DATOC
Data Analysis and Tactical Operations Center

Exposure to *Candidatus Liberibacter asiaticus* in Southern California dooryard citrus trees

A briefing prepared for the California Citrus Pest and Disease Prevention Program Science subcommittee

February 20, 2019

Tania Brenes-Arguedas, Brianna McGuire, Holly Deniston-Sheets, Neil McRoberts

Quantitative Biology & Epidemiology Lab, UC Davis

On behalf of the Data Analysis and Tactical Operations Center

Table of Contents

Executive Summary

On behalf of the Data Analysis and Tactical Operations Center, the McRoberts QBE Lab of UC Davis evaluated distribution patterns of huanglongbing (HLB) infection in dooryard citrus trees in the quarantine zone of Southern California. The remit of the project team was to develop an evidence-based description of exposure to *C*Las among populations of citrus trees in the urban landscape of southern California that could potentially be used to guide a program of mandatory removal of plants "exposed" to the pathogen without confirmatory testing.

Data from surveys of citrus trees and Asian Citrus Psyllid (ACP) from 2012 up to June 2018 were available for analysis. These comprised 100,162 individual trees, including 659 with positive HLB infections, and 134,977 ACP detections, including 145 confirmed with *C*Las. The HLB epidemic showed fairly robust infection gradients. Most HLB+ trees were found close to other HLB+ trees. For instance, 80% of all known positive trees occurred within 80m of another known positive; 85% occurred within 110m of another known positive; 90% within 170m; and 95% within 320m. However, the detection rate of additional HLB-positive trees declined with distance from a known HLB+ tree, indicating decreasing exposure with increasing distance.

In selecting an optimal distance for removal, regulators need to balance the need to remove the largest amount of undetected infections against the risk of unnecessarily culling potentially healthy trees and the sheer numbers of trees which would need to be removed. For instance, a clearing radius of 170m around each infected tree would lead to 90% of all known infected trees being destroyed. However, this would require the removal of an estimated 9,184 citrus trees in an area of 22.5 ac. Of these, 8,590 trees are not currently labeled as infected, although the degree of cryptic infection is still unknown. It should be borne in mind that there is considerable geographic variation in disease density, hence a culling program will vary with location in its efficacy in removing infection and the level of economic loss it causes through removal of healthy trees.

Because HLB is a vector-transmitted disease, the presence of a breeding colony of *C*Las+ ACP in a tree is a certain indication that the tree has been exposed to the disease. However, the ACP surveys conducted by CDFA have detected very few *C*Las+ ACP in the region. As a result, there were many cases where *C*Las+ ACP were not detected in the same location as HLB+ trees.

Culling based on exposure distance may not be as appropriate in citrus groves as in urban landscapes. In groves, ACP and HLB tend to accumulate at the edges of the grove. A culling system could focus on removing trees on the edge that are next to a known HLB+ detection. However, given the strength of the edge effects, the risk of moving the border further into the grove may outweigh the potential benefit of culling trees exposed to the disease.

Because healthy and diseased trees accumulate differently with distance from an infected tree, it may be possible to establish a clearing policy that removes most infected trees while minimizing economic damage from removing healthy trees. This work is not intended to be directive, but instead is to be used as a tool to assist stakeholders in decisions regarding the use of culling of exposed trees as a disease curtailing measure.

Introduction

As of February 2019, Huanglongbing (HLB) disease had been confirmed in at least 1,049 dooryard citrus trees in Orange, Los Angeles, and Riverside counties. As the number of known positive trees increases, the number of trees that must be sampled to test for infection increases geometrically, potentially making the testing of infection impractical in the long term. An alternative tactic would be to make use of language in California state agriculture code that allows for the mandatory removal of plants "exposed" to a quarantine pathogen or pest. If the state has a reliable description of exposure that is based on available scientific knowledge and data, such trees could be removed without the need for testing.

In February 2018, the Citrus Pest and Disease Prevention Program (CPDPP) requested the Data Analysis and Tactical Operations Center (DATOC) define what constitutes exposure to HLB infection in the quarantine zone of Southern California. The McRoberts' Quantitative Biology & Epidemiology (QBE) Lab at the University of California, Davis started working on this question. In November 2018, after seeing the first version of the analysis, the Science Committee of the CPDPC requested the inclusion of a section explaining if a similar definition of exposure could be extended to commercial citrus groves.

The QBE lab used data from California Department of Food and Agriculture (CDFA) surveys to empirically evaluate the spatial structure of HLB infection in the region. CFDA has conducted surveys throughout all regions of California where citrus is grown, tracking the presence of Asian Citrus Psyllid (ACP) and HLB-infected trees since 2012, when HLB was first detected in the state. At the time this study was written, HLB-infected trees had only been detected in the urban landscape of Southern California. This study evaluated exposure in urban landscapes by describing the patterns of disease incidence as a function of distance from known infected trees. To determine if a similar system of exposed tree removal could help contain the epidemic in commercial groves, the QBE lab evaluated the literature and analyzed grove data provided by Dr. Mamoudou Sétamou from Texas A&M. This work is not intended to be directive, but instead is to be used as a tool to assist stakeholders in decisions regarding the use of culling of exposed trees as a disease curtailing measure.

HLB patterns in the urban landscape

Methodology

The dataset provided by CDFA was based on collections through June 2018 of citrus tree leaf petioles which were tested for HLB. Although survey activities extend beyond the counties where HLB had been found, this analysis used data only from the region shown in Figure 1, where HLB had been detected.

Since the CDFA dataset did not uniquely identify individual trees, tree counts were estimated based on the sample address and citrus variety. There is some error in this estimate as the actual number of trees present at any given location cannot always be determined (Appendix A). By this approach, we estimate that there are 100,162 individual trees in the area of infection shown in Figure 1; 659 of them tested positive for HLB.

When a tree is confirmed to be positive for HLB (HLB+), CDFA removes the tree. Also, trees are sprayed with insecticides if ACP is detected. Hence, the results of this analysis describe the dynamics of the epidemic under current control measures.

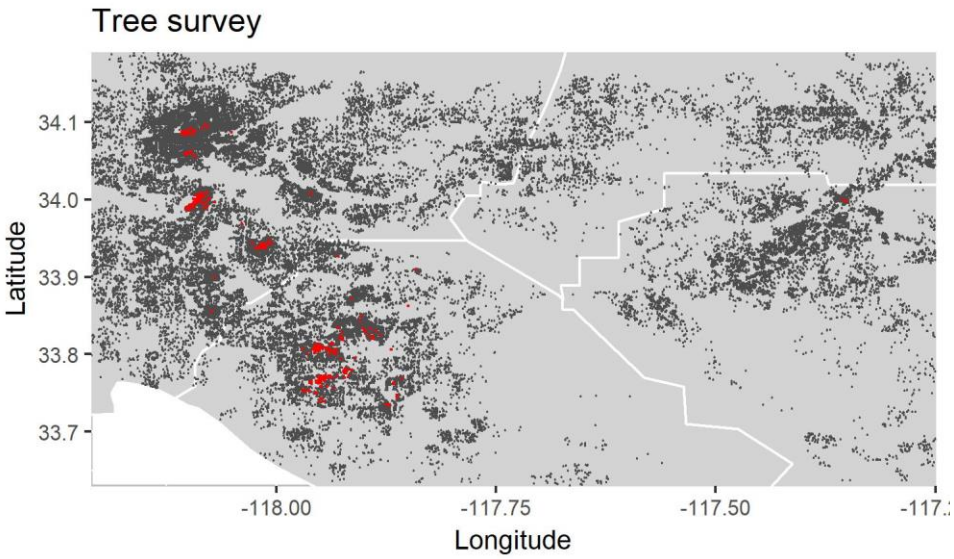

Figure 1. Tree HLB survey data from the Southern California region analyzed in this study. Black dots represent the location of each individual sampling location; HLB+ trees are marked in red dots.

