

1 **Title: Many unreported crop pests and pathogens are probably already present**

2 Running title: Crop pest and pathogen distributions

3

4 Daniel P. Bebber^{1*}, Elsa Field², Gui Heng³, Peter Mortimer³, Timothy Holmes⁴, Sarah J. Gurr¹

5

6 ¹ Department of Biosciences, University of Exeter, Stocker Road, Exeter, EX4 4QD, UK

7 ² Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

8 ³ Kunming Institute of Botany, Chinese Academy of Sciences, 132 Lanhei Road, Kunming 650201,
9 People's Republic of China

10 ⁴ CABI, Nosworthy Way, Wallingford, OX10 8DE, UK

11 * Corresponding author, email d.bebber@exeter.ac.uk, tel. +44 1392 725851

12

13 **Summary**

14 Biotic invasions threaten global biodiversity and ecosystem function. Such incursions present
15 challenges to agriculture where invasive pest species cause significant production losses require major
16 economic investment to control and can cause significant production losses. Pest Risk Analysis (PRA)
17 is key to prioritizing agricultural biosecurity efforts, but is hampered by incomplete knowledge of
18 current crop pest and pathogen distributions. Here we develop predictive models of current pest
19 distributions and test these models using new observations at sub-national resolution. We apply
20 generalized linear models (GLM) to estimate presence probabilities for 1739 crop pests in the CABI
21 pest distribution database. We test model predictions for 100 unobserved pest occurrences in the
22 People's Republic of China (PRC), against observations of these pests abstracted from the Chinese
23 literature. This resource has hitherto been omitted from databases on global pest distributions. Finally,
24 we predict occurrences of all unobserved pests globally. Presence probability increases with host
25 presence, presence in neighbouring regions, per capita GDP, and global prevalence. Presence
26 probability decreases with mean distance from coast and known host number per pest. The models were
27 good predictors of pest presence in Provinces of the PRC, with area under the ROC curve (AUC) values
28 of 0.75 – 0.76. Large numbers of currently unobserved, but probably present pests (defined here as
29 unreported pests with a predicted presence probability > 0.75), are predicted in China, India, southern
30 Brazil and some countries of the former USSR. Our results shows that GLMs can predict presences of
31 pseudo-absent pests at sub-national resolution. The Chinese scientific literature has been largely
32 inaccessible to Western academia but contains important information that can support PRA. Prior
33 studies have often assumed that unreported pests in a global distribution database represents a true
34 absence. Our analysis provides a method for quantifying pseudo-absences to enable improved PRA and
35 species distribution modelling.

36

37 **Keywords**

38 biogeography, crop pathogens, crop pests, Generalized Linear Model, observational bias, Pest Risk
39 Analysis, pseudo-absence, species distribution model.
40

41 **Introduction**

42 The spread of invasive species is homogenizing the biosphere, with wide-ranging implications for
43 natural ecosystems (Baiser *et al.*, 2012; Santini *et al.*, 2013) and agriculture (Fisher *et al.*, 2012; Bebber
44 *et al.*, 2014a; Bebber, 2015). The number of first observations of crop pests and pathogens has
45 accelerated in recent years, driven primarily by global trade (Ding *et al.*, 2008; Bacon *et al.*, 2014), but
46 also potentially by climate change and our improving ability to monitor and identify threats (Bebber *et al.*,
47 2014a; Bebber, 2015). Here, we use the term ‘pest’ to describe any herbivorous arthropod,
48 pathogenic microbe or virus known to attack agricultural crops. Emerging pests can be extremely
49 damaging to agricultural production and the economy, causing both pre-harvest and post-harvest losses
50 (Bebber & Gurr, 2015; Paini *et al.*, 2016; Savary *et al.*, 2017). Recently, for example, sub-Saharan
51 Africa has suffered from the virulent Ug99 strain of the wheat stem rust fungus (*Puccinia graminis*
52 *tritici*) (Patpour *et al.*, 2015), the newly-evolved Maize Lethal Necrosis viral syndrome (Wangai *et al.*,
53 2012), arrival of the Asian citrus psyllid (*Diaphorina citri*) which vectors citrus greening disease
54 (Shimwela *et al.*, 2016), and the appearance of Tropical Race 4 of *Fusarium oxysporum f. sp. cubense*
55 attacking Cavendish bananas (Ordonez *et al.*, 2015). Central America, Europe, East Africa and
56 Australia have been identified as hotspots of new pest invasions, with maize, bananas, citrus and potato
57 as the crops most likely to be affected (Bebber, 2015). Outbreaks of resident pests due to favourable
58 weather conditions, virulence evolution, or crop management factors, add to the burden on farmers. For
59 example, a major outbreak of coffee leaf rust (*Hemileia vastatrix*) in Latin America, likely to have been
60 triggered by a failure in disease management, is reported to have caused large-scale unemployment and
61 social upheaval in recent years (Avelino *et al.*, 2015).

62
63 Despite the expanding ranges of many pests, complete occupation of their potential ranges has not yet
64 occurred (Bebber *et al.*, 2014a) and so there remains a strong impetus for implementation of biosecurity
65 measures at international borders (Fears *et al.*, 2014; Flood & Day, 2016; MacLeod *et al.*, 2016).
66 Control of spread within countries is extremely difficult because of unhindered transport of plants and
67 soils (Ward, 2016), and biosecurity measures focus on quarantine and inspections at borders (MacLeod
68 *et al.*, 2016). A key component of international phytosanitary action is Pest Risk Analysis (PRA), a
69 suite of methods that allow countries to prioritize protective measures against those pests most likely to
70 arrive and cause serious economic damage (Robinet *et al.*, 2012; Baker *et al.*, 2014). PRA involves
71 assessment of the likelihood of pest arrival, the likelihood of establishment, the potential economic
72 impact if uncontrolled, and the prospect of successful control or eradication (Baker *et al.*, 2014). To
73 date, PRA has primarily been based upon expert opinion regarding the likelihood of arrival and potential
74 impact of individual pests. For example, the UK’s recently-established Plant Health Risk Register
75 (PHRR) (Baker *et al.*, 2014) employs simple climate-matching (based on known pest distributions) and
76 host availability to assign qualitative risks of invasion and impact, but not quantitative predictive
77 models. Examples of registered pests include the Oleander aphid *Aphis nerii* which has been assigned

78 very low likelihoods of arrival and establishment, and would cause negligible damage if it did, whereas
79 the Zebra chip phytoplasma *Candidatus liberibacter solanacearum* is thought moderately likely to
80 arrive and would have a very serious impact if it did (DEFRA, 2018).

81

82 While quantitative PRA protocols have been recently developed recently by the European Food Safety
83 Authority (Jeger *et al.*, 2018), examples of quantitative PRA are rare in international phytosanitary
84 legislation and practice. This contrasts with the long and vibrant history of research in predictive species
85 distribution modelling (SDM) for pests (Elith & Leathwick, 2009; Sutherst, 2014). The geographic
86 distributions of species are non-random, determined by their biotic environment (e.g. hosts or prey), the
87 abiotic environment (e.g. climate, edaphic factors), and migration (dispersal to suitable habitat)
88 (Soberón & Peterson, 2005; Soberón, 2007; Soberón & Nakamura, 2009). Thus, pest invasion risk is,
89 in theory, quantifiable. Numerous modelling approaches are now available to predict the likely
90 distributions and impacts of pests (Elith & Leathwick, 2009; Venette *et al.*, 2010; Robinet *et al.*, 2012),
91 ranging from process-based, or mechanistic models, to statistical, or correlative approaches (Dormann
92 *et al.*, 2012). Regional and global databases on known pest distributions are commonly used to
93 parameterize these models, either providing direct estimates of pests' ecological niches (Venette *et al.*,
94 2010; Kriticos, 2012), or indirectly via shared geographic ranges (Paini *et al.*, 2010, 2016; Eschen *et*
95 *al.*, 2014).

