bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Global cropland connectivity: A risk factor for invasion and saturation by emerging pathogens and pests

Y. Xing[1,2,3,4*], J. F. Hernandez Nopsa[1,2,3,4*], J. Andrade-Piedra[5], F. Beed[6], G. Blomme[7], M. Carvajal Yepes[8], D. L. Coyne[9], G. A. Forbes[10], J. Kreuze[5], J. Kroschel[5], P. L. Kumar[11], J. P. Legg[12], M. Parker[13], E. Schulte-Geldermann[13], and K. A. Garrett[1,2,3,4]

[1] Plant Pathology Department, University of Florida, Gainesville, FL, USA

[2] Institute for Sustainable Food Systems, University of Florida, Gainesville, FL, USA

[3] Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA

[4] Department of Plant Pathology, Kansas State University, Manhattan, KS, USA

[5] International Potato Center (CIP), Lima, Peru

[6] World Vegetable Center, P.O. Box 1010 (Kasetsart University), Bangkok 10903, Thailand

[7] Bioversity International, c/o ILRI, Addis Ababa, Ethiopia

[8] International Center for Tropical Agriculture (CIAT), Cali, Colombia

[9] International Institute of Tropical Agriculture (IITA), Nairobi, Kenya

[10] CIP, Beijing, China

[11] IITA, Ibadan, Nigeria

[12] IITA, Dar es Salaam, Tanzania

[13] CIP, Nairobi, Kenya

*These authors made equivalent contributions

Corresponding author: K. A. Garrett (karengarrett@ufl.edu)

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

The geographic pattern of croplands is an important risk factor for the invasion of crop-specific pathogens and arthropods, and saturation by endemic pests. Understanding the structure of cropland networks supports sampling and mitigation strategies. We evaluated global networks of key vegetatively-propagated crops (banana, cassava, potato, sweetpotato, and yam) of particular importance to food security in the tropics. The risk of damage from diseases transmitted through vegetative propagation is a particular concern. We analyzed the structure of cropland networks for each crop, where the existence of a link between geographic location pairs was determined using a gravity model, as a function of the distance between the pair of locations and the product of the harvested crop area in the two locations. Networks were evaluated using a novel index of pathogen or arthropod invasion and saturation risk, based on the role of locations in bridging cropland areas and the degree of connectedness of a location and its neighbors. For example, in addition to locations with high risk due to high cropping density, locations with high risk because of their role as bridges for cassava include South-Central Nigeria, Central Ghana, and Southwestern Democratic Republic of Congo. For potato, bridges include Central and Southern Poland and Northern Ukraine. The highly-linked hub and bridge locations we identified are likely priorities for surveillance and management, and for tracing intra-region movement of pathogens and pests. Integrated analyses of invasion and saturation risk can simultaneously evaluate risk due to cropland connectivity along with other risk factors such as climate and trade routes.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Plant diseases and pests are major threats to food security (Aguayo et al. 2014; Anderson et al. 2004; Fisher et al. 2012; Gonthier and Garbelotto 2013; Woolhouse et al. 2005). Understanding which geographic areas have a high risk of pathogen and arthropod pest invasion is an important first step to designing sampling and mitigation strategies (Fears et al. 2014). Climate effects are one component of this risk, and commonly addressed in species distribution models (Anderson et al. 2004; Bebber et al. 2013; Elith and Leathwick 2009; Garrett et al. 2006; Garrett et al. 2014; Hernandez Nopsa et al. 2014; Jeger and Pautasso 2008; Kroschel et al. 2016; Rodoni 2009; Rosenzweig et al. 2001). Another important risk component is the structure of trade routes, through which pathogens and pests may also move (Anderson et al. 2004; Bebber et al. 2014). Habitat connectivity represents a third component that, integrated with other risk factors, can provide a more complete invasion risk assessment. The connectivity of cropland regions helps to determine whether invasive species dependent on crops will become established before effective actions can be taken to mitigate them (Margosian et al. 2009; Sutherst 2014). Incorporating cropland connectivity risk with other risk factors for invasion supports a number of integrated pest management (IPM) strategies, from improved methods for detecting and mitigating new invasives, to ongoing improvements in policy (Margosian et al. 2009; With 2004).

The invasion of species into new regions is a common research focus, but less attention is given to the process of saturation (Cornell and Lawton 1992; Fox et al. 2000; Lion and Gandon 2009). Bebber et al. (2014) consider saturation in terms of the fraction of potentially habitable regions that are already occupied by a pest species. Similarly, here we define saturation as the process by which a species “fills in” a region, to occupy more and more of the potential habitat.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Defining the difference between invasion and saturation is often a question of the spatial resolution and extent being considered (Fig. 1). From the standpoint of pathogen and arthropod management, a pest may have already invaded a region and be considered endemic, while at the same time there may be some fields it has never reached, and its population may frequently be suppressed by factors such as extreme weather conditions so that it must "re-saturate". Some pathogens continue to "re-emerge" at different time points (Rugalema et al. 2009; Vurro et al. 2010), such as *Phytophthora infestans* (Fry et al. 2015) For pathogens such as *P. infestans*, initial inoculum can be limiting. For example, abundant initial inoculum probably played a key role in the devastating 2009-2010 epidemic in tomato in the Northeastern US (Fry et al. 2013), while high inoculum associated with synergistic virus species interactions was a key driving factor behind the rapid spread of the cassava mosaic disease pandemic in Africa in the 1990s (Harrison et al. 1997; Legg et al. 2006). High cropland connectivity is a risk factor for saturation, as pathogens and arthropods spread from refugia after limiting weather conditions.

Network analysis offers a number of tools for understanding the strengths and vulnerabilities of network structures. In a geographic network analysis of species invasion or saturation, nodes are geographic locations and the links between nodes represent functions such as the probability of movement of a pathogen or pest between the nodes. Characterizing the network structure of cropland areas acting as sinks or sources can inform the selection of key nodes for surveillance, mitigation, and management improvement. Nodes that are linked to many other nodes (nodes that have high degree) and nodes acting as bridges between cropland regions (nodes with high betweenness centrality) may be particularly important for the spread of pathogens and pests and are important for identifying invasion risk (Hernandez Nopsa et al. 2015; Margosian et al. 2009). Network traits such as centrality (how important a particular node

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

or link is (Newman 2010)), local cohesiveness (how well connected a subset of nodes is compared to their connection to other subsets of nodes (Kolaczyk 2009)), and affinity (degree of tendency for nodes to be linked with other nodes of similar centrality (Barrat et al. 2004)) can help to identify locations in networks that may be priorities for attention.

One problem for evaluating the effect of cropland connectivity, in general or for a particular pathogen or pest, is lack of information to parameterize dispersal risk models. More general risk evaluations can draw on models that have proven useful across multiple systems. Gravity models are frequently used to describe the risk of movement between two locations, in applications including zoology, ecology, and epidemiology (Jongejans et al. 2014). In dispersal events, the risk of movement between two locations is often a function of the product of the amount of inoculum potentially produced at the source location and the amount of potential host material at the sink location (Jongejans et al. 2014; Sutrave et al. 2012b; Xia et al. 2004). Here, we use a gravity model to approximate the dispersal risk associated with a pair of locations (nodes).

Here we evaluate the global cropland connectivity risk associated with five crops of particular importance to food security for smallholder farmers in the tropics. These crops are vegetatively-propagated, with the associated high risk of transmission of disease through seed. Cropland connectivity captures some elements of the risk of transmission through movement of pathogens and pests independent of crop plants (through flight or passive dispersal in wind, for example), as well as risk due to movement of planting materials. The objectives of this analysis are to i) characterize the network structure of global cropland for banana and plantain, cassava, potato, sweetpotato, and yam, ii) evaluate the network structure in terms of its potential impact on pest and disease risk due to dispersal, and develop an index for cropland connectivity risk,

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and iii) use the network structure to identify geographic priorities for surveillance and management of emerging pests and diseases and saturation of endemic species.