To define exposure in terms of the current spatial structure of the HLB epidemic in the urban landscapes of Southern California, we estimated the probability of finding more infected trees as a function of distance from other HLB+ infected trees. For each individual HLB+ tree (focal tree), we measured the distance to its nearest HLB+ neighbor independent of time of infection. This means that the nearest HLB+ tree could have been infected before the focal tree was monitored or after. No attempt was made to recover the order or timeframe in which infections took place. While understanding the timeframe of transmission could be useful, neither the survey methods or the CDFA database were designed to study this question.

Spatial structure of the HLB epidemic in Southern California

To understand the spatial relationships among infected trees, we did a cumulative distribution analysis[1]. This analysis quantifies the probability of finding another HLB+ infection as a function of distance from a known infection. The resulting distribution is typical of a population with a strong spatial structure, where most HLB+ trees are near each other. For instance, about 30% of the HLB+ trees are at distance = 0m from another infected tree (Figure 2). Coordinates were taken only once per address during the survey, so 0m represents trees on the same property. This means that there is a 30% probability of detecting an HLB+ tree on the same property as another HLB+ tree. It is unknown how far apart different trees are within a single property.

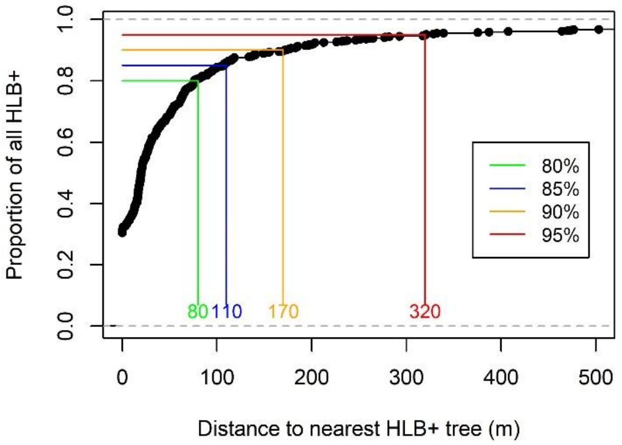

Figure 2. Cumulative distribution of HLB+ trees as a function of their distance to their nearest HLB+ tree. Colored lines highlight 80%, 85%, 90%, and 95% of the HLB+ trees in the region.

In Figure 2, we highlight the distances that would accomplish the culling of 80%, 85%, 90%, and 95% of the HLB+ trees in the region (colored lines in Figure 2). These lines show that 80% of HLB+ trees are within 80m of another HLB+ detection; 85% are within 110m; and 90% are within 170m. After that, there are fewer and fewer trees that are HLB+. It takes an additional 150m to encompass 95% of the HLB+ detections, almost twice what it takes to include 90% of detections.

This information can be used to estimate the impact of culling on containing the spread of the disease. For instance, culling in a circle with an 80m radius around each HLB+ detection would likely eliminate 80% of the undetected HLB+ cases; culling with a 320m radius would

---

[1] To create a cumulative distribution, we sorted all HLB+ trees in order of their distance to their nearest HLB+ neighbor. Then we count HLB+ trees as a function of their separation distance, and convert the counts to proportions (or percent) of the total number of HLB+ trees.

eliminate 95%. This does not, however, consider the number of healthy trees that could be affected by culling.

HLB Detection Rates and Exposure

When healthy trees are exposed to HLB, a proportion of them will become infected, but only a proportion of the infected trees will be detected. Although the qPCR method used by CDFA to detect HLB is very accurate, distribution of the bacteria in the plant is quite heterogeneous, resulting in a hard to quantify proportion of undetected infections. However, it can be reasonably assumed that detection rate is proportional to infection rate, and this in turn is proportional to exposure. In other words, the more HLB+ trees detected at a particular location, the higher the exposure is in that location.

Detection rates are the proportion of HLB+ trees relative to the total tree count. Figure 3 shows total tree counts and HLB+ counts, as a function of distance to an HLB+ tree. Nearly 1400 trees were censused within 20m of all 659 known HLB+ infections in the region; and of these, only 132 were also confirmed to be HLB+ (red bars in Figure 3). This means that the proportion of HLB+ detections within 20m of another HLB+ tree is only 9.5%. If we assume a conservative 25% uncertainty in the count of trees at any given location (Appendix A), then detection rates within 20m would be 7.6 - 12.7%.

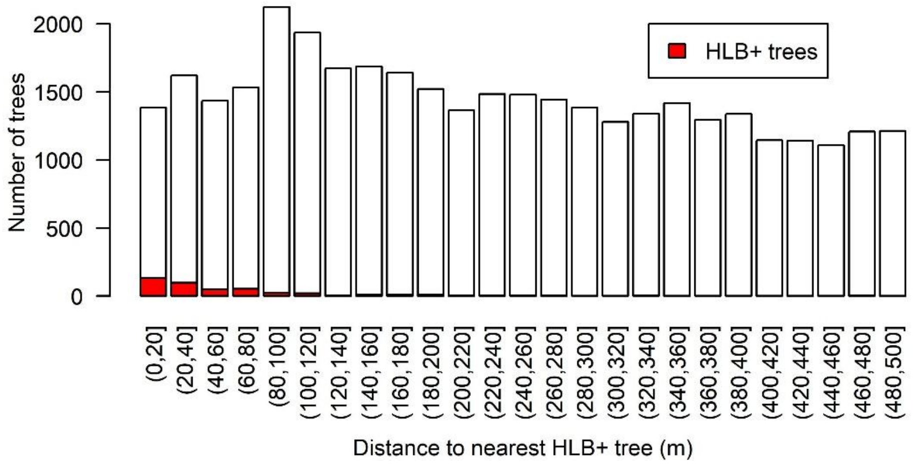

Figure 3. Count of trees at different distances from an HLB+ tree (in meters). Confirmed HLB+ trees are highlighted in red.

In other words, only about one in every ten trees sampled in the immediate vicinity of a known HLB+ tree has also been confirmed as HLB+. Some studies have shown that disease transmission to the ACP vector is inefficient unless the tree is flushing (Sétamou et al. 2016). However, is more likely that low detection rates are explained by the limitations of disease detection. Although qPCR is extremely accurate for true positives, uneven distribution of *C*Las within the tree complicates detection; some researchers have estimated up to a 70% false negative rate in HLB detection from leaf petiole samples. This means that there could be more cryptically infected trees that are currently undetected.

The number of HLB detections decreases with distance (Figure 3, red bars). In other words, fewer HLB detections are found as distance from an HLB infection increases. As a result, HLB+ detection rates decrease rapidly with distance, such that by 40 - 60m they have decreased by more than half, and by 80-100m, the detection rate is only 1% of the trees (Figure 4).

While there is uncertainty in our estimate of detection rates, it is reasonable to assume that if detection rates are higher, incidence and exposure are likely also higher. In this case, higher detection rates close to HLB+ trees indicate higher exposure to the disease. By 100m away from an HLB+ tree, exposure is probably only a tenth of the maximum exposure found in the immediate vicinity (0 - 20m) of an HLB+ tree (Figure 4). Naturally, this is only true if ACP is (or has been) present in the area.

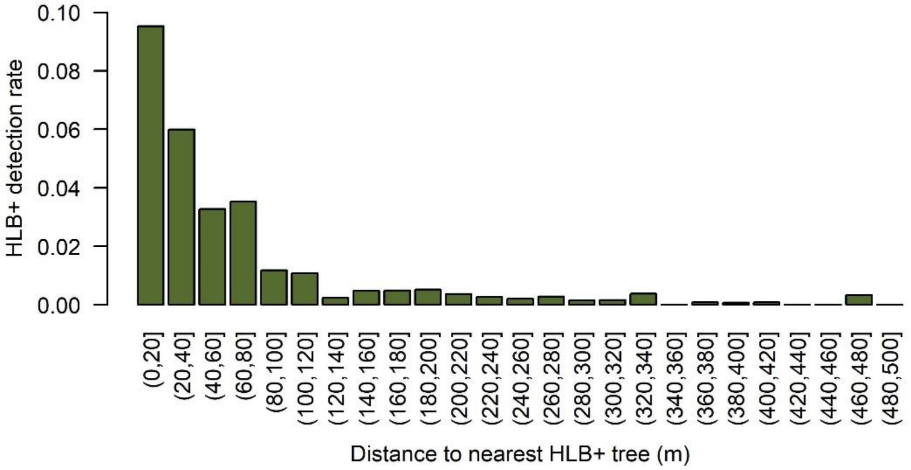

Figure 4. HLB+ detection rate as a function of the distance to the nearest HLB+ tree (meters).

