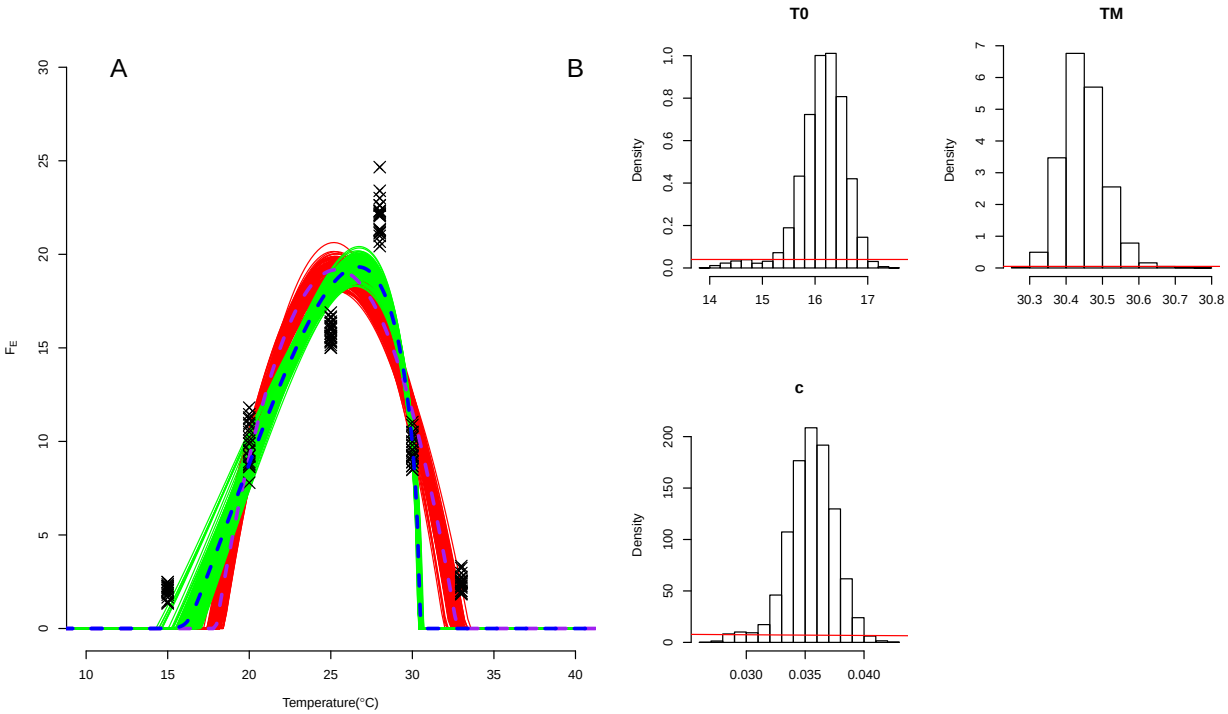

Figure S3.4: (A) The quadratic (red) and Brière (green) trajectories of LT00 F_E against temperature using vague priors. The Brière is the better fit according to DIC. Data points are plotted in black and dashed lines indicate the mean trajectory. (B) Histograms of draws from the posterior distribution for each parameter of the Brière fit. In B, the prior distribution for each parameter is plotted in red. The Brière fit is determined by the equation $cT(T - T0)(T - TM)^{1/2}$ using a Poisson distribution.

Appendix S3.5 Fecundity, F_E, from Hall *et al.* (2011)

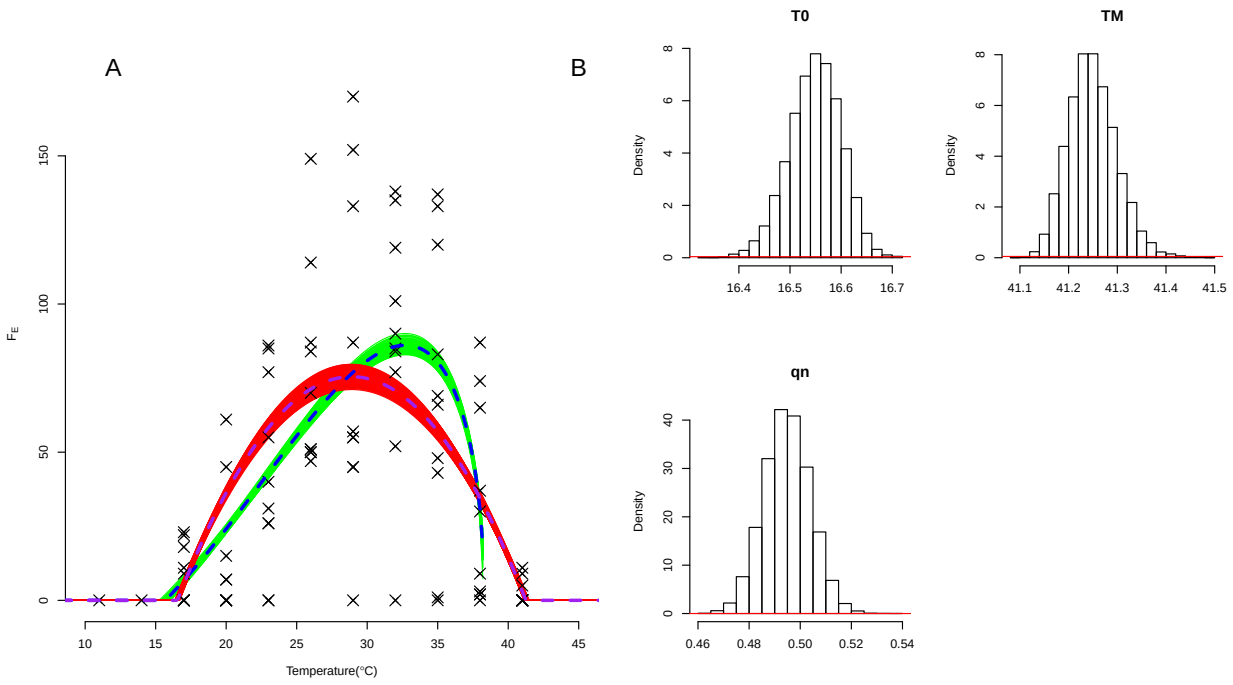

Figure S3.5: (A) The quadratic (red) and Brière (green) trajectories of $1/\mu$ against temperature using vague priors. The quadratic is the better fit according to DIC. Data points are plotted in black and dashed lines indicate the mean trajectory. (B) Histograms of draws from the posterior distribution for each parameter of the quadratic fit. The prior distribution for each parameter is plotted in red. The quadratic fit is determined by the equation $q_n(T - T0)(T - TM)$ using a Poisson distribution.