Materials and Methods

Study system cropland area

We analyzed global cropland area data representing conditions circa 2000 from Monfreda et al. (2008), a standard reference for the geographic distribution of individual crops. We evaluated the harvested cropland area of five crops: banana, cassava, potato, sweetpotato, and yam (*Dioscorea* spp.). Here we use the term "banana" to refer to bananas and plantains (Beed et al. 2012). Data were spatially aggregated by finding the mean harvested area for each crop across 24 x 24 units of the original 5 min x 5 min units, for a resolution of 120 min x 120 min (2° x 2°). We used two methods to calculate the mean of the crop harvested area per grid: the "land mean" (the sum of the harvested area fractions within an aggregated 2° x 2° unit divided by the number of 5 min x 5 min units within the aggregated unit that contain only land) and the "total mean" (the same sum divided by the total number of 5 min x 5 min units within an aggregated 2° x 2° unit). These two formulations of the mean are different primarily on coastlines and for islands; the sensitivity analysis below includes both formulations. To focus on more important production areas, we considered three threshold values for inclusion of nodes in the analysis: 0.0015, 0.002, and 0.0025 mean proportion cropland harvested area. We described the risk for pathogen and pest movement between each pair of nodes as a function of the distance between the nodes and the cropping density associated with the nodes.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Model of risk of movement between geographic nodes

We constructed adjacency matrices where the entry for each pair of nodes was a function of the distance between the two nodes and the cropping density at the two nodes. The distance effect on the risk was calculated as a function of the Euclidean distance between nodes *i* and *j* (d_{ij}) as either a power law function $d_{ij}^{-\beta}$, or an exponential function $exp(-\gamma d_{ij})$ (Campbell and Madden 1990; Gregory 1968; Madden et al. 2007; Mundt et al. 1999; Mundt et al. 2009; Severns et al. 2014). Thus higher values of the parameters β and γ reflect lower likelihood of long-distance dispersal. The risk due to greater cropland area for any two nodes *i* and *j* was accounted for using a gravity model (Jongejans et al. 2014), by multiplying together the mean cropland area (*c*) associated with each of the nodes ($c_i c_j$). Thus in the first step the weights in the adjacency matrix indicating the overall risk of movement between two geographic nodes were $c_i c_j d_{ij}^{-\beta}$ and $c_i c_j exp^{-\gamma d_{ij}}$ for the power law and exponential functions, respectively. We then applied three threshold minimum values (0.001, 0.0001, and 0.00001) for entries in the matrix and set weights below that to be zero. Network models and metrics for the cropland connectivity for each of the five crops were analyzed for the Eastern and Western Hemispheres. We used the igraph package (Csárdi and Nepusz 2006) in the R programming environment (R Core Team 2016) to evaluate the networks. We used the geographic data analysis and modeling raster package (Hijmans et al. 2015b). Other packages used were rgdal (Bivand et al. 2014b), dismo (Hijmans et al. 2015a), expm (Goulet et al. 2013), maptools (Bivand et al. 2015), rrcov (Todorov 2014), rworldmap (South et al. 2013), mapdata (Becker and Wilks 2014a), sp (Pebesma et al. 2014), maps (Becker and Wilks 2014b), rgeos (Bivand et al. 2014a), and RColorBrewer (Neuwirth 2014).

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Index of invasion risk based on network metrics

We propose a new index to represent the component of invasion and saturation risk based on cropland connectivity. This cropland connectivity risk index (CCRI) is constructed as a weighted mean of the components below, giving half the weight to a measure of the importance of the node as a bridge (betweenness centrality), and half the weight to other metrics that measure how well connected a node and its neighbors are. The choice of weighting was determined to emphasize betweenness, because it will capture a role as bridge that is not obvious when individual cropland area is considered alone, and also to include connectedness of a node at different scales. The index was calculated as

- 1/2 Betweenness centrality (the number of shortest paths crossing a node), scaled by the maximum betweenness.
- 1/6 Node strength (the sum of a node's link weights), scaled by the maximum node degree.
- 1/6 Sum of nearest neighbors' node degrees (sum of the number of links associated with nearest neighbors), scaled by the maximum sum of nearest neighbors' node degrees.
- 1/6 Eigenvector centrality (gives each node a score proportional to the sum of the scores of its neighbors) scaled by maximum.

As an overall measure of cropland connectivity risk for pathogen and pest invasion and saturation for these crops, we also evaluated a general index by adding the individual risk indices. This might be useful for evaluating general purpose surveillance and management strategies to control emerging pests and diseases affecting the set of crops, or for special cases where a pathogen or pest uses multiple host species.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Sensitivity analysis and data quality

We performed an analysis of model sensitivity to parameter shifts, to i) evaluate how consistent results were under changes in model parameters, and ii) determine which nodes had high cropland connectivity risk across all sets of assumptions, and which nodes had high risk only for some model scenarios. Based on the combinations of functions, thresholds and parameters used, 144 cropland connectivity risk index maps were generated for each crop (Table 1). For each cell in the maps, we calculated the mean, max, min, and variance across the 144 maps. For reference, Monfreda et al. (2008) summarize the quality of the data available for their analysis.

Illustration of features captured by the cropland connectivity risk index

In this example, we illustrate the cropland connectivity risk index calculation, and the features of a cropland density map that it captures. A hypothetical map of cropland density is presented for a given crop species (Fig. 2A). In this example, most of the cropland units have a low crop proportion (indicated by yellow shading) while one unit has a high crop proportion (indicated by blue shading). A network (Fig. 2B) is constructed using the gravity model described above, from the corresponding data from the cropland map (Fig 2A), where for this particular illustration an exponential function with $\gamma = 0.7$ was used to calculate the link weight, and a threshold of 0.001 was used to determine whether a link exists. The high crop-density location on the map is represented by the blue node in the network while the other nodes represent the other land units where the crop was present. The sum of nearest neighbors' degrees (Fig. 2C) captures how well connected the nodes' neighbors are. Node strength (Fig. 2D) indicates a node's importance in terms of how connected it is to its neighbors. Betweenness centrality values identify nodes acting as bridges in the network (Fig. 2E). Eigenvector centrality (Fig. 2F) shows how well connected a node is through immediate neighbors, their neighbors, and beyond. Finally, the

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cropland connectivity risk index (Fig. 2G) is the weighted mean of these metrics. In addition to the high- risk locations with high crop density, other locations with high risk because of their role as bridges were identified. The results of a sensitivity analysis for the CCRI in this hypothetical map are also illustrated (Fig. 2H, 2J-L), along with an illustration of the land units where the CCRI rank is higher than the rank based on crop density along (Fig. 2I).

Results

In addition to summaries of which regions have particularly high cropland connectivity risk, the summaries here also indicate the locations where the CCRI rank is substantially higher than the harvested crop fraction. These latter cases are ones where looking at harvest crop fraction alone might not identify areas that have higher risk based on their position in the full network.

Banana and plantain. The combination of high mean CCRI and low variance in CCRI was observed for Central, North Central and Southwest Uganda, Northwest Tanzania, Western Rwanda, and Northern Burundi, and for the Inter-Andean valleys in Colombia and all of Central and Western Ecuador (Fig. 3A). The combination of high mean and low variance indicates that these locations have a high risk across all model assumptions. The CCRI rank was substantially higher than the rank based on cropland density alone in multiple locations in Africa, particularly in Tanzania (Fig. 4A).

Cassava. The CCRI was high in South and Southwest Nigeria and South Central Ghana, Eastern Brazil (states of Bahia and Rio Grande do Norte), Southern Brazil (Parana, Mato Grosso, Rio Grande do Sul States) and Southwest Paraguay (Fig. 3B). The CCRI rank was substantially higher than the rank based on cropland density alone in Nigeria and Cameroon (Fig. 4B).

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Potato. The CCRI was high in North Central Europe, including Northern Ukraine, Central Poland, and Central and Southern Belarus, the United States (Idaho, Washington, Colorado, and the Northern Great Lakes region), Southeastern Brazil (Bahia and Sao Paulo states) and the Andes region in South America (Central and Eastern Andes in Colombia, Ecuadorian Andes, Peruvian and Bolivian Andes, as well as Southern Chile and Eastern Argentina) (Fig. 3C). The CCRI rank was substantially higher than the rank based on cropland density alone in multiple locations in Eastern Europe (Fig. 4C).