Defining the area of exposure

To implement a culling plan to manage the HLB epidemic, regulators need to balance the need to remove the largest amount of undetected infections against the risk of unnecessarily culling potentially healthy trees. Although increasing distance from an infected tree increases the probability of removing more infections, the exposure and risk of disease decrease with distance. With decreasing exposure, the proportion of healthy trees affected by culling increases. Also, at larger distances, the sheer numbers of trees which would need to be removed also increases as the area of the circles around infected trees increases in proportion with the square of the circle's radius.

For the purpose of visualization, the following figure (Figure 5) shows the area that would be considered exposed if a conservative radius of 170m from each HLB+ tree was chosen (encompassing 90% of known infections). These areas currently have different infection prevalence and density. Two locations have a large number of HLB+ detections (Garden Grove and Pico Rivera) and two have fewer detections (Santa Ana and Riverside). Most 170m rings overlap due to the presence of multiple HLB+ infected trees within each other's exposure radius. However, some infections are relatively isolated, with no HLB+ neighbors in their exposure radius, and many possibly uninfected trees within the exposure areas.

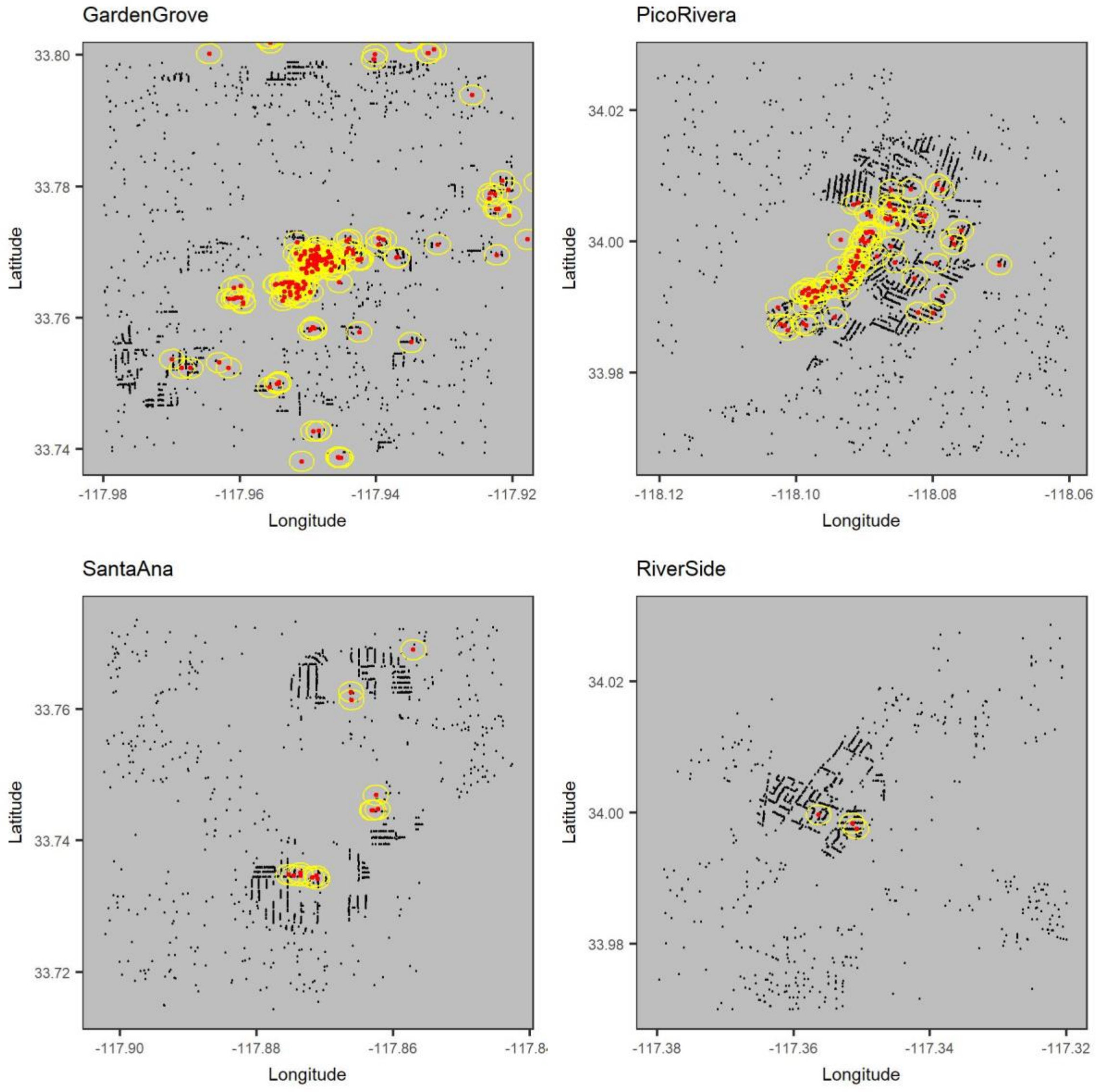

Figure 5. Example of demarcation of 170m exposure areas (yellow rings) in four different infection areas with different incidence and disease density. Red dots represent trees that are confirmed HLB+, and black dots represent trees that have tested negative for HLB.

To estimate the impact of selecting a particular culling radius, the total number of trees inside possible exposure areas, and the proportion of trees within that area that are currently not considered infected, were both calculated. Overlapping exposure rings were merged to avoid counting the same trees multiple times. If all 170m circles around HLB+ trees were merged, for example, the total area that would be considered exposed was 16 km^2 and contained 9,184 trees that would have to be removed. Of these, 8,590 trees (93%) are currently considered uninfected. Other distances are reported in Table 1.

Table 1. Changes in the probability of detection and rate of detection as a function of distance from all current HLB+ trees. The number of HLB+ trees is the same at all distances.

Distance from HLB+ tree (m)	Total exposed area (km^2) [1]	No. of trees in area	No. of currently healthy trees in area	Probability of finding more HLB+ trees	Proportion of trees confirmed HLB+
20	0.59	1163	504	50%	57%
40	1.96	1936	1277	65%	34%
60	3.66	2632	1973	75%	25%
80	5.18	3422	2763	80%	19%
110	8.50	5358	4699	85%	12%
170	16.60	9184	8525	90%	7%
320	38.03	18329	17670	95%	4%

[1] The area is measured by drawing circles around each of the 659 currently known HLB+ detections. Overlapping circles are merged, so area is not counted twice.

Geographical differences in disease incidence

There was considerable variation in infection rates and density of infection throughout the quarantine area of Southern California. To evaluate differences in the density of infection among locations, we counted the number of HLB+ detections inside each 170m exposure rings from each currently known HLB+ tree in the region. In some locations there were over 30 HLB+ trees within 170m from each other, but in many cases there were only 1 or 2 HLB+ trees. Over the whole infection area, there were 67 trees (10% of the detections) that had no HLB+ neighbors within 170m.

Figure 6 shows the number and density of HLB+ detections in each of the 16 cities where HLB has been detected. Infection zones in Anaheim, Garden Grove, and Whittier have a large number of HLB+ detections and high density of infection. In those geographic areas, culling around an infection source can have a large impact in reducing the incidence of the disease, while having a low impact on healthy trees. Other places, like Pico Rivera, are not as densely infected but infections still appear in clusters of two or more trees. Culling will have the least effect in locations with few HLB+ detections or with a lower density of infection. In those cases, most infected trees are isolated and culling will likely affect a larger number of purportedly healthy trees. This information can be used to prioritize disease management plan by the geographic areas based on the density of HLB+ trees.