96

97 One seldom-acknowledged issue with pest distribution data in global databases is geographic bias in
98 the likelihood that a pest will be detected, correctly identified, reported and recorded (Pyšek *et al.*,
99 2008). Analysis of one of the most widely studied global pest distribution databases suggests that
100 hundreds of pests already present in many developing countries have not been reported (Bebber *et al.*,
101 2014b). The total number of observed pests in an administrative area (country, or administrative
102 division for larger countries) can be largely explained by scientific capacity and agricultural production.
103 Under a scenario of globally high scientific and technical capacity (i.e. where all countries have US-
104 level per capita GDP and research expenditure), analysis predicts that many countries across the
105 developing world would report hundreds more pests. This suggests that a large fraction of the *current*
106 agricultural pest burden is unreported and unknown, and that even the best global databases suffer from
107 severe observational bias. This has potentially serious consequences for both plant biosecurity activities
108 and for research based upon these databases. Such observational bias may have implications for SDM
109 methods that infer environmental tolerances from observed distributions. Scientific capacity, economic
110 development, and the ability to detect, identify and report pests, are strongly correlated with latitude, as
111 is climate (Bebber *et al.*, 2014b). Under-reporting of pests at low latitudes will therefore bias estimation
112 of climate tolerances, as occurrence is underreported in warmer regions. Reducing this observational
113 bias by strengthening pest identification efforts in the developing world is therefore critical in
114 improving scientific understanding of pest distributions, and in PRA.

115

116 The People's Republic of China (henceforth referred to as China) has been predicted to harbour the
117 largest number of pests (Bebber *et al.*, 2014b). China produces the largest quantity of crops by tonnage
118 globally, and has the greatest diversity of production. Both factors are strong determinants of recorded
119 pest numbers (Bebber *et al.*, 2014b). Yet, the actual recorded number of pests in China is much smaller
120 than expected (Bebber *et al.*, 2014b). For many countries, under-reporting of agricultural pests is likely
121 to be purely a function of the lack of institutional capacity to detect, identify, and report incidences in
122 the scientific and 'grey' literature used by CABI to populate the distribution database. For China, there
123 is potentially an interesting alternative. The Chinese literature was, until the reforms of 1978, largely
124 inaccessible to Western academia. Even post-reform and the opening of China instigated by Deng
125 Xiaoping, Chinese-language publications are not commonly accessed by English-speaking researchers.
126 A scientifically-important translation of the Chinese literature is the reporting of the anti-malarial
127 compound artemisinin (Klayman, 1985). The Chinese research literature, having developed isolated
128 from the Western literature, therefore provides a potentially independent data source for testing models
129 of pest distributions.

130

131 Here, we develop statistical models of global pest presence using a database of known pest occurrences
132 and confront the predictions of individual pest or pathogen presence in China's provinces with
133 observations from the Chinese literature. In addition, we compare models in which pest absences are
134 treated as true absences with models in which absences are weighted according to estimates of scientific
135 and technical capacity of a given country to report plant health risks, to investigate the effect of
136 observational bias and pseudoabsences in pest distribution modelling. We then apply our distribution
137 models globally to all unreported pests in all regions, to give predicted probabilities of presence. Finally,
138 we list those pests that are probably present, but as yet unreported, around the world.

139

140 **Materials and Methods**

141 We obtained pest distribution data from the CABI Knowledge Bank database in January 2014 with
142 permission (CABI, Wallingford, UK). The database comprised 91,030 records of the observed
143 distributions of 1901 agricultural pests by administrative division of each country, e.g. US States. In
144 total, 384 geographical units were included in the model, comprising 221 countries plus sub-national
145 divisions for Australia (7), Brazil (28), Canada (13), China (31), India (33), and the USA (51).
146 Geographical regions such as Bouvet Island which were smaller than a single pixel (5 arc minute
147 resolution, or approximately 100 sq. km) of the gridded crop distribution database we employed were
148 excluded from the analysis. Host crop spatial distributions for 175 crops at 5 arc minute resolution were
149 obtained from the EarthStat database (<http://www.earthstat.org/>; Monfreda *et al.*, 2008). Known plant
150 hosts of each pest were taken from the CABI Knowledge Bank, and the host genera matched to the
151 genera in the list of 175 crops. Pests without known hosts in this list of 175 crops were excluded from

152 the analysis. Pests from taxonomic groups with fewer than 50 species (e.g. Acari, Gastropoda and
153 various other insect taxa) were also excluded from the analysis. This left a total of 1739 pests comprising
154 124 species, subspecies and pathotypes of Bacteria, 106 Diptera, 215 Coleoptera, 398 Fungi, 233
155 Hemiptera, 248 Lepidoptera, 99 Nematoda, 61 Oomycota and 209 viruses. Assigning reported
156 presences for each pest to each geographical region gave a dataset of 667,776 presences or absences for
157 each pest-region pair. In total, there were 81,821 presences (12.2 per cent of the total) in the final dataset.

158

159 We developed Generalized Linear Models (GLM), using the *glm* function (*MASS* package) in R v.3.4.0
160 with logit link for binomial data (R Development Core Team, 2017), for the presence or (pseudo-)
161 absence of each pest in each region. Model predictors were as follows: log-transformed *per capita* GDP
162 for the country as a whole in 2016 (World Bank data, <http://data.worldbank.org/>); log-transformed total
163 number of crop host genera for the pest (CABI Knowledge Bank, obtained with permission); log-
164 transformed area (ha) of the pest's host crop distribution (summing planted areas of all known host
165 crops in each geographical region); log-transformed host crop area (ha) of neighbouring (i.e. with land
166 border) regions which have reported the pest as present (set to zero if no neighbours have reported the
167 pest); log-transformed total fraction of regions globally that have reported the pest; and log transformed
168 distance (km) of crop area to the coast (calculated as the distance of the centroid of the crop area
169 distribution from the nearest coastline). Log transformations were applied to distribute the predictor
170 variable values more evenly across the sample space. Briefly, the rationale for these predictors was that
171 GDP is a proxy for historical trade (Pyšek *et al.*, 2010) and observational capacity (Bebber *et al.*,
172 2014b), host area indicates the available habitat for each pest, host number indicates the degree of biotic
173 generalism of the pest, neighbouring-region presence indicates the potential for spread across a land
174 border, fraction of regions reporting presence indicates global ubiquity and environmental generalism,
175 and distance to coast indicates proximity to international shipping ports (Chapman *et al.*, 2017).

176

177 We developed two pest distribution models. The 'unweighted' model included geographical and
178 biological predictors and treated all unobserved pests as absent from a region. The 'weighted' model
179 treated unobserved pests as potentially pseudo-absent, using a function of the scientific and technical
180 capacity of each country (Bebber *et al.*, 2014b). Presences were taken as being correct and
181 unambiguous, and given a weighting of unity. Absences were weighted by the logarithm of the
182 agricultural and biological sciences publication output of each country from 1996 – 2016 (Scimago
183 Lab, 2017), normalized to the logarithm of the output of the USA (the world's most scientifically
184 productive country), such that the absence weight $w_0 = \log(s)/\log(s_{USA})$. Thus, pests unreported from
185 scientifically advanced nations were assumed not to be present (or, present at undetectable population
186 density), while pests unreported from developing nations were less informative of absence. China, with
187 the second largest research output, had $w_0 = 0.93$, suggesting that non-reporting of a pest should be
188 relatively strong evidence of its physical absence. However, we hypothesized that non-reporting in the

189 CABI databases could be due to lack of translation from the Chinese literature, therefore we set w_0 to
190 zero for China, effectively omitting these pseudo-absences from the analysis. The weighted and
191 unweighted models were compared with a null model assuming constant presence probability using
192 Likelihood Ratio Tests.