Appendix S4 Further results from the posterior distribution of $S(T)$

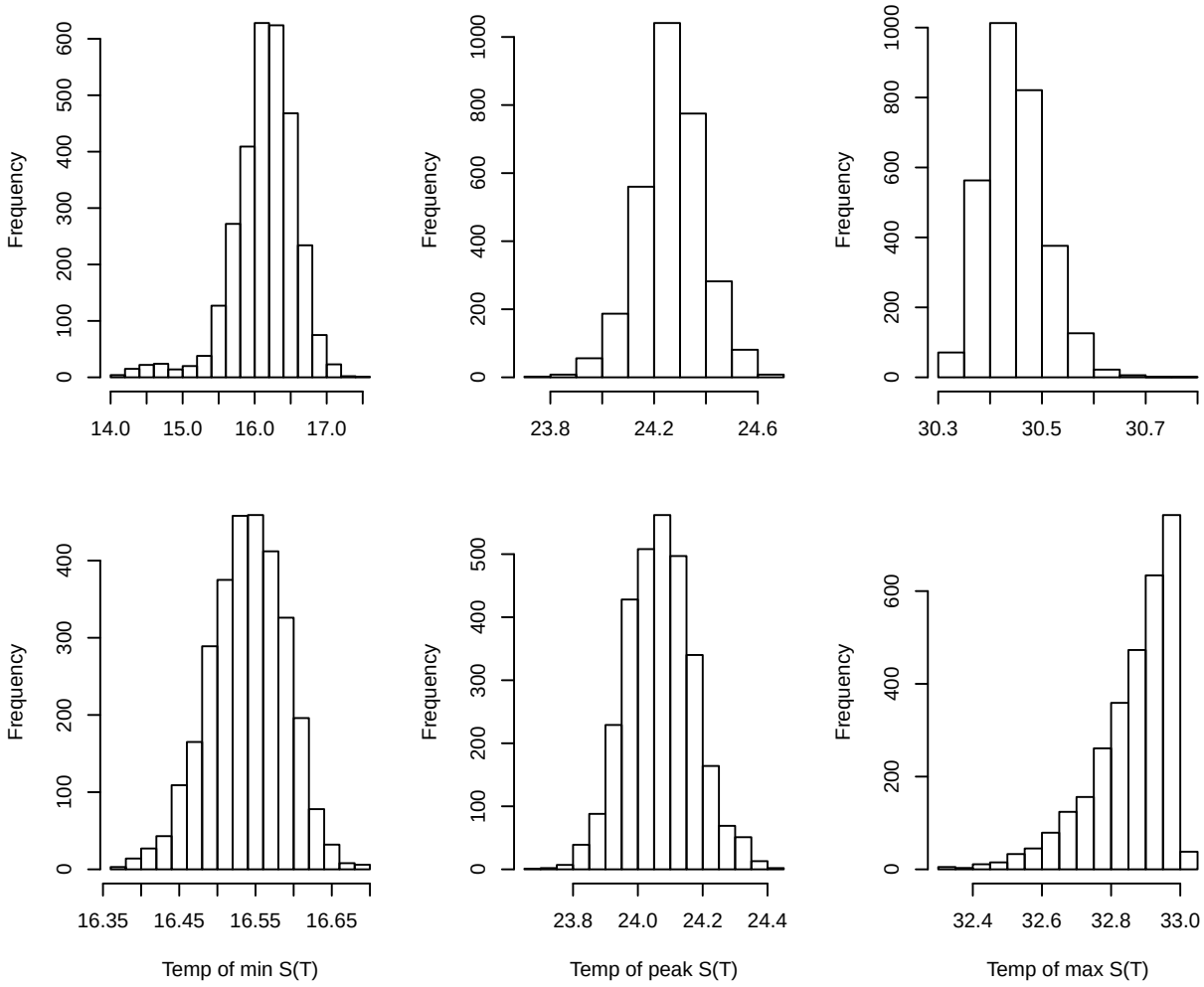

Figure S4.1: Posterior distributions of the lower temperature limit, peak temperature and the upper temperature limit of $S(T)$. Top row: Liu & Tsai (2000), bottom row: Hall *et al.* (2011).

Appendix S5 Quantiles of the suitability metric $S(T)$

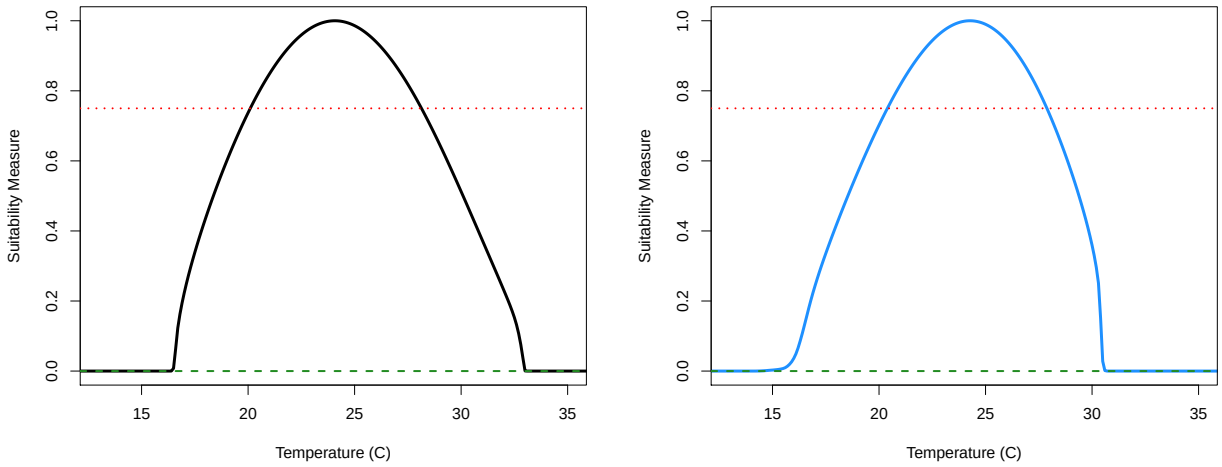

Figure S5.1: The suitability metrics $S(T)$ for Liu & Tsai (2000) (left) and Hall *et al.* (2011) (right). The quantiles indicating where $S(T) > 0$, for permissive suitability, and where $S(T) > 0.75$, for high suitability, are plotted with green dashed and purple dotted lines respectively.