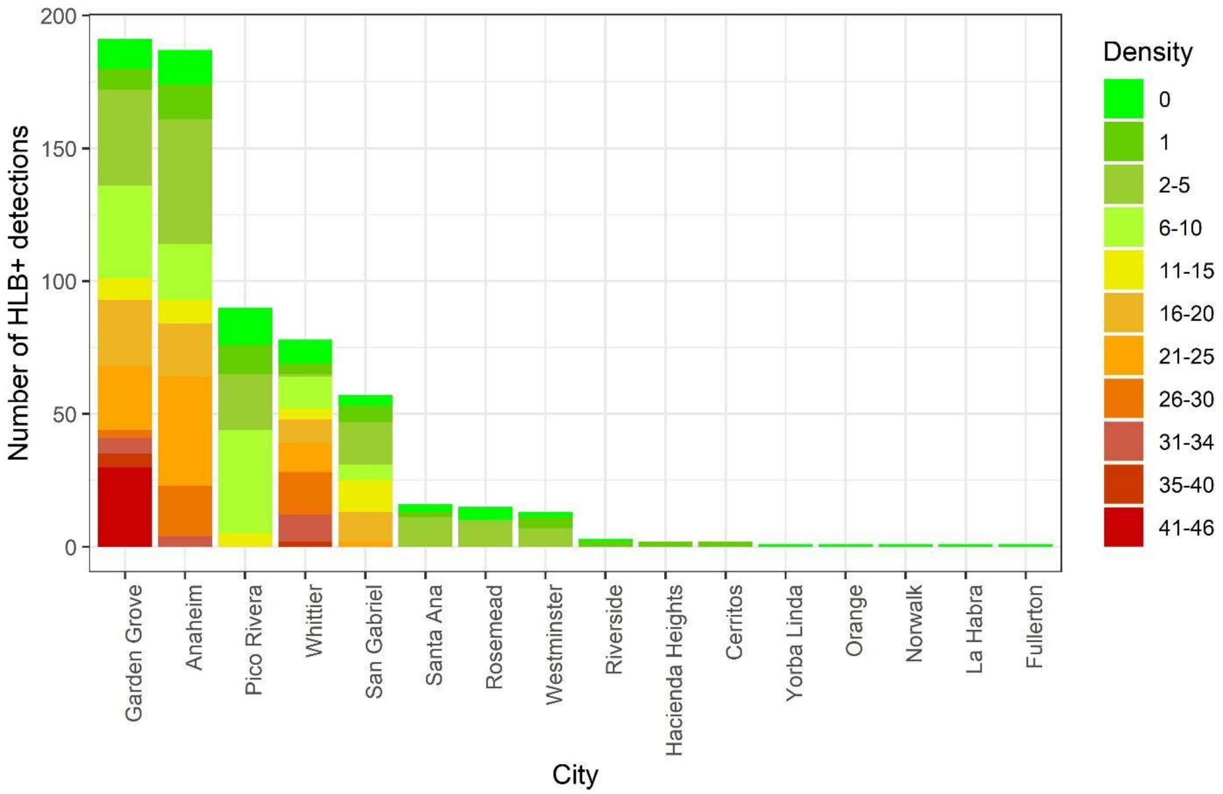

Figure 6. Number of HLB+ detections per city. The bars are color coded to indicate the density of the infections inside a 170m radius around confirmed HLB+ detections.

ACP patterns in the urban landscape

Concurrent with HLB surveys, the CDFA has been tracking the presence of ACP in California dooryard citrus. The QBE lab had access to a dataset provided by CDFA based on collected ACP that were tested for *Candidatus Liberibacter* asiaticus (*C*Las), the bacterium closely associated with HLB infection. Similar to HLB surveys, at the time this analysis was conducted, survey data were available through June 2018. Figure 7 shows the location of all ACP detections in the region where HLB has been detected. Although California legislation, which takes its lead from federal legislation, does not allow for the use of pathogen status of vectors as the basis for regulatory action in plant disease epidemics, vector dynamics remain an important aspect of understanding tree exposure to HLB.

The ACP survey dataset is structured in such a way that we cannot calculate total insect counts at any given location (Appendix A). Hence, instead of insect counts, we addressed ACP population density by counting the number of ACP collection events. Focusing only on the area where HLB has been found, we identified 134,977 ACP collection events; 145 of which had ACP that tested positive for *C*las (*C*las+) (Figure 7).

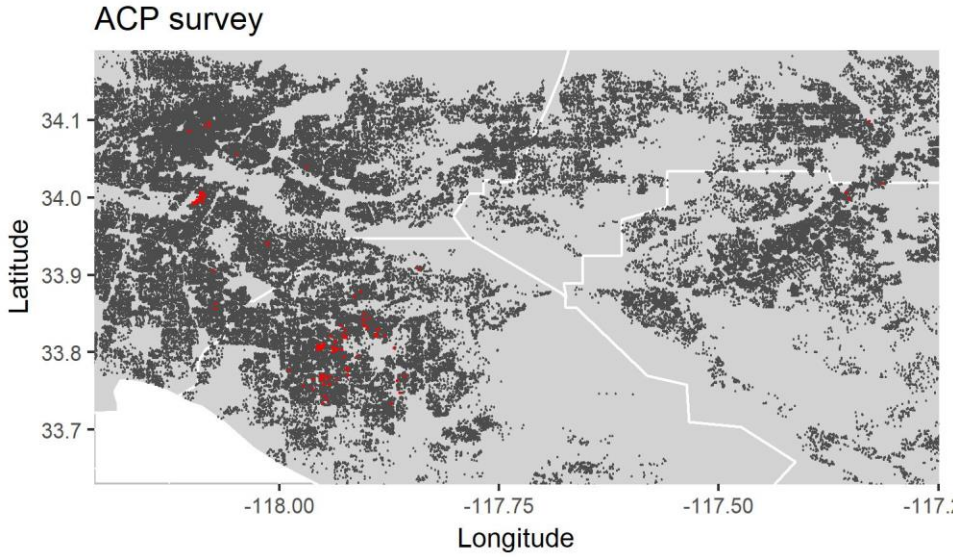

Figure 7. ACP survey data from the Southern California region analyzed in this study. Black dots represent the location of each individual sampling location; C*Las+ ACP are marked in red dots.*

Because HLB is a vector-transmitted disease, the presence of a breeding colony of *C*Las+ ACP in a tree is a certain indication that the tree has been exposed to the disease. However, the ACP surveys conducted by CDFA have detected very few *C*Las+ ACP in the region (145, compared to 659 HLB+ detections). As a result, there are many cases where *C*Las+ ACP were not detected in the same location as HLB+ trees. An absence of *C*Las+ ACP, therefore, should not be considered an indication that no HLB exposure has occurred.

We calculated the distance from each tree in the dataset to the nearest *C*Las+ ACP detection, and evaluated HLB+ detection rates as before. The detection of *C*Las+ ACP in the immediate vicinity of a tree is a good predictor of HLB+ infection. There is a 17.5% probability of detecting an HLB+ tree within 20m of an *C*Las+ ACP (Figure 8). This is nearly twice as high as the probability of detection near another HLB+ tree (9.5%, Figure 4). Similar to HLB, detection rates decrease with increasing distance from *C*Las+ ACP (Figure 8). However, in this case, the decrease is more gradual, probably reflecting the very low frequency of *C*Las+ ACP detections.

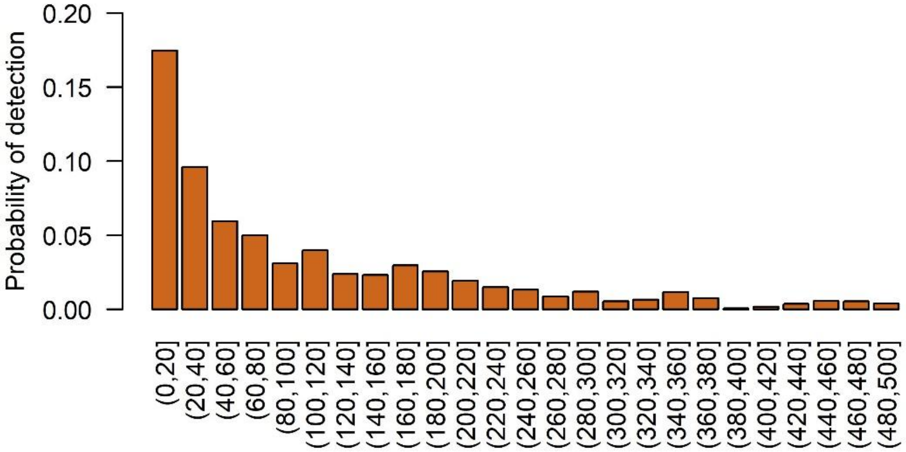

Figure 8. Probability of detecting an HLB+ tree as a function of distance to the nearest infected CLas+ ACP (in meters).