Sweetpotato. The CCRI was high in Central China and in the Caribbean (Haiti and the Dominican Republic) (Fig. 3D). Medium risk levels were observed in Tanzania and Uganda. The CCRI rank was substantially higher than the rank based on cropland density alone in multiple locations in China, as well as Tanzania (Fig. 4D).

Yam. The highest CCRI observed for yam was in South Central Nigeria, Benin, Togo, Ghana, and the Ivory Coast, along with the Caribbean Islands (Fig. 3E). The CCRI rank was substantially higher than the rank based on cropland density alone in Eastern Nigeria, Togo, Western Ivory Coast, and the Dominican Republic (Fig. 4E).

Discussion

We have identified areas with high risk for dispersal of pathogens or pests of banana and plantain, cassava, potato, sweetpotato, and yam, based on cropland connectivity. Cropland connectivity is a risk factor for movement though wind, seed exchange, farm tools, or trade. These locations are candidates for prioritizing surveillance and mitigation programs (Smolinski et al. 2003; Woolhouse et al. 2005), especially if information about weather conduciveness to

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

invasion, and other risk factors such as documented trade patterns, also support the high risk designation.

It might seem as if the cropland connectivity risk index would simply identify as high risk those land units which have high crop fraction. That is, why go to the extra effort to evaluate cropland connectivity if simply evaluating cropland density gives the same result? Land units with high crop fraction *will* tend to be identified as high risk, unless they are particularly isolated. However, the cropland connectivity risk index also identifies locations that have an important role as bridges between cropping regions (pink regions in Figs 2I and 4), even if the cropping density within the bridge land units is not particularly high. Thus, analysis of cropland connectivity can identify additional risk areas based on the larger landscape, beyond those identified through a simple unit-by-unit scan for high cropping density.

Some caveats related to this analysis, which are also potentially useful points for future research, include the following. The analysis is based on data (Monfreda et al. 2008) summarized across global data sets that vary widely in quality from region to region. The resolution we selected for our analyses was intended to represent a compromise – avoiding too high a spatial resolution because it might have little data to back it up, and also avoiding too coarse a resolution that might obscure the roles of different regions. Where more complete data are available or can be collected, more detailed and higher resolution analyses can be performed. Likewise, the current analysis does not take into account geographic features that could have important effects on the likelihood of active or passive movement of pathogens and pests, including wind (Sutrave et al. 2012). Roads and rivers may increase pathogen movement, while other water bodies, deserts, and mountains may isolate nodes (Meentemeyer et al. 2012). And the distribution of individual crop species captures only some aspects of risk for many pathogens

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and pests that can use multiple host species. Conversely, if resistance genes are widely deployed, pathogens and pests may only be able to use a subset of the planted fraction (Brown and Hovmoller 2002; Garrett et al. 2017). Extreme weather patterns may be responsible for many important regional or global introductions of pathogens and pests, such as the potential introduction of soybean rust to the US in hurricane Ivan (Schneider et al. 2005). Flooding may move some soilborne pathogens to new locations. Finally, heterogeneity in time may alter patterns of cropland connectivity. Markets may drive longer term trends in planting patterns, and for shorter season crops such as potato and sweetpotato, geographic heterogeneity in planting seasons may disrupt cropland connectivity observed when seasons are aggregated.

These analyses illustrate general cropland connectivity risk across a large spatial extent and for a fairly coarse spatial resolution. Follow-up analyses for specific locations and particular pathogen or arthropod species may be useful, when more detailed data are available for mapping cropland fraction and for selecting appropriate functions to describe dispersal kernels, and potentially other factors. The results presented here have a greater confidence for certain crops such as potato, and certain regions, based on the quality and quantity of the original data available for assembly by Monfreda et al (2008). Examples of the application of network analysis to invasions of particular species include *Phytophthora ramorum* (Harwood et al. 2009) and *Phakopsora pachyrhizi* (Sanatkar et al. 2015; Sutrave et al. 2012). The role of a land unit (node) will depend on the species of pathogen or pest being considered, and its dispersal kernel. A land unit evaluated for species that tend to only move short distances might be isolated, while for species that tend to move longer distances it might be an important bridge node (Calabrese and Fagan 2004). Individual pathogen or pest analyses can also support Pest Risk Analysis working documents, aiding the elaboration of procedures set up by IPPC and implemented by

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

NPPOs – particularly if produced on a regional basis to share information, in addition to the trade implications. (Beed et al. 2013; Miller et al. 2009). Likewise, areas with low connectivity and otherwise low risk can be identified and prioritized for seed production.

Two regions in Africa, the Great Lakes Region and the region between Ghana and Nigeria, have high cropland connectivity risk for multiple crops. EACMV-Ug emerged in Uganda and caused famine in Eastern, Central and Western Africa (Anderson et al. 2004), and wheat stem rust race Ug-99 also in Uganda (Pretorius et al. 2000). It is an interesting open question if the role of Uganda in cropland networks had some influence on disease emergence, or if it was simply a matter of higher sampling effort that made detection more likely.

The movement of pathogens through the international seed trade is an important risk factor for many crops (Anderson et al. 2004; Rodoni 2009; Wylie et al. 2008). In Sub-Saharan Africa, movement of plant material and farming tools is key to the dispersal of banana diseases such BXW (Beed 2014; Tripathi et al. 2009) and through cuttings for the cassava virus diseases, and particularly CBSD (Bock 1994; Legg et al. 2015). The US late blight pandemic in 2009 was caused by the movement of infected tomato plants via trade from a single national supplier (Fry et al. 2013). During 2009-2010, an epidemic of late blight in tomato was reported in Southwest India with the suggestion that the pathogen was introduced via potato seeds imported from the UK and Europe before 2009 (Chowdappa et al. 2013). Cropland connectivity is likely to capture at least a portion of the risk associated with such movement, to the extent that trade and movement of equipment and agricultural workers tends to be through areas that produce a particular crop. Of course, at the same time that cropland connectivity represents a risk for the spread of pathogens and pests, connectivity may also sometimes confer advantages for efficiency in deployment of equipment and personnel, as well as marketing of products.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Additional background related to disease and pest risk

Banana. Important diseases of banana, particularly in the African Great Lakes region are banana Xanthomonas wilt of banana (BXW) and banana bunchy top disease (BBTD) caused by the *Banana bunchy top virus* (BBTV, genus Babuvirus) (Anderson et al. 2004; Beed et al. 2012; Kumar et al. 2015). Panama disease caused by *Fusarium oxysporum* f. sp. *cubense* (*Foc*) Race 1 and 2 is relatively widespread in Africa but with heterogeneous distribution. However, the emergent *Fusarium* Tropical Race 4 pathogen was recently introduced to Mozambique in commercial Cavendish plantations. Plant parasitic nematodes and banana weevils (*Cosmopolites sordidus*) pose additional important threats to banana production. Dispersal risk of these banana diseases, as well as diseases transmitted in other vegetatively-propagated crops, is strongly influenced by movement of planting materials. Movement of banana suckers may be independent of distance at smaller spatial scales, due to the effects of family, friendship, and ethnic ties. Dispersal of certain diseases (e.g. BXW) is maximized by poor agronomic practices (i.e., non- sterilization of farm tools used to harvest and manage banana fields) and lack of adoption of good practices (removal of male flower buds when last banana cluster forms). Wind direction and speed are also very important for dispersal considering their effects on vectors (e.g., aphids carrying BBTV). At short distances, movement via insect vectors is important for the spread of many viral diseases.