193

194 To validate the models we predicted the probability of presence for a random sample of 100 as-yet
195 unobserved pests in all Chinese Provinces, but excluding Taiwan. The Chinese literature was searched
196 for observations of these unobserved pests in China. We used the text mining methodology designed
197 by CABI for their Plantwise Knowledge Bank. The following rules were followed to locate pest records
198 in the Chinese literature:

- 199 - Include only papers that are primarily about distribution data, not those where distribution is
200 mentioned, but where something else is the primary focus. If this is unclear do not process the
201 paper.
- 202 - Mine only the primary literature (including Masters and Doctoral theses), not meta-analyses,
203 reviews, or non-peer reviewed (“grey”) literature.
- 204 - Pest and host names must be preferred scientific names, following the CAB Thesaurus
205 (www.cabi.org/cabthesaurus/) and the Plant List (<http://www.theplantlist.org/>).
- 206 - Record country and location information given in the paper, including latitude/longitude. CABI
207 uses five levels for location, from the largest scale (i.e., provincial) to the smallest (i.e.,
208 village/town).
- 209 - Record date of observation/collection (entering each year separately) and date of publication.
210 Can be left blank if not given, or use the date of receipt in the diagnostic laboratory as a
211 surrogate for date of collection.
- 212 - Record pest status – present/not found. Only record absence if pest absence is specifically stated
213 in the paper.
- 214 - Record pest status using only the status terms defined by CABI, and only if used in the paper
215 e.g. “widespread”, “restricted” “soil only” “greenhouse only” (see CABI guidelines for
216 complete list).
- 217 - Record if the paper was a first record of that pest or not and details of this (e.g. “first record in
218 <country/location>”, “first record on <host species name>”)
- 219 - Only enter data where the pest/pathogen has been clearly identified, not just symptoms seen.
- 220 - Record only natural infections, not artificial inoculants.

221

222 Combinations of pests and locations were submitted to several search engines. The priority of search
223 engines was: Baidu (www.baidu.com), China National Knowledge Infrastructure (CNKI,
224 <http://www.cnki.net>), Chongqing VIP Information Company (CQVIP, <http://lib.cqvip.com/>), and
225 Wangfang Data (<http://www.wanfangdata.com.cn>). Baidu is the most popular Chinese internet search

226 engine. CNKI is led by Tsinghua University, and supported by ministries of the Chinese Government.
227 CQVIP, formerly known as Database Research Center under the Chongqing Branch of the Institute of
228 Scientific & Technical Information of China (CB-ISTIC), was China's first Chinese journal database
229 research institution. Wanfang Data is an affiliate of the Chinese Ministry of Science & Technology, and
230 provides access to a wide range of database resources.

231

232 Publication titles were searched first, followed by full text interrogation. The first 50 search results were
233 scanned before dismissing a search term. The first search term combination was pest name and location
234 (Province). If this yielded no result, then pest name and various distribution terms were tried. These
235 distribution terms were: "catalogues" OR "checklists" OR "distribution" OR "inventories" OR "new
236 records" OR "surveys" OR "geographical distribution" OR "new geographic records" OR "new host
237 records". Searches included local names in Chinese where these were known or could be identified
238 from the literature, preferred scientific names, and non-preferred scientific names from CAB Thesaurus
239 (<https://www.cabi.org/cabthesaurus/>). Searches continued until one piece of literature was found for
240 that pest in that region, that fitted all of the requirements for CABI text mining.

241

242 If a pest was not found in any of these searches, it was assumed to be absent from the literature and thus
243 effectively absent from the region. We cannot prove, however, that a pest is present at very low
244 population density and has not yet been detected (Crooks, 2005).

245

246 Modelled probabilities of reported pest presence in the global dataset, P_G , were obtained from the
247 predictor variables for each pest-region combination, for each GLM (*predict* function in R). We then
248 compared P_G with the observed presence-absence data for our Chinese sample data using logistic
249 regressions (*glm* function in R) and Receiver-Operator Characteristic (ROC) curves (*pROC* library for
250 R). The logistic regression coefficients c and m determine the probability of pest presence in the Chinese
251 sample as $P_C = 1/(1 + \exp(-(c + mP_G)))$. ROC curves describe the relationship between the true positive
252 rate (sensitivity, the fraction of presences correctly identified as presences) and false positive rate ($1 -$
253 specificity, where specificity is the fraction of absences correctly classified as absences) as the threshold
254 for a binary classifier is decreased from one (classifying any presence probability less than one as
255 absent) to zero (classifying any positive probability as present). A good predictor will have a high true
256 positive rate and low false positive rate for any classification threshold, whereas a poor predictor will
257 have roughly equal true and false positive rates (i.e., be uninformative). The Area Under Curve (AUC)
258 for the ROC curves gives the probability that, for a random pair of presence and absence observations,
259 the presence probability will be greater for the presence than the absence (Jiménez-Valverde, 2012).
260 Models with good discrimination ability should have AUC significantly greater than half.

261

262 For illustration, we identified probably present pests (PPP) as those for which are currently unreported
 263 from a particular region, but for which $P_G > 0.75$ in our weighted model. This threshold was chosen
 264 based on the Kent scale which suggests a probability of 0.75 as an event that would generally be
 265 described as ‘probable’ (Kent, 1994). This is an arbitrary definition but allows us to suggest some of
 266 the pests that PRA and phytosanitary activities may want to focus on.

267

268 **Results**

269 Globally, P_G increased significantly with presence in neighbouring regions, the area of host crops, the
 270 global prevalence of the pest and per capita GDP in both models (Table 1). P_G declined with mean
 271 distance from the coast and known host crop genera per pest. The models explained similar fractions of
 272 the deviance, and had very similar ROC curves with AUC around 88 per cent (Table 1). P_G was always
 273 higher for the weighted model, because absences were down-weighted (i.e. fewer true zeros), but
 274 predictions for the two models were very highly correlated ($r = 0.98$). The models found the highest P_G
 275 for Hemiptera and Lepidoptera, and lowest for Nematoda, Bacteria and Acari, compared with other
 276 taxonomic groupings.

277

278 Table 1. GLMs for global pest presence. The unweighted model treated unobserved pests as true
 279 absences. The weighted model weighted pseudo-absences as a function of country scientific capacity.
 280 The unweighted model had AIC = 339872, AUC = 0.88, Nagelkerke $R^2 = 0.40$, McFadden $R^2 = 0.32$.
 281 The weighted model had AIC = 308171, AUC = 0.88, Nagelkerke $R^2 = 0.37$, McFadden $R^2 = 0.31$.
 282 CoastDist is distance of crop centroid from the coast (km), GDP is per capita GDP (US\$), Hosts is
 283 reported number of host crop genera, HostArea is harvested area of known host crops, NeigArea is
 284 harvested area of host crops in neighbouring regions that have reported the pest, and Prevalence is the
 285 fraction of all regions that have reported the pest.