Appendix S6 Full Validation Results

Appendix S6.1 Data on HLB presence

Non-mountainous regions, only.

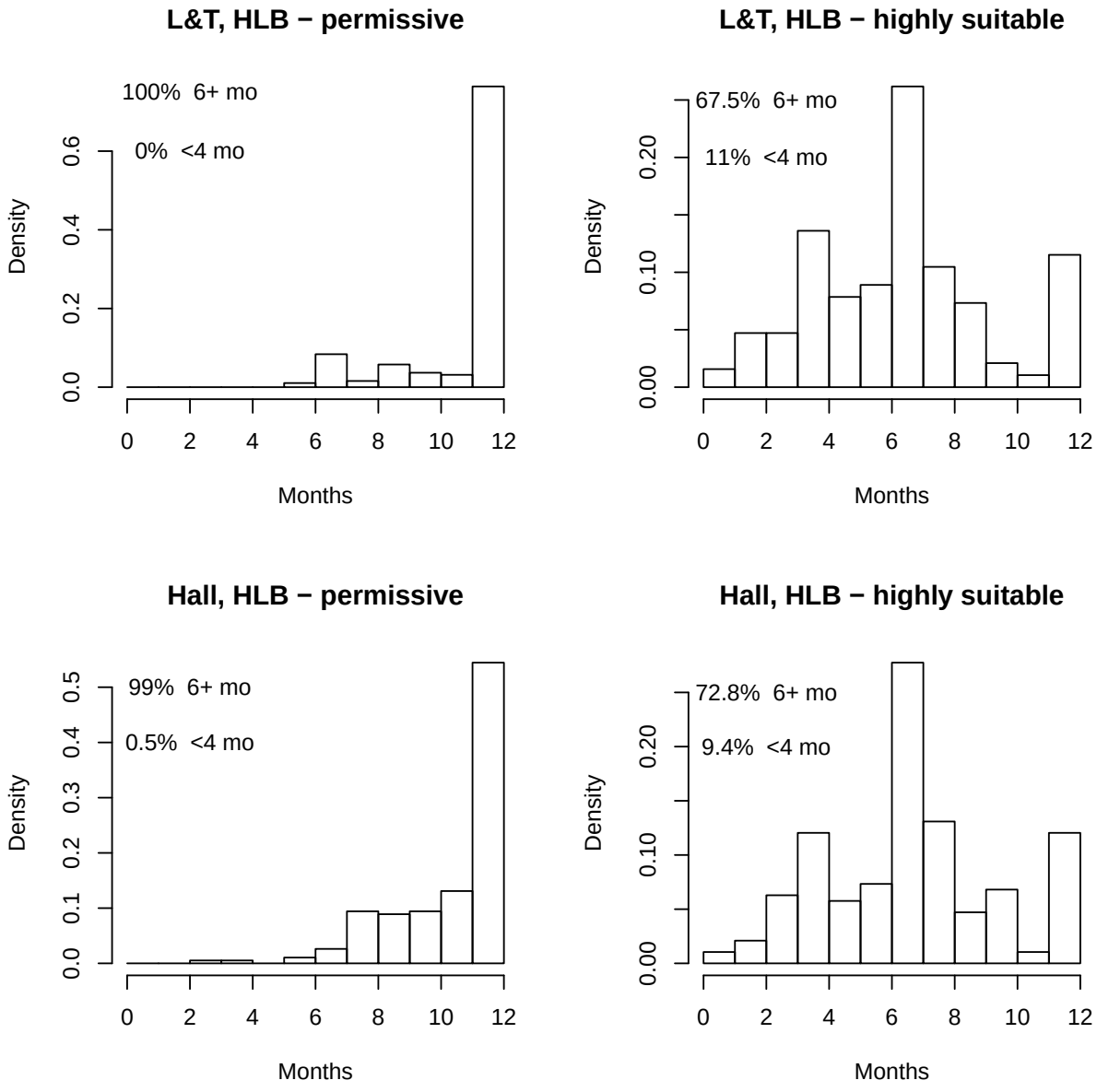

Figure S6.1: The number of months that every location with current HLB presence that is either permissive or highly suitable, **excluding mountainous locations**. We define permissive suitability as $S(T) > 0$ and high suitability as $S(T) > 0.75$. Top row: Predictions based on the model built with the Liu & Tsai data. Bottom row: Predictions based on the model built with the Hall data.

All locations.