Cassava. Cassava's biggest constraints in Africa are virus diseases: Cassava mosaic disease (CMD) caused by Cassava mosaic geminiviruses (CMGs) and Cassava brown streak disease (CBSD) caused by both *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV). Viruses causing both are transmitted by the whitefly vector, *Bemisia tabaci*, and by infected cuttings. CMD is present in all cassava growing areas in Africa

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

while CBSD is confined to Eastern Africa and the Great Lakes region (Legg et al. 2011). There is already ample circumstantial evidence that connectedness is very important for the spread of CMD around Lake Victoria. Interestingly, these two diseases that have the same host and similar cropland areas, have different epidemic patterns, which is a consequence of relative differences in their propensity to be spread through infected cuttings, and differing mechanisms of vector transmission of the causal viruses. In Latin America, other viral diseases affect farmers, such as cassava common mosaic disease (CCMD), caused by *Cassava common mosaic virus* (CsCMV) and transmitted by infected cuttings or mechanically. It can cause yield losses of approximately 30% (Venturini et al. 2016) and recent outbreaks have been reported in Peru and Argentina (Di Feo et al. 2015). Complex viral infections have been associated with disease symptoms in Colombia, as for some other vegetatively-propagated crops (Carvajal-Yepes et al. 2014).

Sweetpotato. Sweetpotato is also the host of important pathogens and pests, including viruses. Several weevils important to yield loss exist worldwide, with different species most abundant in different regions. Some viruses such as SPCSV, SPFMV and some begomoviruses are present globally, whereas other viruses are found only in certain regions, such as SPMMV. They are spread through infected planting material (long or short distances) or arthropod vectors (short distances). Movement of reproductive material (vines) through trade can cover long distances. Some factors can limit the movement of vines such as crop-free periods caused by dry seasons, or temperature cycles. Where cropping is continuous through time, this will maximize the dispersal of inoculum by humans and vectors. Whereas some pests such as viruses are easily spread through planting material (Gibson and Kreuze 2015) and can form permanent reservoirs in wild host species (Tugume et al. 2008; Tugume et al. 2013; Tugume et al. 2016), others, such as weevils are more difficult to spread through planting material (as they mainly affect roots and

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

vine cuttings are the dominant planting material), have no known alternative hosts, and are unable to travel long distances by themselves. Thus both long distance and local spread are much more likely for viruses than for weevils.

Yam. Yam (*Dioscorea* spp.) is a multispecies crop popularly grown for its tubers by millions of smallholder farmers in West Africa. The crop is traditionally propagated using whole tubers or setts cut from tubers. Nearly 94% of the global edible yam production (64 million tons from 7.14 million hectares) is located in West Africa (Benin, Cameroon, Côte d'Ivoire, Ghana, Nigeria and Togo). Nigeria alone produces 66% of global production (FAOstat 2016). *Dioscorea rotundata* (white yam) and *D. cayenensis* (yellow yam) of West Africa origin are the dominant species, whereas introduced Asiatic *D. alata* (water yam) and *D. dumetorum* are other popular food yams in WA. Major constraints to yam production in West Africa are mosaic disease caused by *Yam mosaic virus* and *Yam mild mosaic virus* (genus *Potyvirus*), anthracnose disease caused by *Colletotrichum gloeosporioides*, and damage due to nematodes, *Scutellonema bradys* and *Pratylenchus* spp. (lesion nematode) and *Meloidogyne* spp., (root knot nematodes) responsible for significant pre- and postharvest deterioration of tubers. All these agents are endemic in all the yam production regions in West Africa, however, differential susceptibility was observed between the species. Viruses and nematode spread between generations along with the planting material. Anecdotal evidence clearly links spread of virus and nematodes to the movement of planting material in West Africa.

Acknowledgements

We appreciate helpful input from N. Ramankutty. This research was undertaken as part of, and funded by, the CGIAR Research Program on Roots, Tubers and Bananas (RTB) and supported

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

by CGIAR Fund Donors http://www.cgiar.org/about-us/governing-2010-june-2016/cgiar-fund/fund-donors-2/, the CGIAR Research Program on Climate Change and Food Security (CCAFS), USDA APHIS grant 11-8453-1483-CA, US NSF Grant EF-0525712 as part of the joint NSF-NIH Ecology of Infectious Disease program, US NSF Grant DEB-0516046, and the University of Florida.

References

Aguayo, J., Elegbede, F., Husson, C., Saintonge, F. X., and Marcais, B. 2014. Modeling climate impact on an emerging disease, the *Phytophthora alni* induced alder decline. Global Change Biology 20:3209-3221.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., and Daszak, P. 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends Ecol. Evol. 19:535-544.

Barrat, A., Barthelemy, M., Pastor-Satorras, R., and Vespignani, A. 2004. The architecture of complex weighted networks. Proceedings of the National Academy of Sciences of the United States of America 101:3747-3752.

Bebber, D. P., Ramotowski, M. A. T., and Gurr, S. J. 2013. Crop pests and pathogens move polewards in a warming world. Nature Clim. Change 3:985-988.

Bebber, D. P., Holmes, T., and Gurr, S. J. 2014. The global spread of crop pests and pathogens. Global Ecology and Biogeography 23:1398-1407.

Becker, R. A., and Wilks, A. R. 2014a. mapdata: extra map databases. in: R package version: 2.2-3 http://cran.r-project.org/web/packages/mapdata/index.html.

Becker, R. A., and Wilks, A. R. 2014b. maps: draw geographical maps. in: R package version: 2.3-9 http://cran.r-project.org/web/packages/maps/index.html.

Beed, F., Dubois, T., and Markham, R. 2012. A strategy for banana research and development in Africa. Page 107 in: Scripta Horticulturae, A. Moorhead, ed. International Society for Horticultural Science (ISHS), Leuven, Belgium.

Beed, F., Kubiriba, J., Mugalula, A., Kolowa, H., Bulili, S., Nduwayezu, A., Murekezi, C., Sakayoya, E., Ndayihanzamaso, P., Mulenga, R., Abass, M., Mathe, L., Masheka, B., Onyango, M., Shitabule, E., Nakato, V., Ramathani, I., and Bouwmeester, H. 2013. Processes and partnerships for effective regional surveillance of banana diseases. Pages 210-215 in: Banana Systems in the Humid Highlands of Sub-Saharan Africa: Enhancing Resilience and Productivity. G. Blomme, P. v. Asten and B. Vanlauwe, eds.

Beed, F. D. 2014. Managing the biological environment to promote and sustain crop productivity and quality. Food Security 6:169-186.

Bivand, R., Rundel, C., Pebesma, E., and Hufthammer, K. O. 2014a. rgeos: interface to geometry engine - open source (GEOS). in: R package version: 0.3-8, https://r-forge.r-project.org/projects/rgeos/.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R. J., and Rouault, E. 2014b. rgdal: bindings for the geospatial data abstraction library. in: R package version: 0.9-1, https://r-forge.r-project.org/projects/rgdal/.

Bivand, R., Lewin-Koh, N., Pebesma, E., Archer, E., Baddeley, A., Bibiko, H., Brey, S., Callahan, J., Carrillo, G., Dray, S., Forrest, D., Friendly, M., Giraudoux, P., Golicher, D., Gómez Rubio, V., Hausmann, P., Hufthammer, K. O., Jagger, T., Luque, S., MacQueen, D., Niccolai, A., Pebesma, E., Perpiñán Lamigueiro, O., Short, T., Snow, G., Stabler, B., Stokely, M., and Turner, R. 2015. maptools: Tools for reading and handling spatial objects. in: R package version: 0.8-34, http://r-forge.r-project.org/projects/maptools/.

Bock, K. 1994. Studies on cassava brown streak virus disease in Kenya. Tropical Science 34:134-145.

Brown, J. K. M., and Hovmoller, M. S. 2002. Epidemiology - Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. Science 297:537-541.

Calabrese, J. M., and Fagan, W. F. 2004. A comparison-shopper's guide to connectivity metrics. Frontiers in Ecology and the Environment 2:529-536.

Campbell, C. L., and Madden, L. V. 1990. Introduction to Plant Disease Epidemiology. John Willey & Sons, New York.

Carvajal-Yepes, M., Olaya, C., Lozano, I., Cuervo, M., Castano, M., and Cuellar, W. J. 2014. Unraveling complex viral infections in cassava (*Manihot esculenta* Crantz) from Colombia. Virus Research 186:76-86.