	Unweighted model				Weighted model			
	Mean	SE	Z	Pr(> Z)	Mean	SE	Z	Pr(> Z)
Acari (Intercept)	-3.67	0.051	-72.3	0.000	-0.897	0.055	-16.3	0.000
+ Bacteria	-0.091	0.032	-2.9	0.004	-0.073	0.034	-2.2	0.014
+ Coleoptera	0.036	0.030	1.2	0.240	0.039	0.032	1.2	0.180
+ Diptera	0.092	0.034	2.7	0.006	0.104	0.036	2.9	0.026
+ Fungi	0.027	0.028	1.0	0.337	0.033	0.030	1.1	0.380
+ Hemiptera	0.167	0.029	5.7	0.000	0.150	0.031	4.8	0.000
+ Lepidoptera	0.145	0.029	4.9	0.000	0.134	0.032	4.2	0.000
+ Nematoda	-0.150	0.033	-4.5	0.000	-0.143	0.035	-4.1	0.000
+ Oomycota	0.046	0.034	1.4	0.176	0.061	0.037	1.7	0.151
+ Virus	0.047	0.030	1.6	0.120	0.067	0.033	2.1	0.137

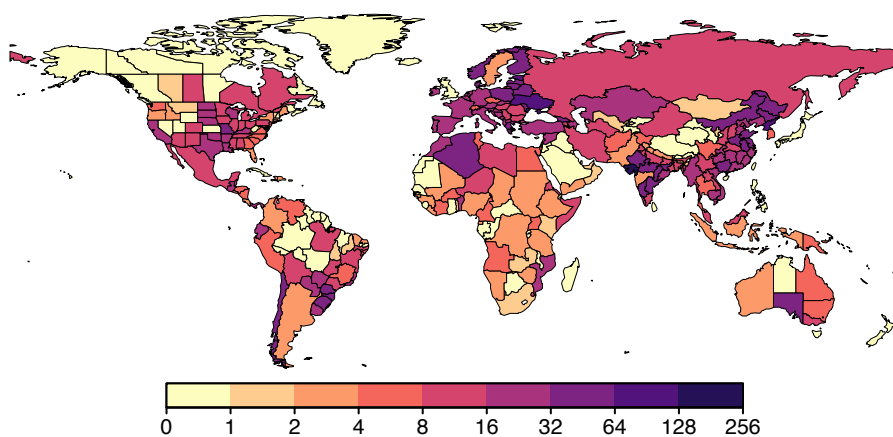
log(CoastDist + 1)	-0.176	0.004	-49.0	0.000	-0.222	0.004	-57.7	0.000
log(GDP + 1)	0.295	0.004	81.5	0.000	0.086	0.004	22.3	0.000
log(Hosts + 1)	-0.300	0.004	-71.4	0.000	-0.297	0.005	-65.6	0.000
log(HostArea + 1)	0.171	0.001	123.1	0.000	0.159	0.001	108.1	0.000
log(NeigArea + 1)	0.140	0.001	181.2	0.000	0.142	0.001	173.5	0.000
log(Prevalence)	0.842	0.007	124.3	0.000	0.867	0.007	121.3	0.000

286

287 We validated the models with reports of pests abstracted from the Chinese literature.

288

289 For illustration, we defined a ‘probably present pest’ (PPP) as one unreported from a region, but with
 290 $P_G > 0.75$ (using the weighted model). Overall, only 4702 of 585955 (0.8 per cent) of all unreported
 291 pest-region combinations fell into this class (Supplementary Table S1). The number of PPPs per pest
 292 category was greatest for Fungi (2052) and Hemiptera (859). Overall, 86 per cent of unreported pest-
 293 region combinations were predicted to be unlikely ($P_G < 0.25$). China, India, the USA and Eastern
 294 Europe had the largest numbers of predicted PPPs, along with other parts of East Asia and Southern
 295 Brazil (Figure 1). The top ten PPPs by number of global regions were *Cochliobolus heterostrophus*
 296 (Ascomycota: Pleosporales, a pathogen of maize), *Rhopalosiphum padi* (Arthropoda: Hemiptera, cereal
 297 pest), *Gibberella fujikuroi* (Ascomycota: Hypocreales, rice pathogen), *Sitophilus zeamais* (Arthropoda:
 298 Coleoptera), maize and rice pest), *Schizaphis graminum* (Arthropoda: Hemiptera, pest of Poaceae
 299 cereals), *Setosphaeria turcica* (Ascomycota: Pleosporales, maize pathogen), *Aphis spiraecola*
 300 (Arthropoda: Hemiptera, wide host range), *Nezara viridula* (Arthropoda: Hemiptera, legume pest),
 301 *Acyrtosiphon pisum* (Arthropoda: Hemiptera, legume pest) and *Rhopalosiphum maidis* (Arthropoda:
 302 Hemiptera, pest of maize and other crops).



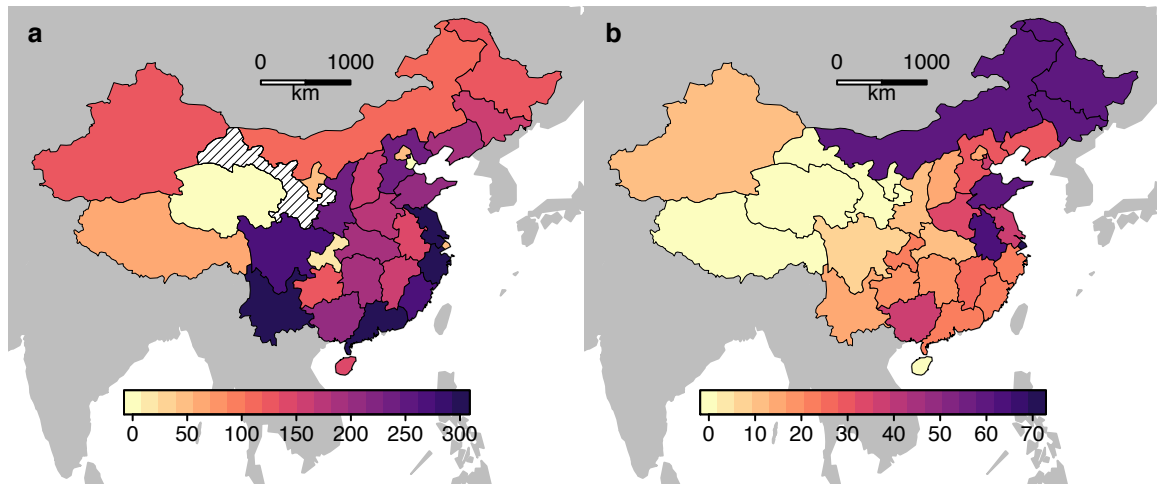
303

304 Figure 1. Total number of probably present pests (PPP) in all countries and sub-national regions.

305

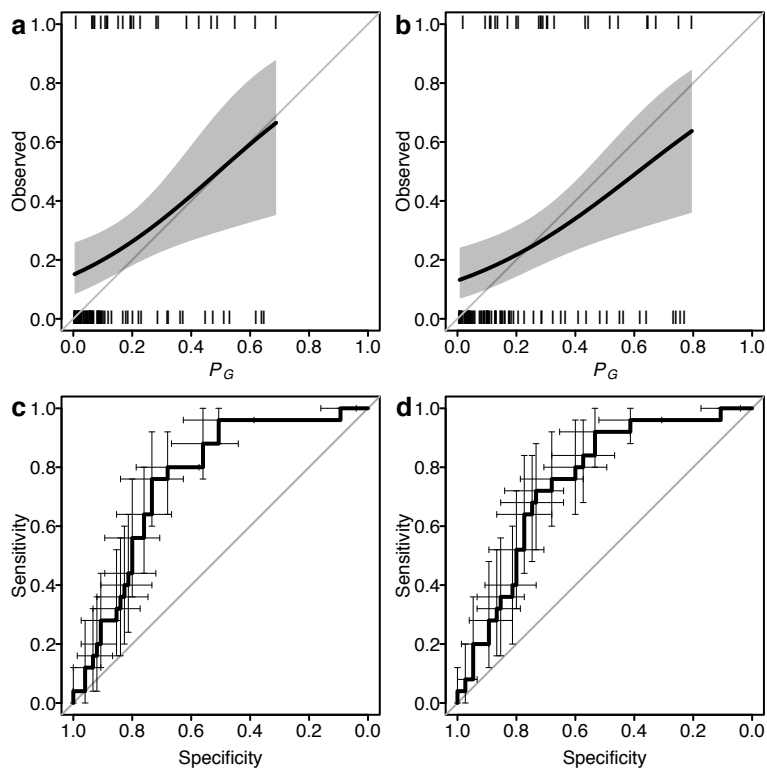
306 Total numbers of recorded pests in China’s Provinces and municipalities increased from northern and
 307 central regions to southern and coast regions (Figure 2a), except for the central province of Gansu which
 308 had 826 reported pests. There is no obvious reason why numbers would be so large in Gansu. Here,