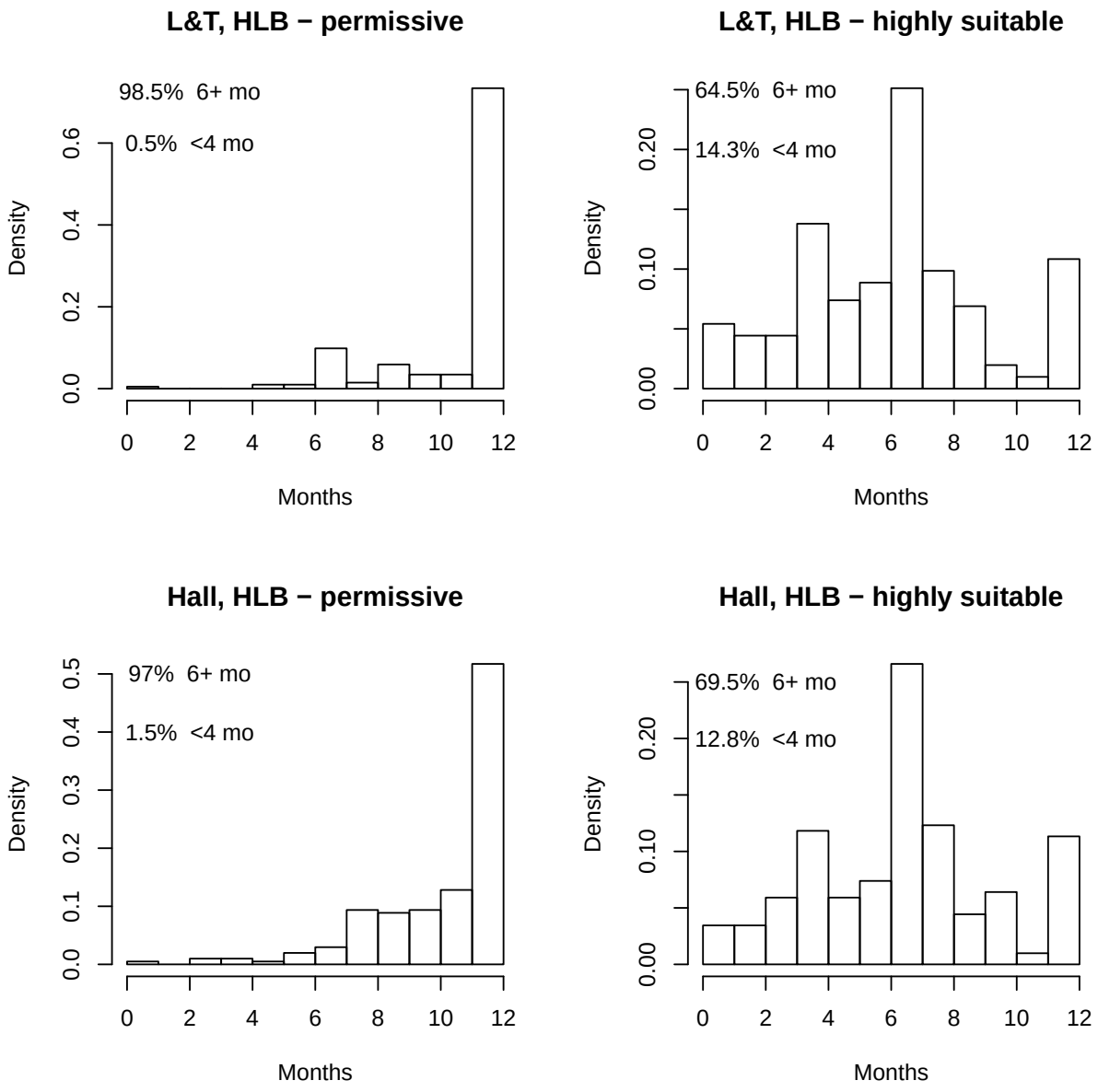

Figure S6.2: The number of months that every location with current HLB presence that is either permissive or highly suitable, **including mountainous locations**. We define permissive suitability as $S(T) > 0$ and high suitability as $S(T) > 0.75$. Top row: Predictions based on the model built with the Liu & Tsai data. Bottom row: Predictions based on the model built with the Hall data.

Appendix S6.2 Data on Asian Citrus Psyllid (ACP) presence

Non-mountainous regions, only.