Chowdappa, P., Kumar, N. B. J., Madhura, S., Kumar, M. S. P., Myers, K. L., Fry, W. E., Squires, J. N., and Cooke, D. E. L. 2013. Emergence of 13_A2 blue lineage of *Phytophthora infestans* was responsible for severe outbreaks of late blight on tomato in South-West India. Journal of Phytopathology 161:49-58.

Cornell, H. V., and Lawton, J. H. 1992. Species interactions, local and regional processes, and limits to the richness of ecological communities - A theoretical perspective. Journal of Animal Ecology 61:1-12.

Csárdi, G., and Nepusz, T. 2006. The igraph software package for complex network research. InterJournal, Complex Systems 1695:http://igraph.org.

Di Feo, L., Zanini, A., Rodríguez Pardina, P., Cuervo, M., Carvajal-Yepes, M., and Cuellar, W. J. 2015. First report of *Cassava common mosaic virus* and *Cassava frogskin-associated virus* infecting cassava in Argentina. Plant Dis. 99:733.

Elith, J., and Leathwick, J. R. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Pages 677-697 in: Annual Review of Ecology Evolution and Systematics, vol. 40.

Fears, R., Aro, E.-M., Pais, M. S., and ter Meulen, V. 2014. How should we tackle the global risks to plant health? Trends in Plant Science 19:206-208.

Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., and Gurr, S. J. 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature 484:186-194.

Fox, J. W., McGrady-Steed, J., and Petchey, O. L. 2000. Testing for local species saturation with nonindependent regional species pools. Ecology Letters 3:198-206.

Fry, W., Birch, P., Judelson, H., Grünwald, N. J., Danies, G., Everts, K. L., Gevens, A. J., Gugino, B., Johnson, D. A., Johnson, S. B., McGrath, M., Myers, K. L., Ristaino, J. B., Secor, G. A., and Smart, C. D. 2015. Five reasons to consider *Phytophthora infestans* a re-emerging pathogen. Phytopathology 105:966-981.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fry, W. E., McGrath, M. T., Seaman, A., Zitter, T. A., McLeod, A., Danies, G., Small, I. M., Myers, K., Everts, K., Gevens, A. J., Gugino, B. K., Johnson, S. B., Judelson, H., Ristaino, J., Roberts, P., Secor, G., Seebold, K., Snover-Clift, K., Wyenandt, A., Grünwald, N. J., and Smart, C. D. 2013. The 2009 late blight pandemic in the eastern United States – causes and results. Plant Disease 97:296-306.

Garrett, K. A., Thomas-Sharma, S., Forbes, G. A., and Nopsa, J. H. 2014. Climate change and plant pathogen invasions. Pages 22-44 in: Invasive Species and Global Climate Change L. H. Ziska and J. S. Dukes, eds. CABI, Wallingford.

Garrett, K. A., Dendy, S. P., Frank, E. E., Rouse, M. N., and Travers, S. E. 2006. Climate change effects on plant disease: genomes to ecosystems. Annual Review of Phytopathology 44:489-509.

Garrett, K. A., Andersen, K., Bowden, R. L., Forbes, G. A., Kulakow, P. A., and Zhou, B. 2017. Resistance genes in global crop breeding networks. bioRxiv: https://doi.org/10.1101/106484.

Gibson, R. W., and Kreuze, J. F. 2015. Degeneration in sweetpotato due to viruses, virus-cleaned planting material and reversion: a review. Plant Pathology 64:1-15.

Gonthier, P., and Garbelotto, M. 2013. Reducing the threat of emerging infectious diseases of forest trees - Mini Review. CAB Reviews 8:1-2.

Goulet, V., Dutang, C., Maechler, M., Firth, D., Shapira, M., and Stadelmann, M. 2013. expm: matrix exponential. in: R package version: 0.99-1.1 http://R-Forge.R-project.org/projects/expm/.

Gregory, P. H. 1968. Interpreting plant disease dispersal gradients. Annual Review of Phytopathology 6:189-212.

Harrison, B., Zhou, X., Otim-Nape, G., Liu, Y., and Robinson, D. 1997. Role of a novel type of double infection in the geminivirus-induced epidemic of severe cassava mosaic in Uganda. Ann. Appl. Biol. 131:437-448.

Harwood, T. D., Xu, X., Pautasso, M., Jeger, M. J., and Shaw, M. W. 2009. Epidemiological risk assessment using linked network and grid based modelling: *Phytophthora ramorum* and *Phytophthora kernoviae* in the UK. Ecol. Model. 220:3353-3361.

Hernandez Nopsa, J. F., Thomas-Sharma, S., and Garrett, K. A. 2014. Climate change and plant disease. Pages 232-243 in: Encyclopedia of Agriculture and Food Systems, N. V. Alfen, ed. Elsevier, San Diego.

Hernandez Nopsa, J. F., Daglish, G. J., Hagstrum, D. W., Leslie, J. F., Phillips, T. W., Scoglio, C., Thomas-Sharma, S., Walter, G. H., and Garrett, K. A. 2015. Ecological networks in stored grain: Key postharvest nodes for emerging pests, pathogens, and mycotoxins. BioScience 65:985-1002.

Hijmans, R. J., Phillips, S., Leathwick, J., and Elith, J. 2015a. dismo: species distribution modeling. in: R package version 1.0-12, http://cran.r-project.org/web/packages/dismo/index.html.

Hijmans, R. J., van Etten, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., Lamigueiro, O. P., Bevan, A., Racine, E. B., and Shortridge, A. 2015b. raster: geographic data analysis and modeling. in: R package version 2.3-24, http://cran.r-project.org/web/packages/raster/.

Jeger, M. J., and Pautasso, M. 2008. Plant disease and global change─the importance of long-term data sets. New Phytologist 177:8-11.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Jongejans, E., Skarpaas, O., Ferrari, M., Long, E., Dauer, J., Schwarz, C., Rauschert, E. J., Jabbour, R., Mortensen, D., Isard, S., Lieb, D., Sezen, Z., Hulting, A., and Shea, K. 2014. A unifying gravity framework for dispersal. Theoretical Ecology.

Kolaczyk, E. D. 2009. Statistical Analysis of Network Data: Methods and Models. Springer.

Kroschel, J., Mujica, N., Carhuapoma, P., and Sporleder, M., eds. 2016. Pest Distribution and Risk Atlas for Africa. Potential global and regional distribution and abundance of agricultural and horticultural pests and associated biocontrol agents under current and future climates. International Potato Center (CIP), Lima, Peru.

Kumar, P. L., Selvarajan, R., Iskra-Caruana, M.-L., Chabannes, M., and Hanna, R. 2015. Biology, etiology, and control of virus diseases of banana and plantain. Advances in Virus Research 91:229-269.

Legg, J., Owor, B., Sseruwagi, P., and Ndunguru, J. 2006. Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. Advances in Virus Research 67:355-418.

Legg, J. P., Kumar, P. L., Makeshkumar, T., Tripathi, L., Ferguson, M., Kanju, E., Ntawuruhunga, P., and Cuellar, W. 2015. Cassava virus diseases: biology, epidemiology, and management. Advances in Virus Research 91:85-142.

Legg, J. P., Jeremiah, S. C., Obiero, H. M., Maruthi, M. N., Ndyetabula, I., Okao-Okuja, G., Bouwmeester, H., Bigirimana, S., Tata-Hangy, W., Gashaka, G., Mkamilo, G., Alicai, T., and Kumar, P. L. 2011. Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. Virus Research 159:161-170.

Lion, S., and Gandon, S. 2009. Habitat saturation and the spatial evolutionary ecology of altruism. Journal of Evolutionary Biology 22:1487-1502.

Madden, L. V., Hughes, G., and van den Bosh, F. 2007. The Study of Plant Disease Epidemics. The American Phytopathological Society, St. Paul, Minnesota.

Margosian, M. L., Garrett, K. A., Hutchinson, J. M. S., and With, K. A. 2009. Connectivity of the American agricultural landscape: Assessing the national risk of crop pest and disease spread. BioScience 59:141-151.