309 agricultural production is moderate, and there are no particular academic centres which could account
310 for observational bias. Hence, the Gansu values appear to be an artefact of the CABI database. The
311 smallest numbers of recorded pests were in the mountainous provinces of Qinghai (0) and Xizang
312 Zizhiqu (Tibet) (73), the central provinces of Ningxia (48), and the municipalities of Chongqing (24),
313 Tianjin (3), Beijing (50) and Shanghai (55). Total numbers were largest in the coastal provinces of
314 Guangdong (301), Zeijiang (294), Jiangsu (293), Fujian (263), and also in the southern provinces of
315 Yunnan (291) and Sichuan (259).
316



317
318 Figure 2. a) Total number of pests recorded in the CABI pest distribution database by China Province
319 (excluding Taiwan). Hatched region is Gansu, see text for details. b) Total number of probably present
320 pests (PPP) in China Provinces.
321

322 We validated our models using published pest observations from the Chinese literature. Both models
323 were significant predictors of pest presence/absence for 100 randomly-sampled pest-Province
324 combinations, of which 27 were found to be present (Figure 3, Supplementary Table S2). For the
325 unweighted model, the coefficients of the logistic function were $c = -1.73 \pm 0.34$ and $m = 3.52 \pm 1.25$
326 (likelihood ratio test vs null model, $p = 0.0043$). For the weighted model, the coefficients were $-1.90 \pm$
327 0.38 and 3.19 ± 0.96 (likelihood ratio test, $p = 0.0006$). The predictive power of the models was also
328 tested using ROC curves, demonstrating significant discriminant ability with AUC of 0.76 (95 per cent
329 Confidence Interval 0.66 – 0.86) for the unweighted model, and AUC 0.75 (0.64 – 0.86) for the
330 weighted model (Figure 3). Our analysis revealed gaps in the CABI database, which is commonly used
331 for analyses of global pest distributions. Taking one important potato pest, late blight *Phytophthora*
332 *infestans* (Oomycota), as an example, high presence probabilities (> 0.75) were predicted for ten
333 provinces listed as not reporting this pest in the CABI database. However, this pathogen has been
334 reported present throughout the potato-growing regions of China, including Guangdong (Guo *et al.*,
335 2010).
336



337

338 Figure 3. Model prediction tests. Observed presence/absence of 100 pest-province combinations vs. P_G
339 from a) unweighted model and b) weighted model. Curves and shaded areas show mean and 95% CI
340 for logistic regression fits. Tick marks show observed data. Grey diagonals show identity relationship.
341 ROC curves for c) unweighted and d) weighted models. Error bars show 95% CI for specificity and
342 sensitivity derived from 2000 bootstrap replications.

343

344 For China, the total number of PPPs increased from west to east (Figure 2b), and was greatest in the
345 north eastern provinces of Jilin (59), Heilongjiang (58), and Inner Mongolia (58), the eastern provinces
346 of Shandong (60) and Anhui (61), well as the ports of Shanghai (71) and Tianjin (51). The eastern
347 provinces of Xizang Zizhiqu (Tibet) (1), Qinghai (1), Gansu (0) and Ningxia (2) had the lowest numbers,
348 along with the island of Hainan (0) (Figure 3). The total number of PPPs in China was 827, the majority
349 being Fungi (332) and Hemiptera (175). The top ten most-common PPPs in China were (in decreasing
350 order) *Gibberella fujikuroi* (Ascomycota: Hypocreales, rice pathogen), *Aphis spiraecola* (Arthropoda:
351 Hemiptera, generalist), *Delia platura* (Arthropoda: Diptera, pest of legumes), *Acyrtosiphon pisum*
352 (Arthropoda: Hemiptera, legume pest), *Rhopalosiphum padi* (Arthropoda: Hemiptera, cereal pest),
353 *Schizaphis graminum* (Arthropoda: Hemiptera, pest of Poaceae), *Curvularia* sp. (Fungi: Ascomycota,
354 generalist pathogen), *Rhopalosiphum maidis* (Arthropoda: Hemiptera, pest of maize and other crops),
355 *Agrotis ipsilon* (Arthropoda: Lepidoptera, generalist pest), *Lasiodiplodia theobromae* (Ascomycota:
356 Botryosphaerales, generalist pathogen). Thus, many of the most common PPPs in China were also
357 common globally.

358

359 **Discussion**

360 The Chinese literature provided strong and significant support for the predictions of pest distribution
361 models based upon host distribution, pest prevalence, and other socioeconomic factors. China's
362 growing economy is expected to lead to large influxes of invasive species, including pests, in coming
363 years (Ding *et al.*, 2008). Analysis of temporal trends in CABI pest observations show a relatively
364 smooth increase in pests from 1950-2000, but the pattern for China is more complex, with a slow
365 increase from 1950 until the late 1970s, a step increase, and then a more rapid growth in pest numbers
366 from 1980 onwards (Bebber *et al.*, 2014a). One potential determinant of this sudden acceleration is the
367 strong support for science and technology given by Deng Xiaoping in 1978, which lead to an increase
368 in funding and academic freedom following the anti-intellectualism of the Cultural Revolution. China
369 now ranks second only to the USA in annual R&D expenditure (IMF, 2013) and scientific output
370 (Scimago Lab, 2017).

371

372 We identified a number of pests that were very likely to be present, and the majority of these PPPs were
373 globally distributed and had wide host ranges. Their distributions commonly spanned wide latitudinal
374 ranges, indicating broad climatic tolerances. *C. heterostrophus*, or Southern Leaf Spot, is primarily
375 known as a pathogen of maize but has a wide host range. It has a wide geographic distribution both
376 latitudinally and across continents, resulting in a high likelihood of occurrence in other regions where
377 hosts are present. For example, *C. heterostrophus* is currently recorded only in eastern regions of North
378 America, where most maize is grown. The lack of reported observations in the western regions of North
379 America may be due to the fact that maize, the major host, is uncommon, and hence the disease currently
380 has little impact. *C. sativus*, causing root and foot rot, also has a very wide geographic distribution, but
381 an even wider host range. It is reported from Texas, Oklahoma, Mississippi, Illinois and Tennessee, but
382 not from neighbouring Arkansas or Missouri. Hence, the high presence probability in these States. A
383 similar pattern is seen for the maize pathogen *S. turcica*. Another global species, *R. maidis*, the green
384 corn aphid, is reported across Europe and in Russia, but, like many other pests, not from the former
385 Soviet states of Ukraine, Belarus, Lithuania, Latvia and Estonia. It is plausible that reporting from these
386 nations was less likely when they were part of the USSR.

387

388 Predictors like host availability, presence in neighbouring territories and global prevalence were
389 expected to have positive relations with presence probability. The negative relation with distance from
390 coast is likely to be related to import via shipping ports (Huang *et al.*, 2012; Liebhold *et al.*, 2013), and
391 supports the observation that islands report more pests than countries with land borders (Bebber *et al.*,
392 2014b). Detailed modelling of individual pest climate responses (Bregaglio *et al.*, 2012; Kriticos *et al.*,
393 2013) for such a large number of pests was beyond the scope of this study. Implicitly, we can assume
394 that the presence of the host crop indicates that the climate is suitable for the pest (Paini *et al.*, 2016),
395 though we acknowledge that this is not necessarily the case (Berzitis *et al.*, 2014). The negative

396 relationship with number of host genera per pest might suggest that host specialists are more likely to
397 invade and establish than host generalists, once host availability has been taken into account. For the
398 practical purposes of PRA, our models provide reliable probability estimates for the presence of
399 unreported pests at subnational resolution, and we have provided a global list of the unreported pests
400 whose presence is most likely (Table S2).