HLB spatial dynamics in citrus groves

In California, HLB has not yet been detected in commercial groves, but the CPDPC requested that the QBE lab also define exposure distances in commercial groves. A wealth of information in the literature regarding HLB dynamics focuses on commercial and in experimental groves in Brazil, Florida, and Texas. Hence, there is more information about the disease dynamics in these systems than in residential citrus. For example, in Brazil, a grove is deemed fully exposed and regulatory action is triggered if over 28% of trees are infected (Craig et al. 2018). In Florida citrus groves, models have been developed to predict disease spread (Parry et al. 2014). A separate task force in DATOC is working on developing control recommendations for growers, and it is not the goal of this study to duplicate those efforts.

One important difference between disease spread in groves and urban landscapes is the edge effect. Numerous studies have documented this effect. For instance, Sétamou and Bartels (2015) showed how, even in the absence of disease control, ACP is more abundant in grove edges. Unpublished data from the same study by Sétamou, shows that HLB infections are also most abundant in the edges (Figure 9). We used this dataset to model the probability of infection from one census to the next as a function of the distance to the nearest HLB+ tree and distance to the edge using a binomial regression. The probability of infection as predicted by the model are shown in Figure 10.

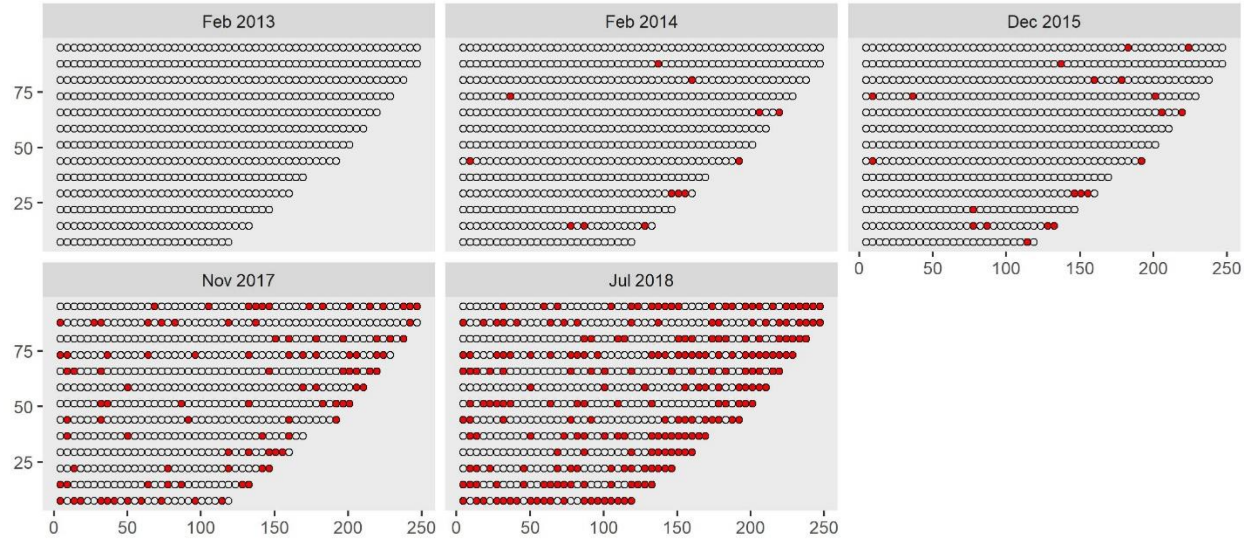

Figure 9. HLB spread over time in an untreated citrus grove in Texas. Distances in both axis are in meters. Red dots represent infected trees based on HLB symptoms and PCR analysis. Data courtesy of: Dr Mamoudou Sétamou, Texas A&M.

Overall, edge effects were more important than neighbor effects. Proximity to an HLB+ tree only increased the probability of infection for trees in the grove edge (Figure 10, A), and only in the 2018 census, when the epidemic was very advanced (40% infection rate). Away from the edge, probability of infection was independent of distance to infected trees in any of the censuses (Figure 10, B). On the other hand, trees in the edge had higher probability of infection than trees farther from the edge (Figure 10, C, D). This effect was observed during 2017, independent of the distance to HLB+ neighbors (Figure 10, C, D). During the 2018 census the edge effect was stronger when close to an HLB+ tree (Figure 10, C), but could not be tested on panel D, as there were few trees farther than 20m from other HLB+ trees.

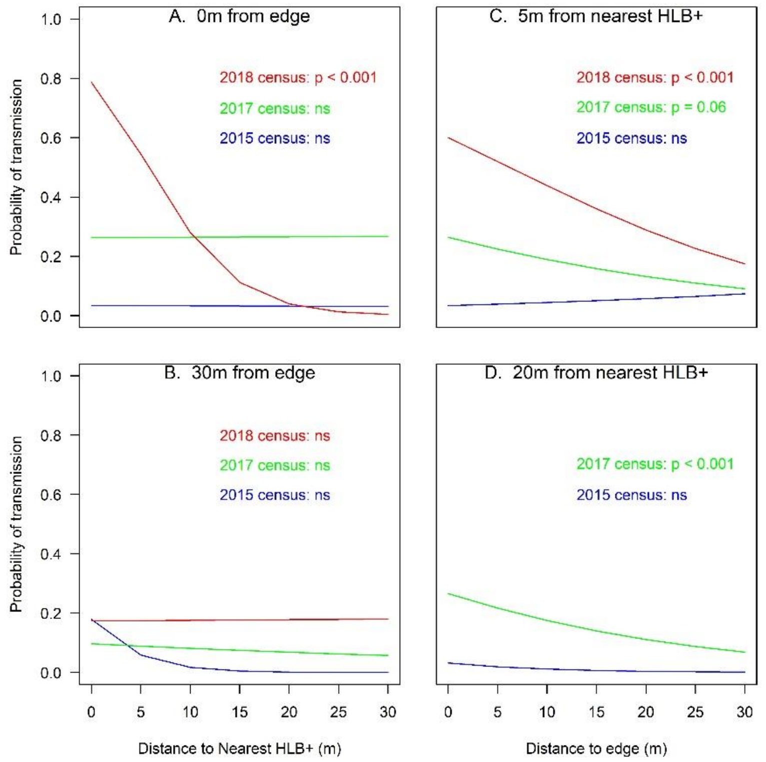

Figure 10. Probability of HLB+ infection transmission as a function of distances to the nearest HLB+ tree in the previous census (A, B), and distance from the edge of the grove (C, D) for the censuses in 2015 (blue), 2017 (green) and 2018 (red). The red line is not included in panel D, because there are few trees left 20m from another HLB+ tree. Data courtesy of: Dr Mamoudou Sétamou, Texas A&M.

Despite clear edge effects, infection rates through the grove are high. By the third census, there is a 10% infection rate in the inner trees, and this goes up to around 20% in the last census. Away from the edge there was very little effect of distance to the nearest HLB+ tree (Figure 10, B) indicating that ACP dispersal is relatively random. This is consistent with other studies showing that psyllid vectors can disperse to trees at considerable distances (Gottwald, 2010).

Conclusions

This report uses data from CDFA surveys to empirically describe disease exposure in the urban and suburban dooryard trees of Southern California. The goal is to provide relevant information that can be used by regulators to make decisions on management of the disease via culling of exposed trees.