Meentemeyer, R. K., Haas, S. E., and Vaclavik, T. 2012. Landscape epidemiology of emerging infectious diseases in natural and human-altered ecosystems. Annual Review of Phytopathology 50:379-402.

Miller, S. A., Beed, F. D., and Harmon, C. L. 2009. Plant disease diagnostic capabilities and networks. Annual Review of Phytopathology 47:15-38.

Monfreda, C., Ramankutty, N., and Foley, J. A. 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Glob. Biogeochem. Cycle 22.

Mundt, C. C., Ahmed, H. U., Finckh, M. R., Nieva, L. P., and Alfonso, R. F. 1999. Primary disease gradients of bacterial blight of rice. Phytopathology 89:64-67.

Mundt, C. C., Sackett, K. E., Wallace, L. D., Cowger, C., and Dudley, J. P. 2009. Long-distance dispersal and accelerating waves of disease: empirical relationships. American Naturalist 173:456-466.

Neuwirth, E. 2014. RColorBrewer: colorbrewer palettes. in: R package version: 1.1-2, http://colorbrewer2.org.

Newman, M. E. J. 2010. Networks: An Introduction. OUP Oxford.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Pebesma, E., Bivand, R., Rowlingson, B., Gomez-Rubio, V., and Hijmans, R. J. 2014. sp: classes and methods for spatial data. in: R package version: 1.0-17 http://rspatial.r-forge.r-project.org/.

Pretorius, Z. A., Singh, R. P., Wagoire, W. W., and Payne, T. S. 2000. Detection of virulence to wheat stem rust resistance gene Sr31 in *Puccinia graminis*. f. sp. *tritici* in Uganda. Plant Disease 84:203-203.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rodoni, B. 2009. The role of plant biosecurity in preventing and controlling emerging plant virus disease epidemics. Virus Research 141:150-157.

Rosenzweig, C., Iglesias, A., Yang, X., Epstein, P. R., and Chivian, E. 2001. Climate change and extreme weather events; implications for food production, plant diseases, and pests. Global Change & Human Health 2:90-104.

Rugalema, G., Muir, G., Mathieson, K., Measures, E., Oehler, F., and Stloukal, L. 2009. Emerging and re-emerging diseases of agricultural importance: why local perspectives matter. Food Security 1:441-455.

Sanatkar, M. R., Scoglio, C., Natarajan, B., Isard, S., and Garrett, K. A. 2015. History, epidemic evolution, and model burn-in for a network of annual invasion: Soybean rust. Phytopathology 105:947-955.

Schneider, R. W., Hollier, C. A., Whitam, H. K., Palm, M. E., McKemy, J. M., Hernandez, J. R., Levy, L., and DeVries-Paterson, R. 2005. First report of soybean rust caused by *Phakopsora pachyrhizi* in the continental United States. Plant Disease 89:774-774.

Severns, P. M., Estep, L. K., Sackett, K. E., and Mundt, C. C. 2014. Degree of host susceptibility in the initial disease outbreak influences subsequent epidemic spread. Journal of Applied Ecology 51:1622-1630.

Smolinski, M. S., Hamburg, M. A., and Lederberg, J. 2003. Microbial Threats to Health: Emergence, Detection, and Response. National Academies Press, Washington, DC.

South, A., Scutt-Phillips, J., Rowlingson, B., Bivand, R., and Foster, P. 2013. rworldmap. in: R package version: 1.3-1, http://cran.r-project.org/web/packages/rworldmap

Sutherst, R. W. 2014. Pest species distribution modelling: origins and lessons from history. Biological Invasions 16:239-256.

Sutrave, S., Scoglio, C., Isard, S. A., Hutchinson, J. M. S., and Garrett, K. A. 2012. Identifying highly connected counties compensates for resource limitations when evaluating national spread of an invasive pathogen. PLoS ONE 7:e37793.

Todorov, V. 2014. rrcov: scalable robust estimators with high breakdown point. in: R package version: 1.3-8 http://cran.r-project.org/web/packages/rrcov/index.html.

Tripathi, L., Mwangi, M., Abele, S., Aritua, V., Tushemereirwe, W. K., and Bandyopadhyay, R. 2009. Xanthomonas wilt: a threat to banana production in East and Central Africa. Plant Disease 93:440-451.

Tugume, A., Mukasa, S., and Valkonen, J. 2008. Natural wild hosts of *Sweet potato feathery mottle virus* show spatial differences in virus incidence and virus-like diseases in Uganda. Phytopathology 98:640-652.

Tugume, A. K., Mukasa, S. B., and Valkonen, J. P. T. 2016. Mixed infections of four viruses, the incidence and phylogenetic relationships of *Sweet potato chlorotic fleck virus* (Betaflexiviridae) isolates in wild species and sweetpotatoes in Uganda and evidence of distinct isolates in East Africa. PLoS ONE 11:e0167769.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tugume, A. K., Amayo, R., Weinheimer, I., Mukasa, S. B., Rubaihayo, P. R., and Valkonen, J. P. T. 2013. Genetic variability and evolutionary implications of RNA silencing suppressor genes in RNA1 of *Sweet potato chlorotic stunt virus* isolates infecting sweetpotato and related wild species. PLoS ONE 8:e81479.

Venturini, M. T., Araújo, T. d. S., Abreu, E. F. M., Andrade, E. C. d., Santos, V. d. S., Silva, M. R. d., and Oliveira, E. J. d. 2016. Crop losses in Brazilian cassava varieties induced by the *Cassava common mosaic virus*. Sci. Agric. 73:520-524.

Vurro, M., Bonciani, B., and Vannacci, G. 2010. Emerging infectious diseases of crop plants in developing countries: impact on agriculture and socio-economic consequences. Food Security 2:113-132.

With, K. A. 2004. Assessing the risk of invasive spread in fragmented landscapes. Risk Anal. 24:803-815.

Woolhouse, M. E. J., Haydon, D. T., and Antia, R. 2005. Emerging pathogens: the epidemiology and evolution of species jumps. Trends in Ecology & Evolution 20:238-244.

Wylie, S. J., Coutts, B. A., Jones, M. G. K., and Jones, R. A. C. 2008. Phylogenetic analysis of Bean yellow mosaic virus isolates from four continents: relationship between the seven groups found and their hosts and origins. Plant Disease 92:1596-1603.

Xia, Y. C., Bjornstad, O. N., and Grenfell, B. T. 2004. Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics. Am. Nat. 164:267-281.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. Components of the cropland connectivity risk index (CCRI) that were varied in sensitivity analyses. Each possible combination of the levels of the values indicated was evaluated. The possible combinations included varying the form of mean (total mean or land mean) and varying the dispersal model (power law or exponential), as well as the parameters of the model selected.

	Method	Parameter	Levels	Interpretation
Total mean	$\sum_{i=1}^{576} c_{5\,\mathrm{mingrid}_i}/576$ $c_{5\,\mathrm{mingrid}_i}$ is the cropland proportion of the i_{th} 5 min grid within a 2 degree grid	$c_{5\,\mathrm{mingrid}_i}$		The sum of cropland proportion of all 5 min x 5 min grids within a 2° x 2° grid is divided by the total number of 5 min x 5min grids aggregated in a 2° x 2° per grid
Land mean	$\sum_{i=1}^{576} c_{5\,\mathrm{mingrid}_i}/(576 - \#water)$ $c_{5\,\mathrm{mingrid}_i}$ is the cropland proportion of the i_{th} 5 min grid within a 2 degree grid	$c_{5\,\mathrm{mingrid}_i}$		The sum of cropland proportion of all 5 min x 5 min grids within a 2° x 2° grid is divided by the total number of 5 min x 5min grids containing only land (5 min x 5 min grids with water are excluded) aggregated in a 2° x 2° per grid; #water denotes the total number of 5 min grids with water
Dispersal Risk Model (DRM) **DRM=$c_i c_j d_{ij}^{-\beta}$**	Power law model $d_{ij}^{-\beta}$ d_{ij} is the distance between nodes i and j c_i is the fraction harvested area with the crop of interest at the i_{th} node	β	β_1= 0.5 β_2= 1.0 β_3= 1.5	Potential changes in model to describe different types of pests and dispersal mechanisms
DRM=$c_i c_j \exp^{(-\gamma d_{ij})}$	Exponential model $\exp^{(-\gamma d_{ij})}$ d_{ij} is the distance between nodes i and j c_i is the fraction harvested area with the crop of interest at the i_{th} node	γ	γ_1= 0.05 γ_2= 0.1 γ_3= 0.2 γ_4= 0.3 γ_5= 1.0	Potential changes in model to describe different types of pests and dispersal mechanisms
Cropland proportion	Minimum cropland proportion for inclusion of node in analysis	p_c	$p_{c1}>0.0015$ $p_{c2}>0.002$ $p_{c3}>0.0025$	Smaller thresholds would result in more nodes retained in the network
Link weight	Minimum link weight for inclusion of link in network	p_l	$p_{l1}>0.001$ $p_{l2}>0.0001$ $p_{l3}>0.00001$	Smaller thresholds would allow more links in the network