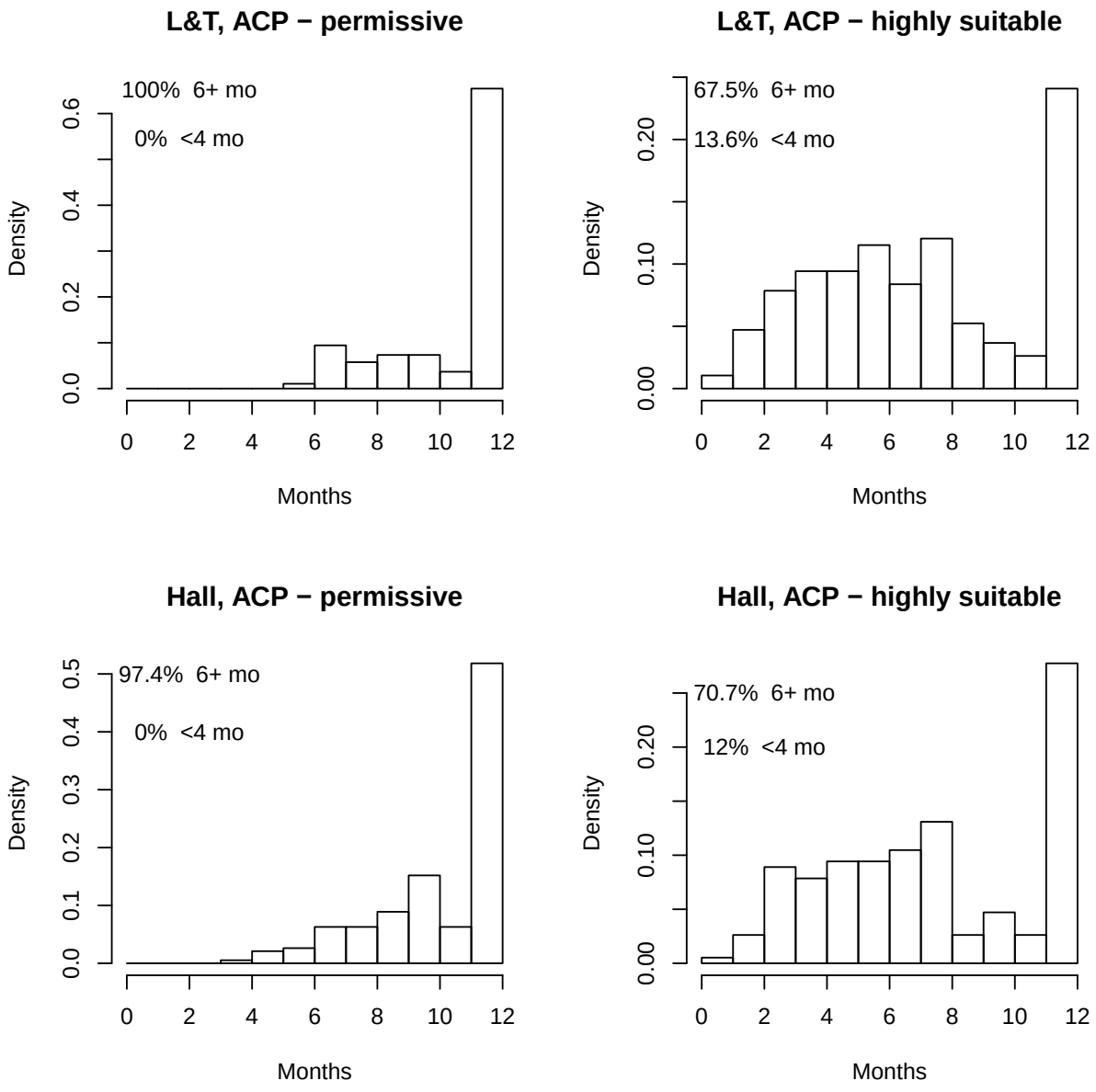

Figure S6.3: The number of months that every location with current ACP presence that is either permissive or highly suitable, **excluding mountainous locations**. We define permissive suitability as $S(T) > 0$ and high suitability as $S(T) > 0.75$. Top row: Predictions based on the model built with the Liu & Tsai data. Bottom row: Predictions based on the model built with the Hall data.

All locations.

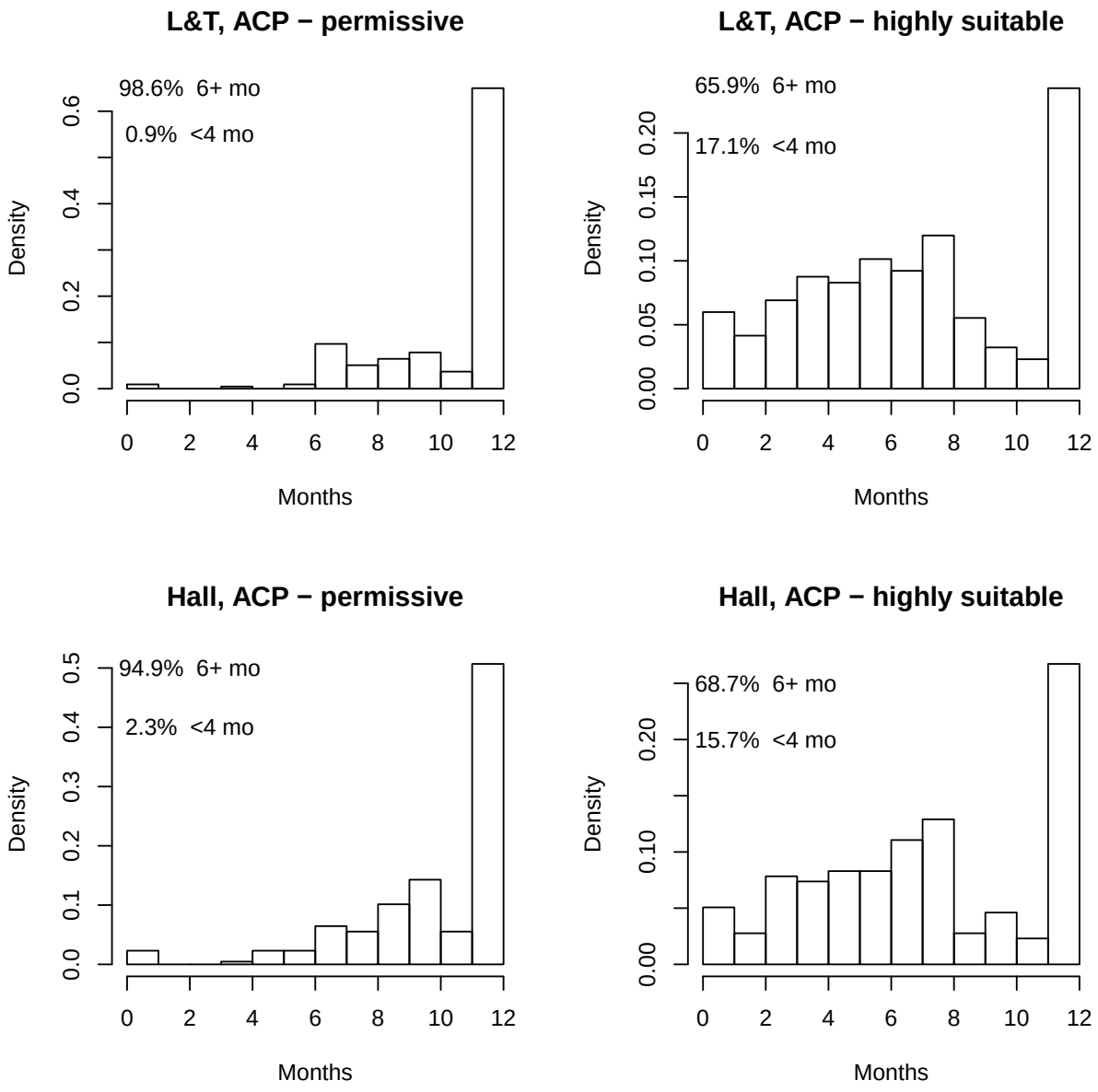

Figure S6.4: The number of months that every location with current ACP presence that is either permissive or highly suitable, **including mountainous locations**. We define permissive suitability as $S(T) > 0$ and high suitability as $S(T) > 0.75$. Top row: Predictions based on the model built with the Liu & Tsai data. Bottom row: Predictions based on the model built with the Hall data.