401

402 We addressed the issue of pseudo-absences in the CABI data by statistically weighting missing pest
403 observations in proportion to the scientific output of the reporting nation, since scientific output had
404 been confirmed as a strong determinant of total reported pest numbers (Bebber *et al.*, 2014b). Often,
405 unreported pests are treated as true absences in pest risk analyses (Paini *et al.*, 2016). The positive
406 relation of GDP with presence probability supports our hypothesis that wealthy countries are more
407 likely to detect and report pests (Bebber *et al.*, 2014b). Once observational bias is controlled for using
408 scientific capacity-based weighting, per capita GDP becomes a weaker determinant. Our weighted
409 model has similar overall explanatory power to our unweighted model. Nevertheless, the issue of
410 observational biases related to country-level socioeconomic variation has been raised several times for
411 various classes of organism (Jones *et al.*, 2008; Pyšek *et al.*, 2008; Westphal *et al.*, 2008; Boakes *et al.*,
412 2010; Bebber *et al.*, 2013, 2014b), and we therefore recommend the application of appropriate statistical
413 controls when analysing datasets produced from reports of species presences (as opposed to
414 distributional datasets derived from rigorous sampling protocols).

415

416 Our SDM was statistical, fitting response functions for various predictors to the probability of pest
417 presence. Many SDM approaches exist, from highly mechanistic models based on pest biology and
418 ecology (Bregaglio *et al.*, 2012; Skelsey *et al.*, 2016) to purely statistical models that utilize only
419 patterns in known distributions (Paini *et al.*, 2010). The rarity of quantitative model input into PRAs is
420 partly due to the scarcity of empirical data available on pest biology and epidemiology required to
421 parameterize mechanistic models, and so key biological parameters are often inferred from known
422 distributions (Robinet *et al.*, 2012). This is particularly the case for newly emergent pathogens for which
423 experimental investigations have not yet been conducted. The European Food Safety Authority (EFSA)
424 has developed quantitative PRA guidelines that recommend modelling approaches and data sources for
425 assessing invasion and establishment risk (Jeger *et al.*, 2018), and application of these methods was
426 attempted for *Diaporthe vaccinii*, a pest of blueberries (Jeger *et al.*, 2017). However, most of the
427 epidemiological data required for this pest was unavailable, and the risk assessment was thus based on
428 expert opinion or data from related pests (Jeger *et al.*, 2017). Epidemiological parameters can be poorly
429 constrained even for long-established pests. For example, coffee leaf rust fungus (*Hemileia vastatrix*)
430 has affected coffee production for more than a century, but a recent infection model relied upon
431 temperature response functions derived from the single available study published three decades
432 previously (Bebber *et al.*, 2016). Initiatives such as the EU-funded PRATIQUE project (2008-11) have

433 attempted to fill this knowledge gap and enable modelling by collating available ecophysiological data
434 for insect pests (Baker, 2012). While the advantages and disadvantages of the many different pest
435 distribution and impact models continue to be researched and debated (Venette *et al.*, 2010; Dormann
436 *et al.*, 2012; Robinet *et al.*, 2012; Sutherst, 2014), it is clear that practical application of these methods
437 in PRA remains limited.

438

439 SDM for pests has direct policy implications for PRA and plant biosecurity. PRA is guided by
440 International Standards for Phytosanitary Measures (ISPM), which are part of the International Plant
441 Protection Convention (IPPC) (MacLeod *et al.*, 2010). ISPMs tend to rely on expert judgement for
442 PRA, rather than quantitative modelling to support decision making. ISPM No. 21 “Pest Risk Analysis
443 for Regulated Non-Quarantine Pests”, endorsed in 2004, mentions use of pest and host life-cycle and
444 epidemiological information, but not quantitative modelling (FAO, 2004). Individual PRAs similarly
445 employ a qualitative approach. For example, the Australian Government’s PRA for *Drosophila suzukii*
446 references only a single unpublished report on SDM for this species, conducted for North America.
447 Probabilities of *D. suzukii* spread within Australia are qualitatively assessed by comparison with
448 observations in other countries (Department of Agriculture, Fisheries and Forestry, 2013). The
449 European and Mediterranean Plant Protection Organization (EPPO) PRAs occasionally include model
450 results. For example, a climate matching for the bacterium *Xanthomonas axonopodis* pv. *allii* was
451 undertaken using the CLIMEX model, to identify areas at risk within the EPPO region (EPPO, 2008).
452 However, as discussed previously, appropriate empirical studies are rare (Jeger *et al.*, 2017). Our results
453 contribute to the quantification of risk within PRA by providing probabilistic estimates for the presence
454 of hundreds of unreported pests around the world, thereby improving understanding of the threats to
455 global agriculture. With growing evidence that pest ranges are shifting poleward in response to global
456 climate change (Bebber *et al.*, 2013), our poor knowledge of pest distributions, particularly in the
457 developing world, is troubling, both because of the burden these organisms place on farmers who have
458 little access to detection and control technologies, and because invasions of temperate regions are likely
459 to occur from warmer regions. Improved targeting of phytosanitary measures through quantitative PRA
460 is therefore vital to crop protection.

461

462 **Acknowledgements**

463 We thank the British Society for Plant Pathology for an Undergraduate Vacation Bursary for Elsa Field.
464 We thank Le Yei and Li Huili (Kunming Institute of Botany) for assistance in text mining the Chinese
465 literature and Ho Wai Yim for translations. CABI is an international intergovernmental organisation
466 and we gratefully acknowledge the core financial support from our member countries (and lead
467 agencies) including the United Kingdom (Department for International Development), China (Chinese
468 Ministry of Agriculture), Australia (Australian Centre for International Agricultural Research), Canada
469 (Agriculture and Agri-Food Canada), Netherlands (Directorate-General for International Cooperation),

470 and Switzerland (Swiss Agency for Development and Cooperation). See [https://www.cabi.org/about-](https://www.cabi.org/about-cabi/who-we-work-with/key-donors/)
471 [cabi/who-we-work-with/key-donors/](https://www.cabi.org/about-cabi/who-we-work-with/key-donors/) for full details.

472

473 **Author contribution**

474 DB conducted the analyses and wrote the manuscript. EF and GH searched the Chinese literature. TH
475 assisted with CABI data acquisition. All authors contributed ideas and edited the manuscript.