Our analysis shows a strong spatial aggregation of HLB infections in urban areas. This means that exposure is highest near another HLB+ tree and decreases rapidly with distance. Selecting the optimal distance for managing the epidemic via culling of exposed trees needs to balance three factors: the probability of capturing the highest number of infected trees, the reduction in exposure away from the infection, and the cost of removal of a large number of trees. We can estimate the probability that culling at certain distances from an infection source would contain the current epidemic. For example, 80% of future HLB+ detections would likely be removed if all trees within 80m of an HLB+ tree were removed; using a 170m radius would likely remove 90% of the epidemic. However, the larger the radius used to define exposure, the greater the number of total trees that would have to be removed, increasing costs. For example, the total number of trees affected by the 80m removal radius is 3,422, while the 170m radius would require the removal of 9,184 trees. Because exposure decreases with distance from an infection source, the benefit of tree removal also decreases with distance. Only about 1% of all trees within 170m of an HLB+ tree are themselves also known HLB+.

Decisions makers should use Table 1 as an aid in the visualization of the pros and cons of each possible culling distance. The selection of culling distance can also consider the differences in the density of infection across the quarantine zone. Some urban areas have a higher number of HLB+ cases than others, and among these areas with high numbers of infection, the density of the infections also changes. Thus, there were some locations that only had a few nearly isolated HLB+ infections, while some locations had over 30 HLB+ trees within a 170m radius of each other. A culling program may decide to prioritize areas by their disease incidence or density.

Interestingly, even in the immediate vicinity of an HLB+ tree, where exposure is highest, HLB+ detection rates were low. In fact, less than one in ten trees within 20m of an HLB+ tree was also confirmed HLB+. Decision makers should keep in mind that there is an unknown degree of uncertainty in our estimates of the detection rates. This uncertainty is due to a number of factors: 1) As individual trees were not identified in the dataset, there could be errors in the tree counts used to estimate detection rates; 2) There are probably more trees in the geographical area that were not sampled due to limited accessibility; and 3) Given the large heterogeneity of *C*Las colonization and proliferation in trees, there could be more undetected and cryptically infected trees. Because exposure decreases with distance, it is possible that cryptic infections are also more common closer to an HLB+ tree.

The presence of *C*Las+ ACP is a better predictor of exposure to HLB infection. Unfortunately, *C*Las+ ACP counts are not a consistent source of information as the overall number of *C*Las+ ACP detections in the region have been very low. In addition, at this time, ACP cannot be used to trigger regulatory action in the form of tree removal by state or national law.

It is the opinion of the QBE lab that culling based on exposure distance might not be as appropriate in citrus groves as it could be in urban landscapes. Based on the information from the literature, and on our analysis of a citrus grove from Texas, the best culling system could focus on removing trees on the edge that are next to a known HLB+ detection. However, this could have the side effect of moving the edge inwards, thereby increasing the risk of infection in rows further inside the grove. Given the strength of the edge effects, the risk of removing the border may outweigh the potential benefit of culling trees exposed to the disease. Aggressive ACP control may be a more effective way of controlling the spread of the disease from a cryptic or very early infection. Further research, outside the scope of this study, is required to identify the best ways to control the disease in citrus groves.

There are considerable social and budgetary concerns to account for in determining if and how to implement a tree removal program based on this description of exposure. While tree removal may eliminate the anticipated local infection, it may also remove a high number of healthy trees, which could have social and economic implications that this report does not attempt to quantify.

Literature cited

Craig, Andrew P., Nik J. Cunniffe, Matthew Parry, Francisco F. Laranjeira, and Christopher A. Gilligan. 2018. “Grower and Regulator Conflict in Management of the Citrus Disease Huanglongbing in Brazil: A Modelling Study.” *Journal of Applied Ecology* 55 (4): 1956–65. https://doi.org/10.1111/1365-2664.13122.

Gottwald, Tim R. 2010. “Current Epidemiological Understanding of Citrus Huanglongbing.” *Annual Review of Phytopathology* 48 (1): 119–39. https://doi.org/10.1146/annurev-phyto-073009-114418.

Parry, M., G. J. Gibson, S. Parnell, T. R. Gottwald, M. S. Irey, T. C. Gast, and C. A. Gilligan. 2014. “Bayesian Inference for an Emerging Arboreal Epidemic in the Presence of Control.” *Proceedings of the National Academy of Sciences* 111 (17): 6258–62. https://doi.org/10.1073/pnas.1310997111.

Sétamou, Mamoudou, Olufemi J. Alabi, Madhurababu Kunta, John L. Jifon, and John V. da Graça. 2016. “Enhanced Acquisition Rates of ‘ Candidatus Liberibacter Asiaticus’ by the Asian Citrus Psyllid (Hemiptera: Liviidae) in the Presence of Vegetative Flush Growth in Citrus.” *Journal of Economic Entomology* 109 (5): 1973–78. https://doi.org/10.1093/jee/tow171.

Sétamou, Mamoudou, and David W. Bartels. 2015. “Living on the Edges: Spatial Niche Occupation of Asian Citrus Psyllid, Diaphorina Citri Kuwayama (Hemiptera: Liviidae), in Citrus Groves.” *PLOS ONE* 10 (7): e0131917. https://doi.org/10.1371/journal.pone.0131917.

Appendix A

Data preparation and analysis methods

Description of the data

The data analyzed in this study includes two datasets, the tree HLB survey and the ACP survey. Survey activities used here were completed by CDFA from October 2012 to June 2018. Both datasets include a combination of CDFA activities such as risk-based surveying and delimitation surveying. To facilitate analysis, the full dataset was cropped to the geographic area where infection is found, specifically to the region between longitudes -118.2077 and -117.2508, and latitudes 33.63420 and 34.19627.

For the spatial analysis conducted here it is important to identify individual trees, their location and infection status. In the CDFA dataset, some trees were sampled repeatedly over time. In addition, since December 2017, some trees were sampled in quadruplicate to increase the chance of detecting HLB infections. If repeated samples of the same tree are erroneously assumed to be different trees, this can skew the tree counts making them seem larger. Unfortunately, individual trees are not tracked or tagged at each property. Without individual tree identifiers, we resorted to identifying trees by address and citrus variety (i.e., mandarin, lemon, orange). Hence, in some cases it was not possible to know with precision the actual number of trees sampled at any given location.

Before identifying individual trees by location, data were cleaned to ensure standardized naming conventions amongst varieties, cities, and addresses. It is important that these be standardized as much as possible, as spelling errors can erroneously result in two samples from the same tree/location appearing as different trees/locations. For example, a “101 Bethany Rd” is the same as “101 Bethany Road”, however, these would be considered two different trees because of the differences in spelling of the address. Similarly, “Blood Orange” would be considered different from “Blood Oranges”. Because, the original dataset had 233,521 records this cleaning cannot be done visually, and is limited to what can be corrected with data cleaning tools and scripts. Though most obvious errors in spelling and naming were corrected, it is estimated that approximately 0.01% of the dataset may still contain errors.

We identify individual trees by their address and variety. Multiple samples from the same address (from the same or from different dates) are all considered a single tree unless they are different varieties. Hence, if a lime tree has been sampled four times (e.g., quadrant samples) those four records will have the same address and variety and will be correctly counted as one single lime tree. Unfortunately, if a property has multiple trees of the same variety, they cannot be differentiated from one another. For example, if a property contained two mandarins and four limes, our approach identifies them as only two trees, one mandarin and one lime. Hence in properties with multiple trees, there is a potential to underestimate the tree count. On the other hand, any remaining typos in the entry of addresses or varieties could result in overestimating the number of trees.

We have no way to quantify how much error there is in our tree counts based on this approach. However, we identified that about 75% of the addresses had only one sample. Any error differentiating multiple trees in one property could only occur in the other 25% of the samples. Also, we corrected spelling errors in the address in 11% of the records. Thus, for the purpose of quantifying errors, we chose the largest value and assume that total tree counts have an error of ± 25%. Table 2 shows the total counts of records in the dataset and the count of trees after data cleaning and manipulation.

Table 2. Total number of records, tree counts, and ACP detection events, in the two CFDA survey datasets.