Appendix S7 Liu & Tsai (2000) model maps

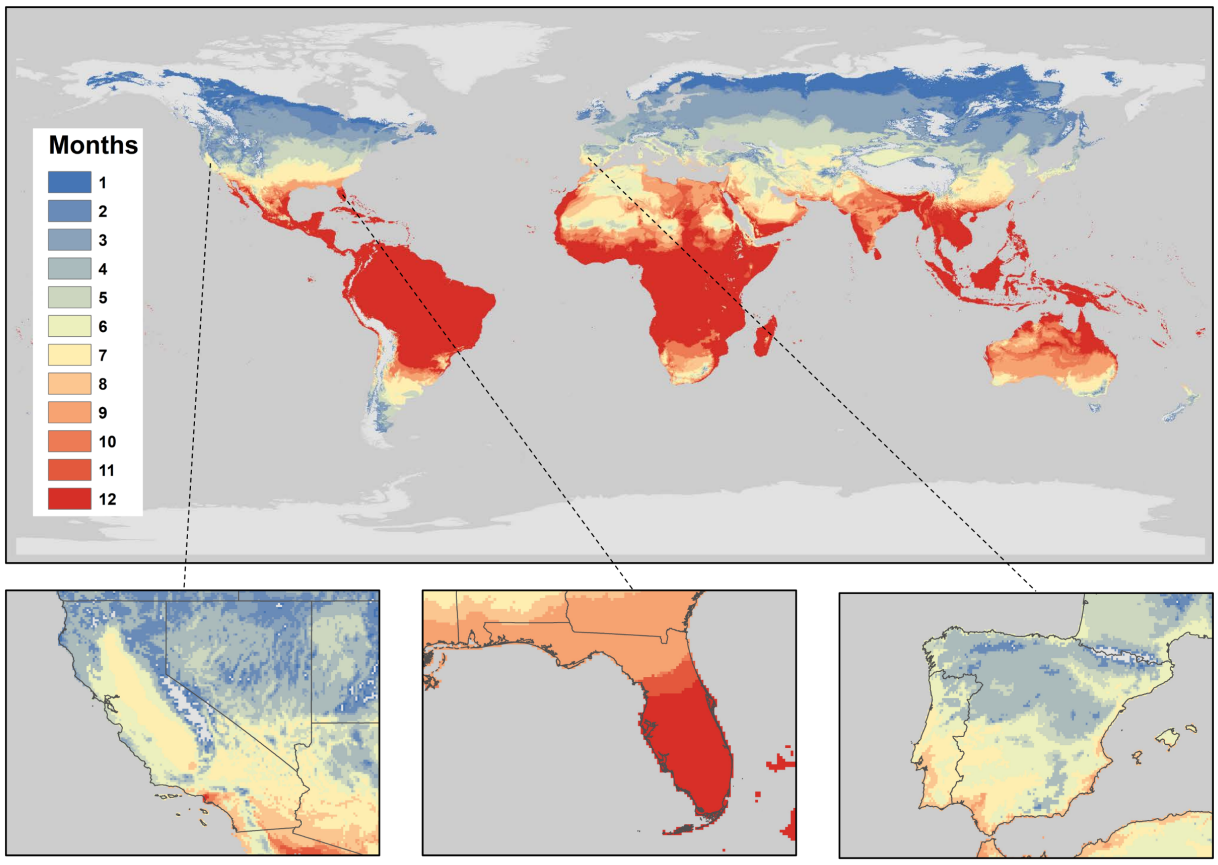

Figure S7.1: The number of months a year that locations have permissive temperatures according to the Liu & Tsai (2000) model. Inset plots of California, Florida and the Iberian peninsula, respectively, are included. We define permissive temperatures for suitability as $S(T) > 0$. Locations in grey have zero months suitable for HLB transmission.