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures and figure captions

Figure 1. Contrasting the processes of invasion and saturation over time, where the pathogen or pest species is present in red nodes and absent in green nodes. A. In an invasion, a region is initially free of the species. During the process of invasion, the species enters the region and spreads over time. B. In the process of saturation, a species is already present in a region but not in all potential locations. A restricted subset of nodes in the region may act as refugia for overwintering or oversummering, or for persistence of the species during years with weather less conducive to the species. From these locations, the species can spread to linked nodes when conditions are more conducive. Note that the difference between invasion and saturation is partially a difference in the spatial resolution being considered. For consideration of the risk due to cropland connectivity, even when a species is already present in the region, higher cropland connectivity increases the risk of saturation.

Figure 2 in Box 1. An illustration of the evaluation of cropland connectivity risk for a simple scenario. A. The map of cropland density indicates the fraction harvested area for crop species in a hypothetical small region, where white areas have none of the crop species, yellow areas (1-16 and 18-19) have a low fraction of land planted to the crop species, and the blue area (17) has a high fraction planted to the crop species. B. The network of cropland connectivity that corresponds to the map in A, indicating the links for one set of parameters. The high density region (blue node 17) and bridging region (green node 10) are indicated. C. The map of nearest neighbors' degrees for the network in B. Nearest neighbors are those with direct links to a reference node, and node degree is the number of links to that node. D. The map of node strength for the network in B. Node strength is the sum of the weights of links to a reference node. E. The map of betweenness centrality for the network in B. Betweenness centrality indicates the number of shortest paths in the network that pass through a reference node. F. Eigenvector centrality for the network in B. Eigenvector centrality is a measure of how well connected a node is in terms of immediate neighbors, their neighbors, etc. G. The cropland connectivity risk index (CCRI) for the network in B. CCRI is a weighted mean of the four measures of connectedness in maps C through F. Note that this is the CCRI for one parameter combination. H. The map of the mean CCRI from a sensitivity analysis of networks corresponding to map A for a range of parameter combinations (Table 1). I. The difference between ranked values of the mean CCRI from a sensitivity analysis, and the ranked values in map A. This difference indicates locations where the fraction of land planted to the crop species does not capture all the features of connectivity in the CCRI. Pink indicates regions where the CCRI is higher than the ranked values of map A, and green indicates regions where the CCRI is lower. J. Map of the maximum CCRI from the sensitivity analysis. K. Map of the minimum CCRI from the sensitivity analysis. L. Map of the variance in CCRI across realizations in the sensitivity analysis.

Figure 3. The mean and variance observed for cropland connectivity risk index (CCRI) components in a sensitivity analysis for banana/plantain, cassava, potato, sweetpotato, and yam.

Figure 4. Maps of the difference in cell rank between harvested area fraction and the cropland connectivity risk index (CCRI) for banana/plantain, cassava, potato, sweetpotato, and yam.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