	----------- Number of samples -----------			
	Trees	HLB+	ACP	CLas+
Original dataset	233,521	701	318,646	173
Cropped region	168,101	701	189,606	169
Grouped by location	100,162	659	134,977	145

The ACP dataset did not require the same cleaning protocol, as it was only necessary to identify the location of individual ACP detection events. CDFA protocol dictates that ACP adults discovered at a given property are pooled for all trees in the property, but ACP nymphs (the juvenile form of ACP) are counted separately per tree. That means that when there are multiple trees in a property, there are multiple records for that collection event: one for each tree/nymph collection. All nymphs and adults discovered at every property are further summed into a single “Insect count” per location. However, this insect count value is repeated for all records in that property. Thus, insect counts are often duplicated in properties with multiple trees. For example, say 12 nymphs and four adults were found on a property with two trees. Each tree appears as a single record with six nymphs, four adults and 16 total insects. It would be incorrect to assume that there were eight adults or 32 insects collected at that location. Hence, the sum of all insects does not always equal the individual counts of nymphs and adults. To avoid summing duplicated insect counts as if they were separate events, this study instead evaluates the number of ACP detections independent of the number of insects collected. Individual ACP detection events were identified by their coordinates and date of collection. Hence, each location and date in the dataset was counted as a single event of ACP adult or nymph detection, without analyzing the number of ACP detected in that event.

One important limitation with the ACP dataset is that it is not corrected by collection effort. In other words, if there was a peak in ACP detections, it is not possible to differentiate if this was caused by actual increases in ACP numbers in the area, or by an increase in collection effort (i.e., more samplers on the ground).

Some changes in the data collection and entry methodologies in the future may resolve some of issues that required cleaning for this analysis. These may include recording quadrant samples at their times of collection, or quantifying both adults and nymphs per tree at each address of detection.

Analytical methods

For each individual tree location in the region (focal tree) we identified the infection status (if an infected tree was detected at that location at any time) and the distance to the nearest HLB+ tree or location. In addition, we calculated the distance to the nearest CLas+ ACP. Nearest neighbors (trees or ACPs) were identified using the spDistsN1 function from the R sp package.

The analysis was done independent of time. Hence, the nearest infection could have been the source of the disease on the focal tree or could have become infected after the focal tree. This distinction in the direction of the infection is not important for our conclusions.

Similar nearest neighbor analysis has previously been used by stakeholders in the California citrus community, most notably by the Gottwald Lab to refine delimitation zones for CDFA surveys around positive finds.

Appendix B

Supplemental Analyses

Timeline of infection

HLB+ infection was first detected in the city of Hacienda Heights in Los Angeles county in 2012, and has been spreading very rapidly since 2017 (Figure 11). CLas+ ACP detections are less frequent than HLB+ trees but have also been spreading over time (Figure 12). Figure 11 and Figure 12 show that the locations of CLas+ ACP detections do not always match the locations of HLB+ detections.

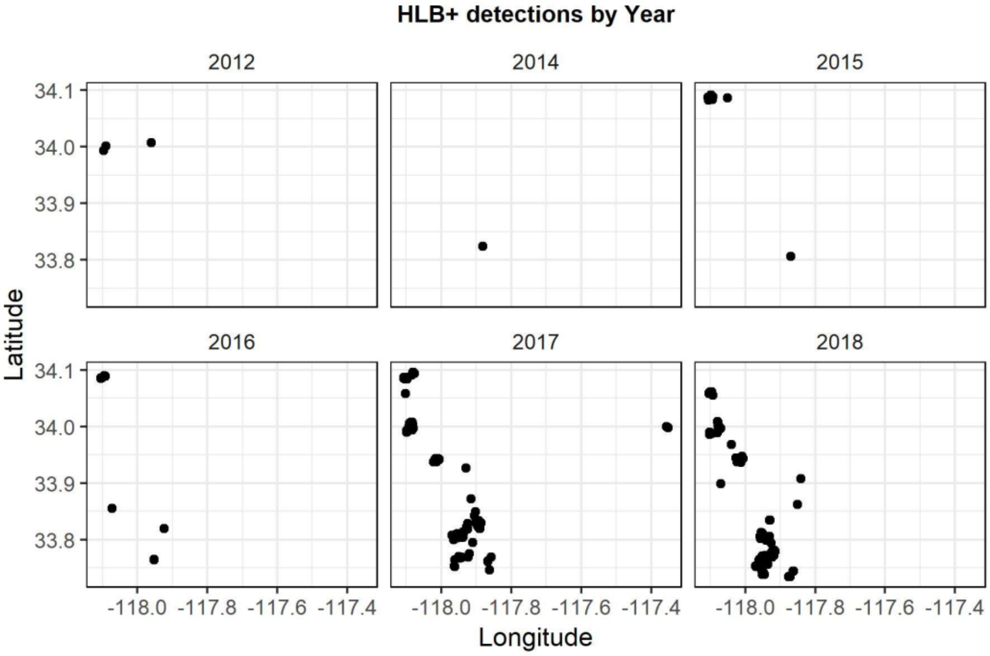

Figure 11. Detections of HLB+ infection in trees in Southern California by year.

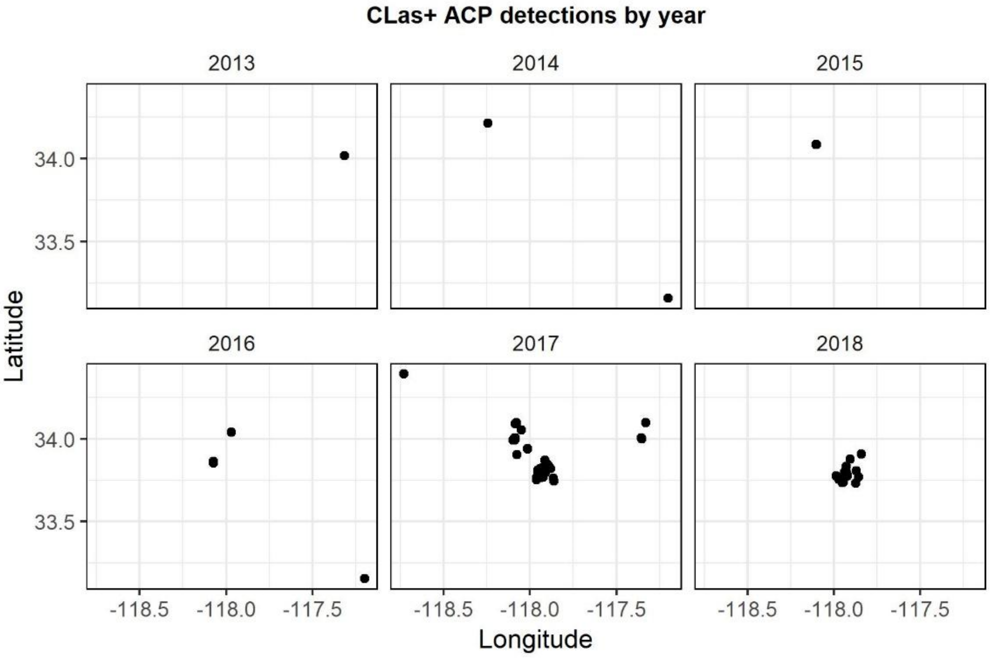

Figure 12. Detection of CLas+ ACP in Southern California by year.

Here we evaluate the number of ACP detections over time. An ACP detection is any event were ACP (adult or nymph) was collected, independent of how many insects were collected. Because this is a CLas testing dataset, it only includes cases were ACP was collected. This means that the counts are not corrected for collection effort. In other words, if there is a peak in ACP detections, it is not possible to differentiate if this was caused by actual increases in ACP numbers in the area, or by an increase in collection effort (more samplers on the ground). Figure 13 shows the timeline of ACP detections and CLas+ infection for three regions in the infected area. There are many months with no data. It is possible that means that there were no ACP finds during those dates, however, that is not specified in the data. These timelines show cases were ACP was present in the area years before there were detections of HLB+ trees. Also, in the case of the city of Santa Ana, CLas+ ACP has not been detected, despite the presence of HLB+ trees.