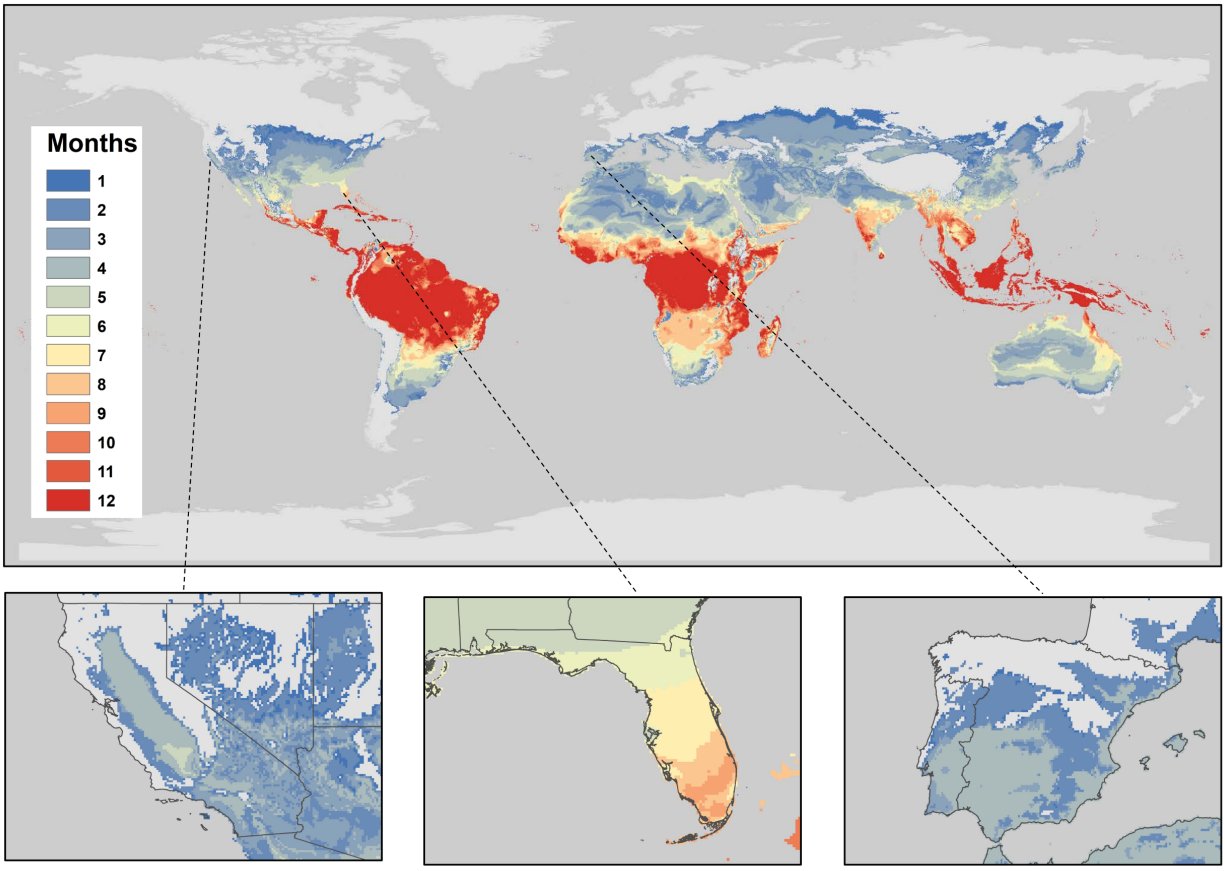

Figure S7.2: The number of months a year that locations have highly suitable temperatures according to the Liu & Tsai (2000) model. Inset plots of California, Florida and the Iberian peninsula, respectively, are included. We define highly suitable temperatures as $S(T) > 0.75$. Locations in grey have zero months suitable for HLB transmission.

Appendix S8 Thermal Suitability Maps with Validation Points

We plot the permissive suitability maps for both Liu & Tsai (2000) and Hall *et al.* (2011) with first the HLB occurrence locations and then the ACP occurrence locations from Narouei-Khandan *et al.* (2016) indicated.

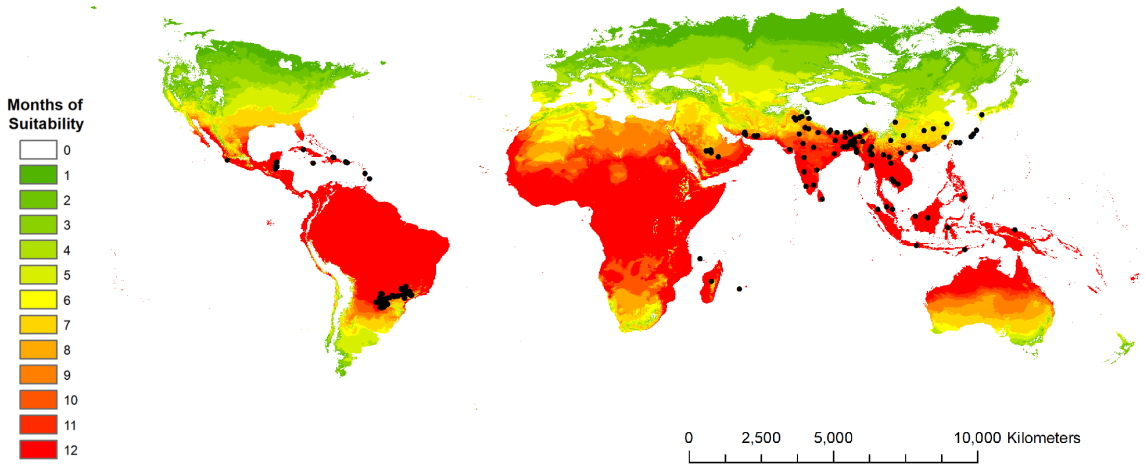

Figure S8.1: The number of months a year that locations have permissive temperatures according to the Hall *et al.* (2011) model. Black dots represent confirmed locations of HLB. We define permissive temperatures for suitability as $S(T) > 0$.

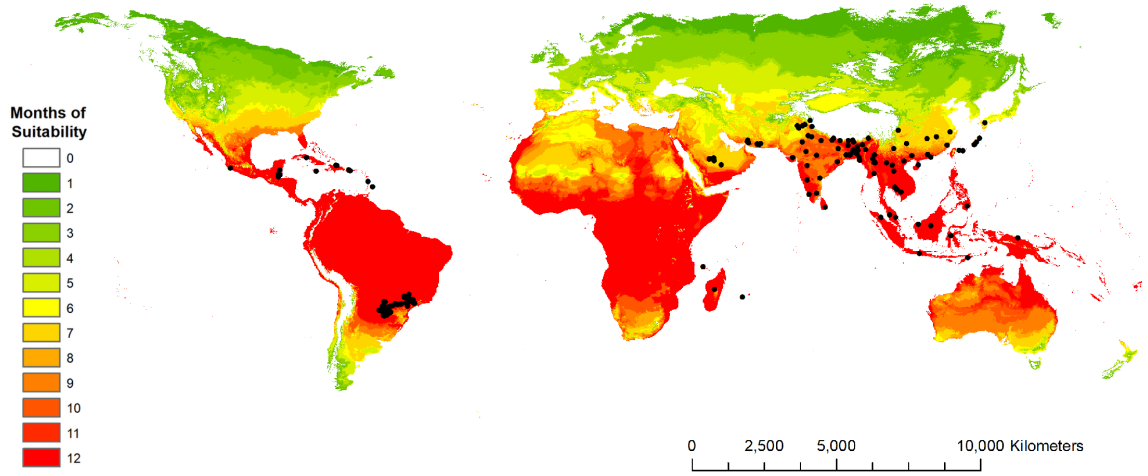

Figure S8.2: The number of months a year that locations have permissive temperatures according to the Liu & Tsai (2000) model. Black dots represent confirmed locations of HLB. We define permissive temperatures for suitability as $S(T) > 0$.