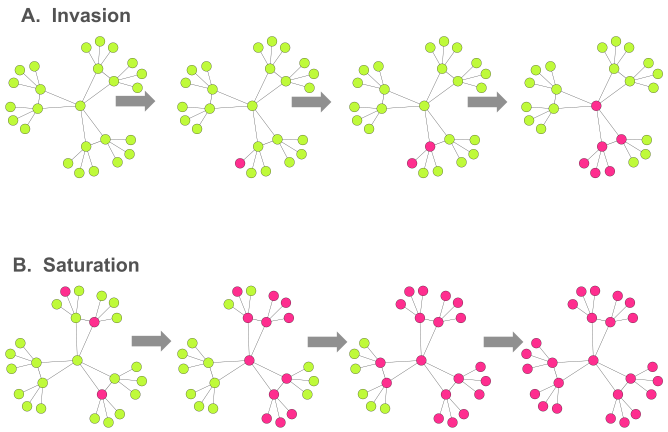

Figure 1.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

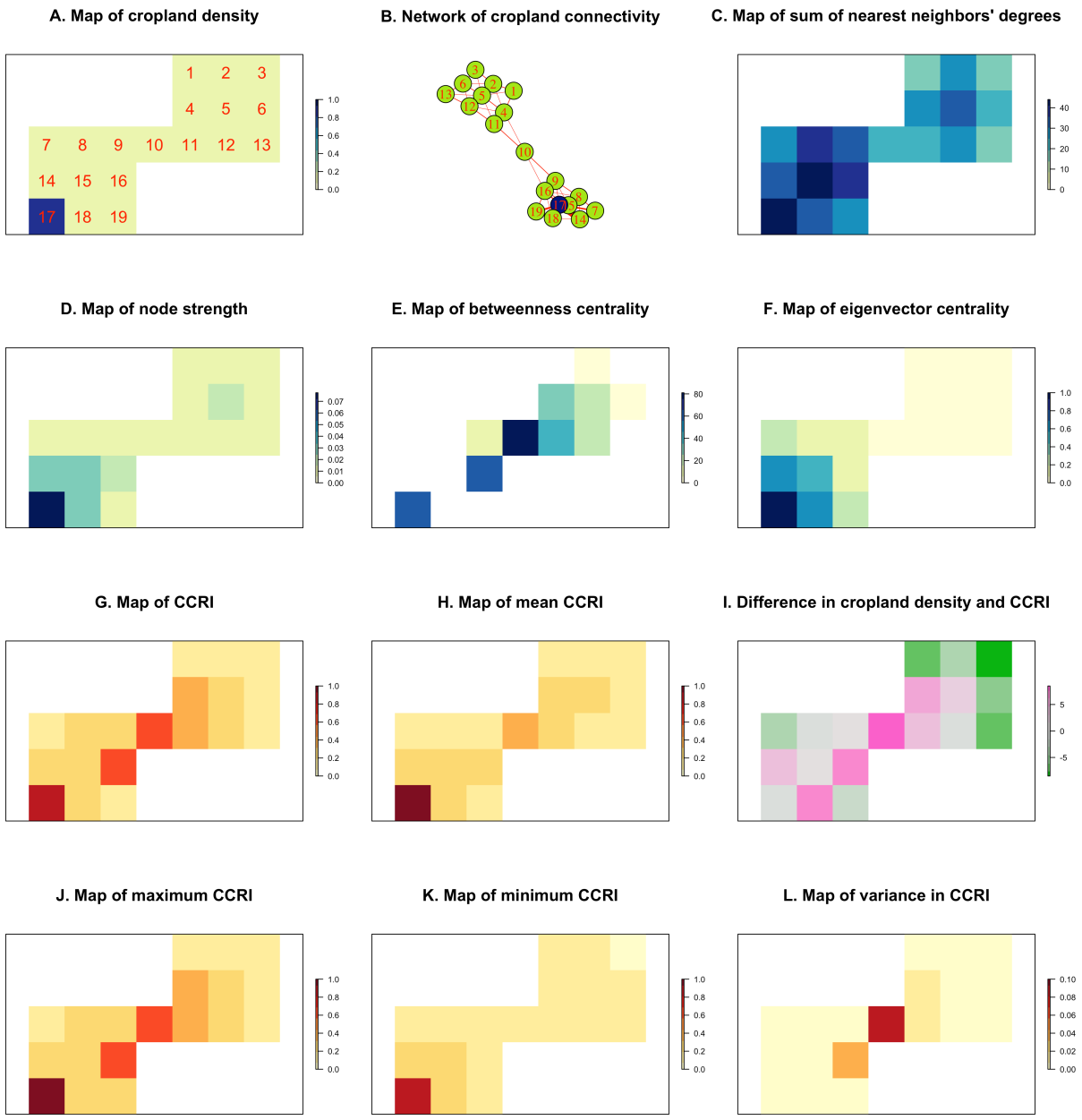

Figure 2.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

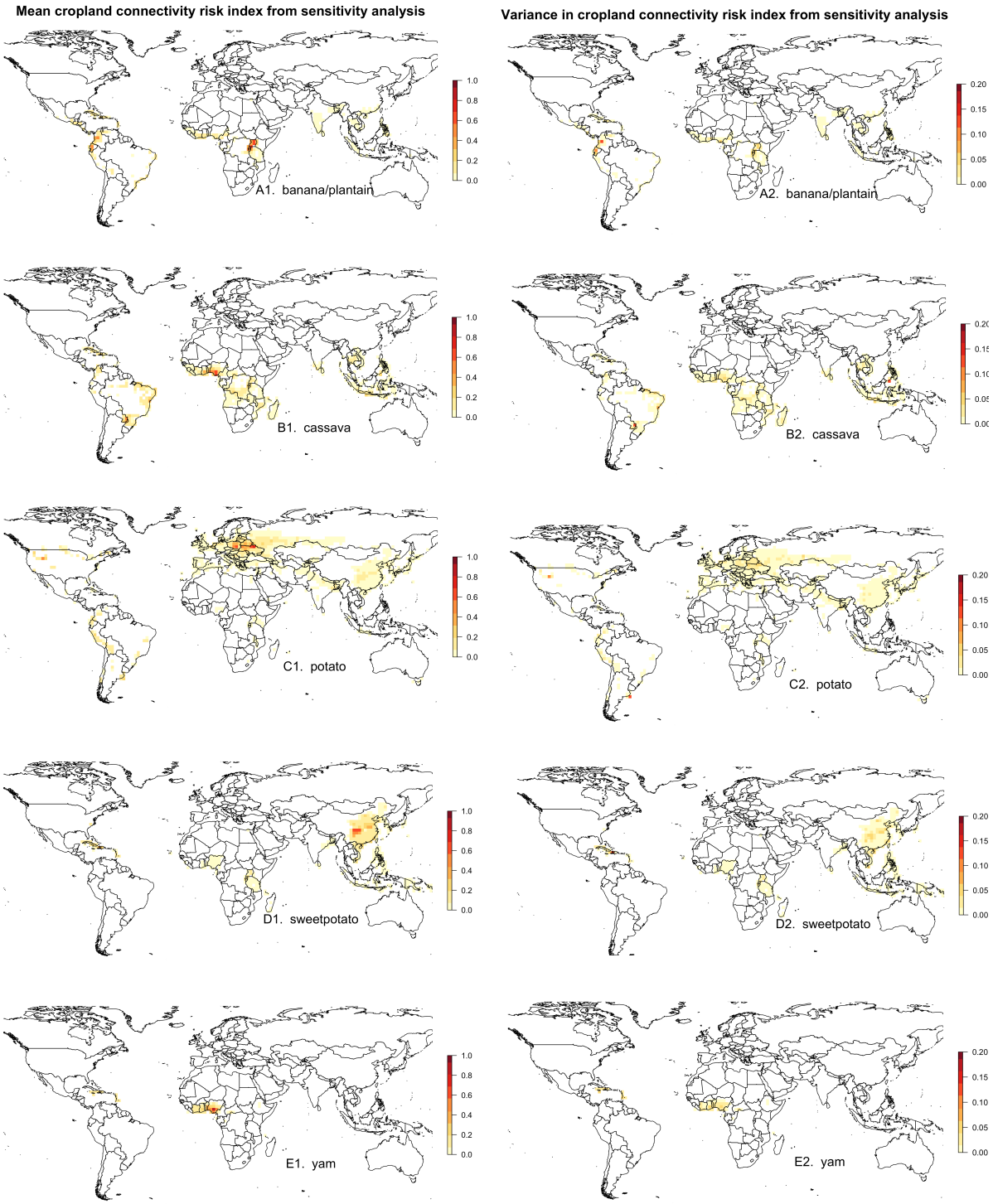

Figure 3.

bioRxiv preprint doi: https://doi.org/10.1101/106542; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

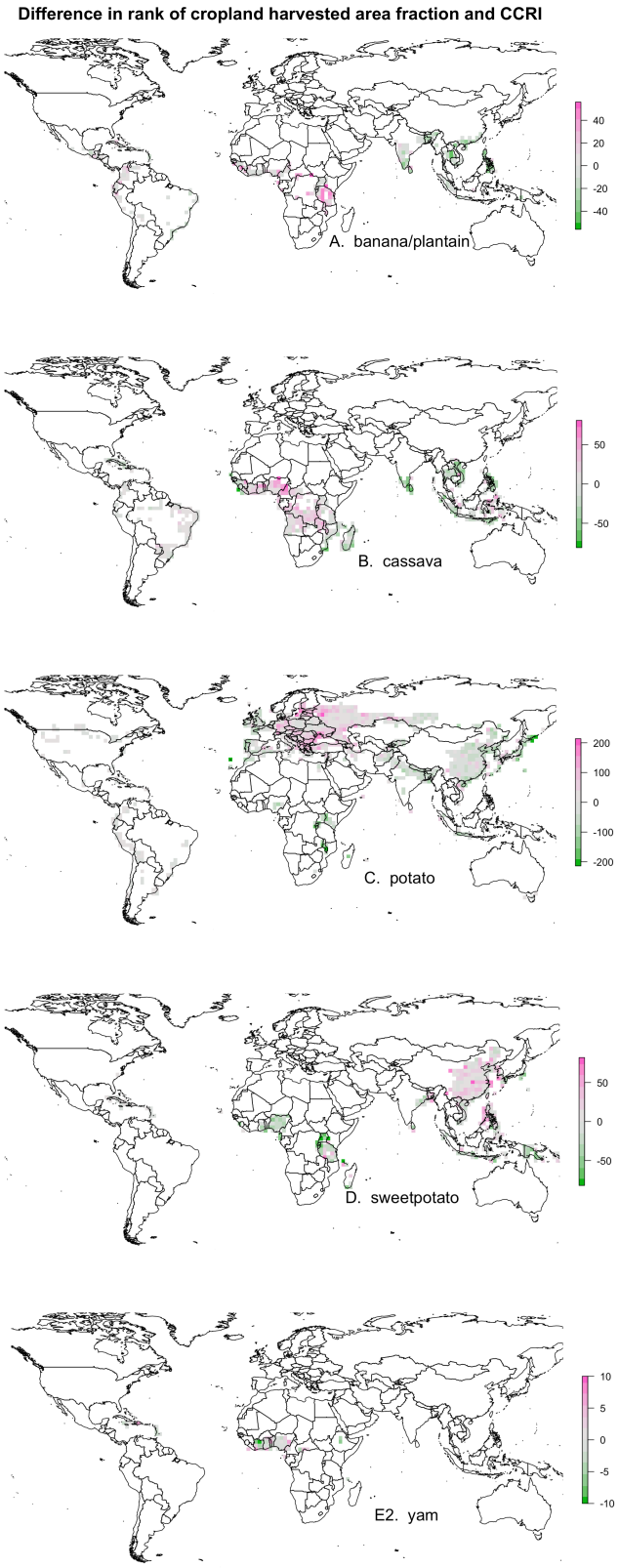

Figure 4.