476

477 **Data accessibility**

478 Pest distribution data are available with permission from CABI, Nosworthy Way, Wallingford, OX10
479 8DE, UK

480 Sources for other datasets used in the analysis are given in the text.

481

482

483 **References**

- 484 Avelino, J., Cristancho, M., Georgiou, S., Imbach, P., Aguilar, L., Bornemann, G., Läderach, P.,
485 Anzueto, F., Hruska, A.J. & Morales, C. (2015) The coffee rust crises in Colombia and
486 Central America (2008–2013): impacts, plausible causes and proposed solutions. *Food*
487 *Security*, **7**, 303–321.
- 488 Bacon, S.J., Aebi, A., Calanca, P. & Bacher, S. (2014) Quarantine arthropod invasions in Europe: the
489 role of climate, hosts and propagule pressure. *Diversity and Distributions*, **20**, 84–94.
- 490 Baiser, B., Olden, J.D., Record, S., Lockwood, J.L. & McKinney, M.L. (2012) Pattern and process of
491 biotic homogenization in the New Pangaea. *Proceedings of the Royal Society B: Biological*
492 *Sciences*, **279**, 4772–4777.
- 493 Baker, R.H.A. (2012) An introduction to the PRATIQUE Research Project. *EPPO Bulletin*, **42**, 1–2.
- 494 Baker, R.H.A., Anderson, H., Bishop, S., MacLeod, A., Parkinson, N. & Tuffen, M.G. (2014) The
495 UK Plant Health Risk Register: a tool for prioritizing actions. *EPPO Bulletin*, **44**, 187–194.
- 496 Bebber, D.P. (2015) Range-Expanding Pests and Pathogens in a Warming World. *Annual Review of*
497 *Phytopathology*, **53**, 335–356.
- 498 Bebber, D.P., Castillo, A.D. & Gurr, S.J. (2016) Modelling coffee leaf rust risk in Colombia with
499 climate reanalysis data. *Philosophical Transactions of the Royal Society B: Biological*
500 *Sciences*, **371**, 20150458.
- 501 Bebber, D.P. & Gurr, S.J. (2015) Crop-destroying fungal and oomycete pathogens challenge food
502 security. *Fungal Genetics and Biology*, **74**, 62–64.
- 503 Bebber, D.P., Holmes, T. & Gurr, S.J. (2014a) The global spread of crop pests and pathogens. *Global*
504 *Ecology and Biogeography*, **23**, 1398–1407.
- 505 Bebber, D.P., Holmes, T., Smith, D. & Gurr, S.J. (2014b) Economic and physical determinants of the
506 global distributions of crop pests and pathogens. *New Phytologist*, **202**, 901–910.
- 507 Bebber, D.P., Ramotowski, M.A.T. & Gurr, S.J. (2013) Crop pests and pathogens move polewards in
508 a warming world. *Nature Climate Change*, **3**, 985–988.
- 509 Berzitis, E.A., Minigan, J.N., Hallett, R.H. & Newman, J.A. (2014) Climate and host plant availability
510 impact the future distribution of the bean leaf beetle (*Cerotoma trifurcata*). *Global Change*
511 *Biology*, **20**, 2778–2792.
- 512 Boakes, E.H., McGowan, P.J.K., Fuller, R.A., Chang-qing, D., Clark, N.E., O’Connor, K. & Mace,
513 G.M. (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence
514 data. *PLoS Biol*, **8**, e1000385.
- 515 Bregaglio, S., Cappelli, G. & Donatelli, M. (2012) Evaluating the suitability of a generic fungal
516 infection model for pest risk assessment studies. *Ecological Modelling*, **247**, 58–63.
- 517 Chapman, D., Purse, B.V., Roy, H.E. & Bullock, J.M. (2017) Global trade networks determine the
518 distribution of invasive non-native species. *Global Ecology and Biogeography*, **26**, 907–917.
- 519 Crooks, J.A. (2005) Lag times and exotic species: The ecology and management of biological
520 invasions in slow-motion. *Ecoscience*, **12**, 316–329.
- 521 DEFRA (2018) UK Plant Health Risk Register.
- 522 Department of Agriculture, Fisheries and Forestry (2013) *Final pest risk analysis report for*
523 *Drosophila suzukii*, Australian Government, Canberra, Australia.
- 524 Ding, J., Mack, R.N., Lu, P., Ren, M. & Huang, H. (2008) China’s booming economy is sparking and
525 accelerating biological invasions. *BioScience*, **58**, 317.
- 526 Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin,
527 X., Römermann, C., Schröder, B. & Singer, A. (2012) Correlation and process in species
528 distribution models: bridging a dichotomy. *Journal of Biogeography*, **39**, 2119–2131.
- 529 Elith, J. & Leathwick, J.R. (2009) Species Distribution Models: Ecological Explanation and
530 Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*,
531 **40**, 677–697.
- 532 EPPO (2008) *Report of a Pest Risk Analysis for Xanthomonas axonopodis pv. allii*, European and
533 Mediterranean Plant Protection Organization, Paris.
- 534 Eschen, R., Holmes, T., Smith, D., Roques, A., Santini, A. & Kenis, M. (2014) Likelihood of
535 establishment of tree pests and diseases based on their worldwide occurrence as determined
536 by hierarchical cluster analysis. *Forest Ecology and Management*, **315**, 103–111.

- 537 FAO (2004) *ISPM 21. Pest risk analysis for regulated non-quarantine pests*, IPPC, FAO, Rome.
- 538 Fears, R., Aro, E.-M., Pais, M.S. & ter Meulen, V. (2014) How should we tackle the global risks to
539 plant health? *Trends in Plant Science*, **19**, 206–208.
- 540 Fisher, M.C., Henk, D.A., Briggs, C.J., Brownstein, J.S., Madoff, L.C., McCraw, S.L. & Gurr, S.J.
541 (2012) Emerging fungal threats to animal, plant and ecosystem health. *Nature*, **484**, 186–194.
- 542 Flood, J. & Day, R. (2016) Managing risks from pests in global commodity networks – policy
543 perspectives. *Food Security*, **8**, 89–101.
- 544 Guo, L., Zhu, X.-Q., Hu, C.-H. & Ristaino, J.B. (2010) Genetic Structure of *Phytophthora infestans*
545 Populations in China Indicates Multiple Migration Events. *Phytopathology*, **100**, 997–1006.
- 546 Huang, D., Zhang, R., Kim, K.C. & Suarez, A.V. (2012) Spatial Pattern and Determinants of the First
547 Detection Locations of Invasive Alien Species in Mainland China. *Plos One*, **7**, e31734.
- 548 IMF (2013) IMF Data and Statistics.
- 549 Jeger, M., Bragard, C., Caffier, D., Candresse, T., Chatzivassiliou, E., Dehnen-Schmutz, K., Gilioli,
550 G., Grégoire, J.-C., Jaques Miret, J.A., MacLeod, A., Navarro, M.N., Niere, B., Parnell, S.,
551 Potting, R., Rafoss, T., Rossi, V., Urek, G., Van Der Werf, W., West, J., Winter, S., Gardi, C.,
552 Mosbach-Schulz, O., Koufakis, I. & Van Bruggen, A. (2017) Pest risk assessment of
553 *Diaporthe vaccinii* for the EU territory. *EFSA Journal*, **15**, e04924.
- 554 Jeger, M., Bragard, C., Caffier, D., Candresse, T., Chatzivassiliou, E., Dehnen-Schmutz, K., Grégoire,
555 J.-C., Jaques Miret, J.A., MacLeod, A., Navajas Navarro, M., Niere, B., Parnell, S., Potting,
556 R., Rafoss, T., Rossi, V., Urek, G., Van Bruggen, A., Van Der Werf, W., West, J., Winter, S.,
557 Hart, A., Schans, J., Schrader, G., Suffert, M., Kertész, V., Kozelska, S., Mannino, M.R.,
558 Mosbach-Schulz, O., Pautasso, M., Stancanelli, G., Tramontini, S., Vos, S. & Gilioli, G.
559 (2018) Guidance on quantitative pest risk assessment. *EFSA Journal*, **16**, e05350.
- 560 Jiménez-Valverde, A. (2012) Insights into the area under the receiver operating characteristic curve
561 (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and*
562 *Biogeography*, **21**, 498–507.
- 563 Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L. & Daszak, P. (2008)
564 Global trends in emerging infectious diseases. *Nature*, **451**, 990–993.
- 565 Kent, S. (1994) Words of estimative probability. *Studies in Intelligence*, **8**, 49–65.
- 566 Klayman, D.L. (1985) Qinghaosu (Artemisinin): An Antimalarial Drug from China. *Science*, **228**,
567 1049–1055.
- 568 Kriticos, D.J. (2012) Regional climate-matching to estimate current and future sources of biosecurity
569 threats. *Biological Invasions*, **14**, 1533–1544.
- 570 Kriticos, D.J., Morin, L., Leriche, A., Anderson, R.C. & Caley, P. (2013) Combining a Climatic
571 Niche Model of an Invasive Fungus with Its Host Species Distributions to Identify Risks to
572 Natural Assets: *Puccinia psidii* Ssensu Lato in Australia. *PLoS ONE*, **8**, e64479.
- 573 Liebhold, A.M., McCullough, D.G., Blackburn, L.M., Frankel, S.J., Von Holle, B. & Aukema, J.E.
574 (2013) A highly aggregated geographical distribution of forest pest invasions in the USA.
575 *Diversity and Distributions*, **19**, 1208–1216.
- 576 MacLeod, A., Jones, G.D., Anderson, H.M. & Mumford, R.A. (2016) Plant health and food security,
577 linking science, economics, policy and industry. *Food Security*, **8**, 17–25.
- 578 MacLeod, A., Pautasso, M., Jeger, M.J. & Haines-Young, R. (2010) Evolution of the international
579 regulation of plant pests and challenges for future plant health. *Food Security*, **2**, 49–70.
- 580 Monfreda, C., Ramankutty, N. & Foley, J.A. (2008) Farming the planet: 2. Geographic distribution of
581 crop areas, yields, physiological types, and net primary production in the year 2000. *Global*
582 *Biogeochemical Cycles*, **22**, GB1022.
- 583 Ordonez, N., Seidl, M.F., Waalwijk, C., Drenth, A., Kilian, A., Thomma, B.P.H.J., Ploetz, R.C. &
584 Kema, G.H.J. (2015) Worse Comes to Worst: Bananas and Panama Disease—When Plant
585 and Pathogen Clones Meet. *PLOS Pathogens*, **11**, e1005197.
- 586 Paini, D.R., Sheppard, A.W., Cook, D.C., Barro, P.J.D., Worner, S.P. & Thomas, M.B. (2016) Global
587 threat to agriculture from invasive species. *Proceedings of the National Academy of Sciences*,
588 **113**, 7575–7579.
- 589 Paini, D.R., Worner, S.P., Cook, D.C., De Barro, P.J. & Thomas, M.B. (2010) Threat of invasive
590 pests from within national borders. *Nature Communications*, **1**, 115.