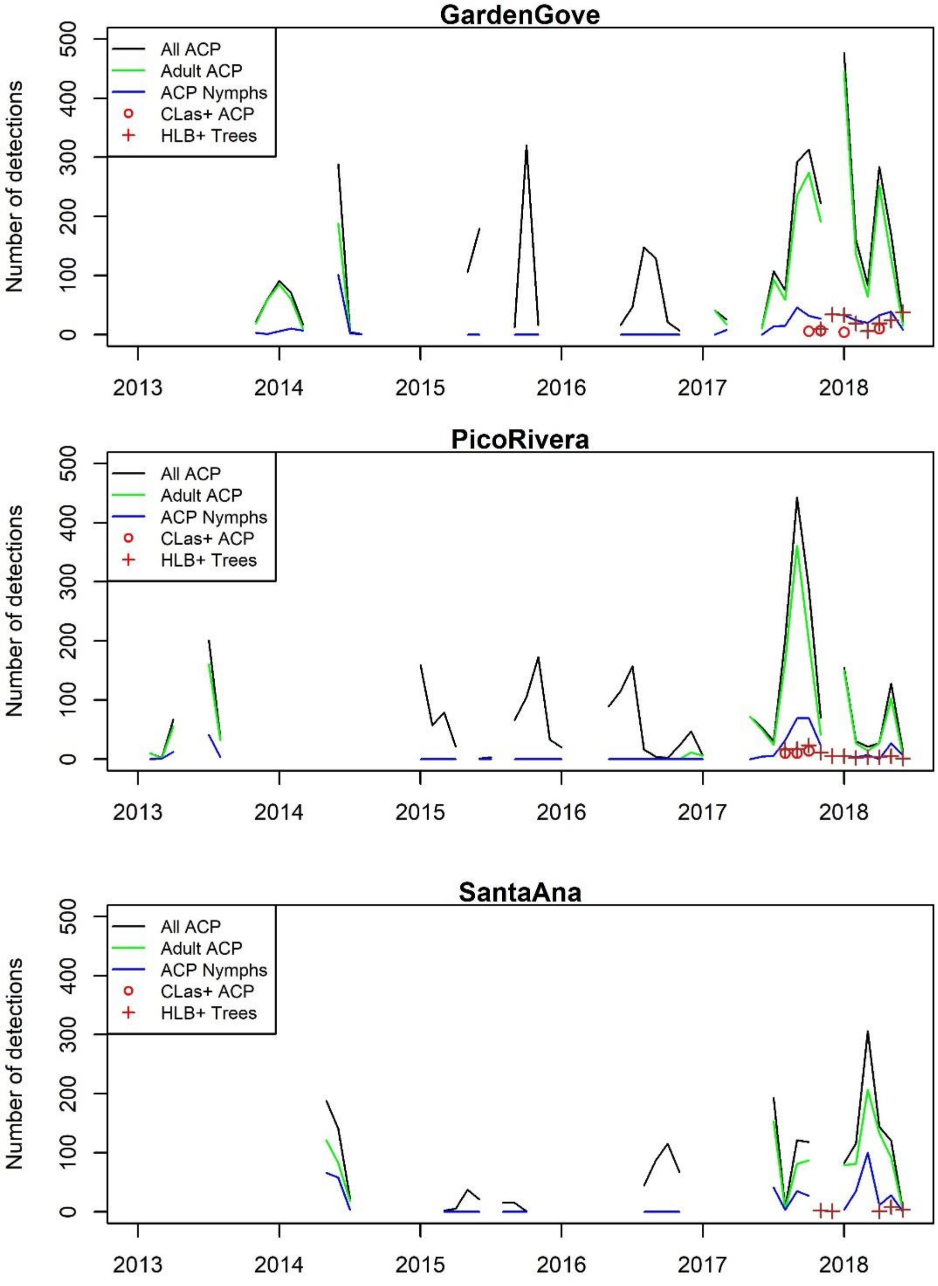

Figure 13. Timeline of ACP and HLB detections per month not corrected for collection effort.

Limitations in survey coverage

Figure 14 shows the proximity of multiple HLB+ trees (shown in black) to CLas- ACP and HLB- trees in the densest zone of infestation found in Garden Grove. This also displays the potential lacking data in the dataset. There were few trees measured in inset oval on the maps, yet many ACP are discovered there, indicating that there are likely far more citrus trees in Southern California that are not a part of this CDFA dataset. As stated in the main body of the report, there was no control in our analysis over sampling effort or sampling error, and figures such as those below indicate that sampling may be heavily heterogeneous in ways that may have impacted our analysis.

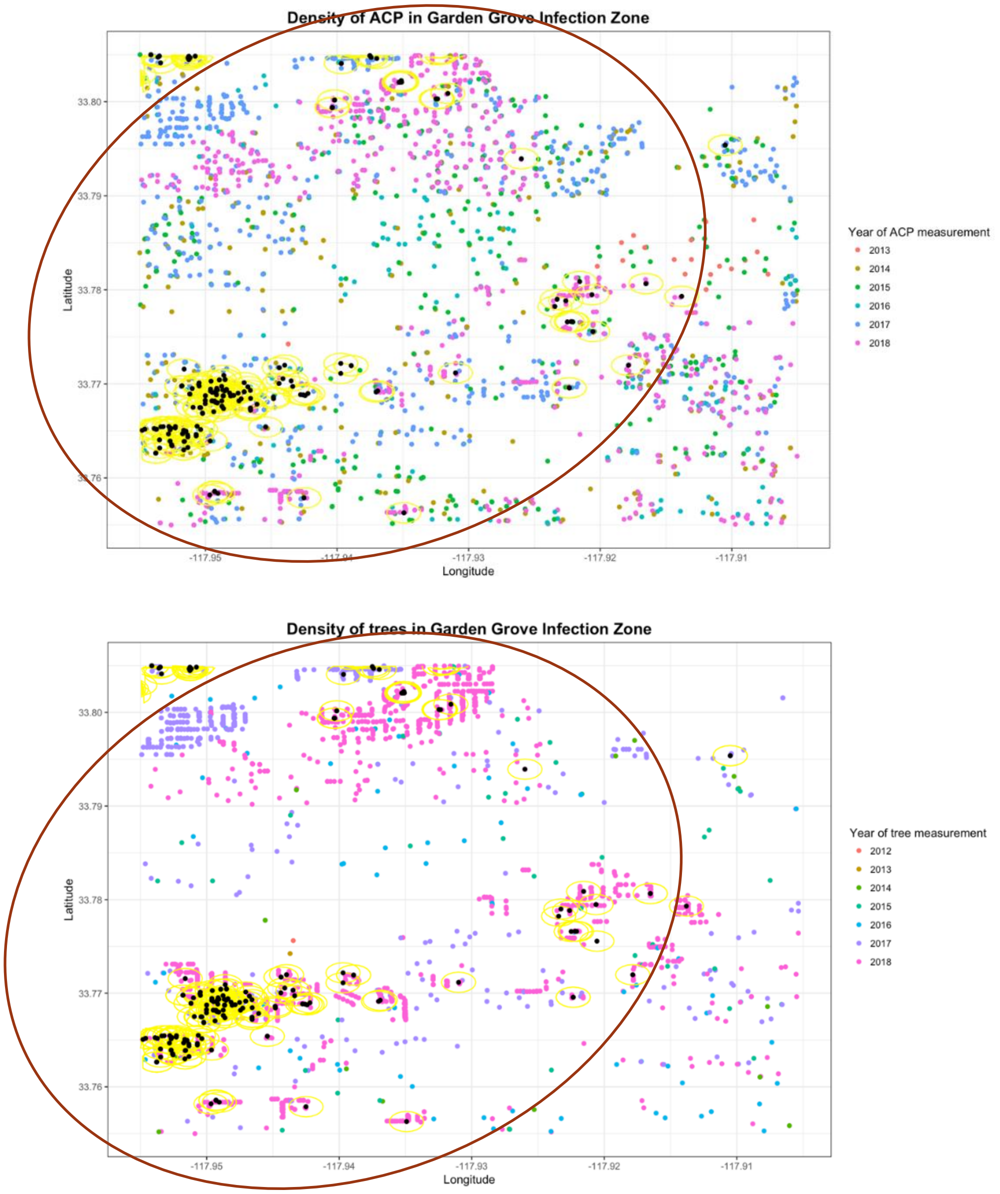

Figure 14. Location of ACP (top) and trees (bottom) locations in the Garden Grove Infection Zone color coded by the year of the latest survey. Black dots are the HLB+ trees, and yellow rings represent the 170m exposure area.