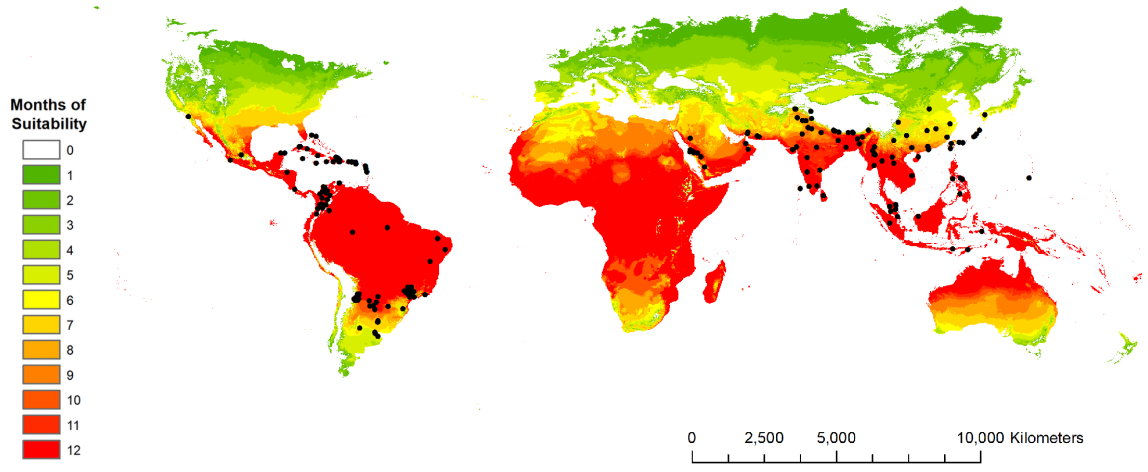

Figure S8.3: The number of months a year that locations have permissive temperatures according to the Hall *et al.* (2011) model. Black dots represent confirmed locations of ACP. We define permissive temperatures for suitability as $S(T) > 0$.

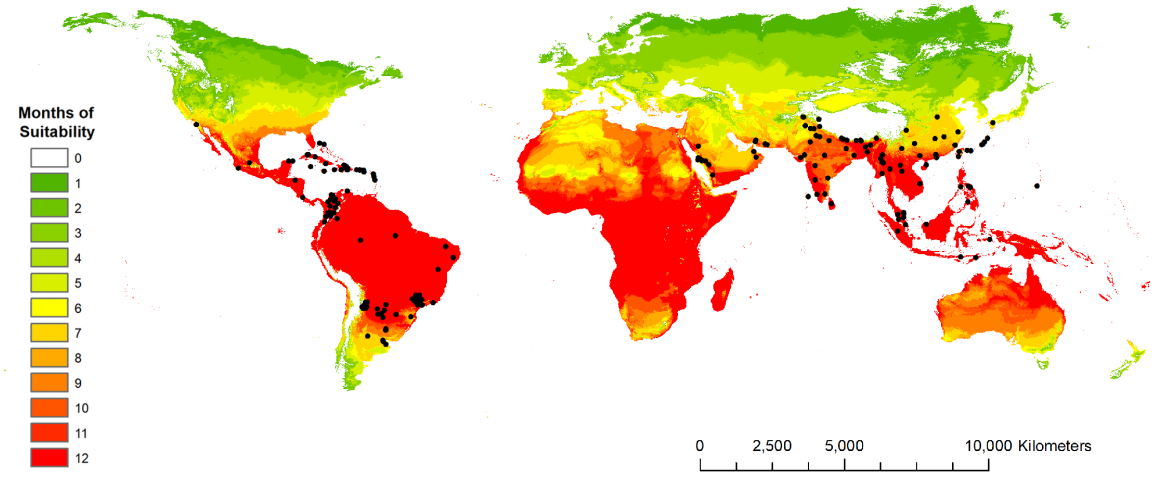

Figure S8.4: The number of months a year that locations have permissive temperatures according to the Liu & Tsai (2000) model. Black dots represent confirmed locations of ACP. We define permissive temperatures for suitability as $S(T) > 0$.

References

Clark, J.S. (2007). *Models for Ecological Data*. Princeton University Press, Princeton, NJ.

Diekmann, O. & Heesterbeek, J.A.P. (2000). *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. vol. 5. John Wiley & Sons.

Diekmann, O., Heesterbeek, J.A.P. & Roberts, M.G. (2009). The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface*, p. rsif20090386.

Gottwald, T.R. (2010). Current epidemiological understanding of citrus Huanglongbing. *Annual Review of Phytopathology*, 48, 119–139.

Hall, D.G., Wenninger, E.J. & Hentz, M.G. (2011). Temperature studies with the Asian citrus psyllid, *Diaphorina citri*: Cold hardiness and temperature thresholds for oviposition. *Journal of Insect Science*, 11, 83.

Johnson, L.R., Ben-Horin, T., Lafferty, K.D., McNally, A., Mordecai, E., Paaijmans, K.P., Pawar, S. & Ryan, S.J. (2015). Understanding uncertainty in temperature effects on vector-borne disease: a Bayesian approach. *Ecology*, 96, 203–213.

Kuang, Y. (1993). *Delay differential equations: with applications in population dynamics*. vol. 191. Academic press.

Liu, Y.H. & Tsai, J.H. (2000). Effects of temperature on biology and life table parameters of the Asian citrus psyllid, *Diaphorina citri* Kuwayama (Homoptera: Psyllidae). *Annals of Applied Biology*, 137, 201–206.

Lloyd, A.L. (2001). Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theoretical Population Biology*, 60, 59–71.

Mordecai, E.A., Paaijmans, K.P., Johnson, L.R., Balzer, C., Ben-Horin, T., Moor, E., McNally, A., Pawar, S., Ryan, S.J., Smith, T.C. & Lafferty, K.D. (2013). Optimal temperature for malaria transmission is dramatically lower than previously predicted. *Ecology Letters*, 16, 22–30.

Narouei-Khandan, H.A., Halbert, S.E., Worner, S.P. & van Bruggen, A.H.C. (2016). Global climate suitability of citrus Huanglongbing and its vector, the Asian citrus psyllid, using two correlative species distribution modeling approaches, with emphasis on the USA. *European Journal of Plant Pathology*, 144, 655–670.

Parham, P.E. & Michael, E. (2010). Modeling the effects of weather and climate change on malaria transmission. *Environmental Health Perspectives*, 118, 620.

Plummer, M. (2003). Jags: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March*. pp. 20–22.

Plummer, M. (2013). `rjags`*: Bayesian graphical models using MCMC*. R package version 3.10.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Taylor, R.A., Mordecai, E.A., Gilligan, C.A., Rohr, J.R. & Johnson, L.R. (2016). Mathematical models are a powerful method to understand and control the spread of Huanglongbing. *PeerJ*, 4:e2642.