- 591 Patpour, M., Hovmøller, M.S., Shahin, A.A., Newcomb, M., Olivera, P., Jin, Y., Luster, D., Hodson,
592 D., Nazari, K. & Azab, M. (2015) First Report of the Ug99 Race Group of Wheat Stem Rust,
593 *Puccinia graminis* f. sp. *tritici*, in Egypt in 2014. *Plant Disease*, **100**, 863–863.
- 594 Pyšek, P., Jarošík, V., Hulme, P.E., Kühn, I., Wild, J., Arianoutsou, M., Bacher, S., Chiron, F.,
595 Didžiulis, V., Essl, F., Genovesi, P., Gherardi, F., Hejda, M., Kark, S., Lambdon, P.W.,
596 Desprez-Loustau, M.-L., Nentwig, W., Pergl, J., Poboljšaj, K., Rabitsch, W., Roques, A.,
597 Roy, D.B., Shirley, S., Solarz, W., Vilà, M. & Winter, M. (2010) Disentangling the role of
598 environmental and human pressures on biological invasions across Europe. *Proceedings of*
599 *the National Academy of Sciences*, **107**, 12157–12162.
- 600 Pyšek, P., Richardson, D.M., Pergl, J., Jarošík, V., Sixtová, Z. & Weber, E. (2008) Geographical and
601 taxonomic biases in invasion ecology. *Trends in Ecology & Evolution*, **23**, 237–244.
- 602 R Development Core Team (2017) *R: A Language and Environment for Statistical Computing*, R
603 Foundation for Statistical Computing, Vienna.
- 604 Robinet, C., Kehlenbeck, H., Kriticos, D.J., Baker, R.H.A., Battisti, A., Brunel, S., Dupin, M., Eyre,
605 D., Faccoli, M., Ilieva, Z., Kenis, M., Knight, J., Reynaud, P., Yart, A. & van der Werf, W.
606 (2012) A Suite of Models to Support the Quantitative Assessment of Spread in Pest Risk
607 Analysis. *PLoS ONE*, **7**, e43366.
- 608 Santini, A., Ghelardini, L., De Pace, C., Desprez-Loustau, M.L., Capretti, P., Chandelier, A., Cech,
609 T., Chira, D., Diamandis, S., Gaitniekis, T., Hantula, J., Holdenrieder, O., Jankovsky, L.,
610 Jung, T., Jurc, D., Kirisits, T., Kunca, A., Lygis, V., Malecka, M., Marcais, B., Schmitz, S.,
611 Schumacher, J., Solheim, H., Solla, A., Szabò, I., Tsopelas, P., Vannini, A., Vettraino, A.M.,
612 Webber, J., Woodward, S. & Stenlid, J. (2013) Biogeographical patterns and determinants of
613 invasion by forest pathogens in Europe. *New Phytologist*, **197**, 238–250.
- 614 Savary, S., Bregaglio, S., Willocquet, L., Gustafson, D., D’Croz, D.M., Sparks, A., Castilla, N.,
615 Djurle, A., Allinne, C., Sharma, M., Rossi, V., Amorim, L., Bergamin, A., Yuen, J., Esker, P.,
616 McRoberts, N., Avelino, J., Duveiller, E., Koo, J. & Garrett, K. (2017) Crop health and its
617 global impacts on the components of food security. *Food Security*, **9**, 311–327.
- 618 Scimago Lab (2017) Scimago Journal & Country Rank.
- 619 Shimwela, M.M., Narouei-Khandan, H.A., Halbert, S.E., Keremane, M.L., Minsavage, G.V.,
620 Timilsina, S., Massawe, D.P., Jones, J.B. & van Bruggen, A.H.C. (2016) First occurrence of
621 *Diaphorina citri* in East Africa, characterization of the *Ca. Liberibacter* species causing
622 huanglongbing (HLB) in Tanzania, and potential further spread of *D. citri* and HLB in Africa
623 and Europe. *European Journal of Plant Pathology*, **146**, 349–368.
- 624 Skelsey, P., Cooke, D.E.L., Lynott, J.S. & Lees, A.K. (2016) Crop connectivity under climate change:
625 future environmental and geographic risks of potato late blight in Scotland. *Global Change*
626 *Biology*, n/a-n/a.
- 627 Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology*
628 *Letters*, **10**, 1115–1123.
- 629 Soberón, J. & Nakamura, M. (2009) Niches and distributional areas: concepts, methods, and
630 assumptions. *Proceedings of the National Academy of Sciences*, **106**, 19644–19650.
- 631 Soberón, J. & Peterson, A.T. (2005) Interpretation of models of fundamental ecological niches and
632 species’ distributional areas. *Biodiversity Informatics*, **2**, 1–10.
- 633 Sutherst, R.W. (2014) Pest species distribution modelling: origins and lessons from history.
634 *Biological Invasions*, **16**, 239–256.
- 635 Venette, R.C., Kriticos, D.J., Magarey, R.D., Koch, F.H., Baker, R.H.A., Worner, S.P., Gómez
636 Raboteaux, N.N., McKenney, D.W., Dobesberger, E.J., Yemshanov, D., De Barro, P.J.,
637 Hutchison, W.D., Fowler, G., Kalaris, T.M. & Pedlar, J. (2010) Pest Risk Maps for Invasive
638 Alien Species: A Roadmap for Improvement. *BioScience*, **60**, 349–362.
- 639 Wangai, A.W., Redinbaugh, M.G., Kinyua, Z.M., Miano, D.W., Leley, P.K., Kasina, M., Mahuku, G.,
640 Scheets, K. & Jeffers, D. (2012) First report of maize chlorotic mottle virus and Maize Lethal
641 Necrosis in Kenya. *Plant Disease*, **96**, 1582–1582.
- 642 Ward, M. (2016) Action against pest spread—the case for retrospective analysis with a focus on
643 timing. *Food Security*, **8**, 77–81.
- 644 Westphal, M.I., Browne, M., MacKinnon, K. & Noble, I. (2008) The link between international trade
645 and the global distribution of invasive alien species. *Biological Invasions*, **10**, 391–398.

