



HLB Research and
Extension Outputs Network

Fostering informed decision-making for citrus growers: A research project database to combat Huanglongbing (HLB) disease

Ulianova Vidal-Gomez¹, Kaity Coleman², Sandra Olkowski², Neil McRoberts², David McClure³, Mike Hill³, David Laughlin⁴, Dharmendra Kalauni⁵, Sebastian Galindo⁵, Lauren M. Diepenbrock¹, Tripti Vashisth¹, Yu Wang¹, and Megan M. Dewdney¹

¹Citrus Research and Education Center, University of Florida, Lake Alfred, FL. ²University of California, Davis, Davis, CA. ³Purdue University, West Lafayette, IN. ⁴Texas A&M University, Kingsville, Weslaco, TX. ⁵University of Florida, Gainesville, FL.

INTRODUCTION

Huanglongbing (HLB) has greatly damaged the United States citrus industry, causing a major reduction in production over the past decade, particularly in Florida (USDA 2023). Despite extensive efforts, the HLB community struggles with inconsistent terminology and limited access to critical information.

To address these obstacles, this study aims to develop a comprehensive database of USDA*-funded research projects. Using text mining, we analyzed 573 titles and 72 summaries from HLB-focused grants to identify knowledge gaps, key developments and critical topics such as disease/vector management, new control strategies, and monitoring. Our findings reveal distinct priorities across USDA agencies. NIFA** emphasizes integrated management, ARS*** focuses on innovation and resistance, while multiagency initiatives address regional needs. These insights will guide the development of web tools to enhance the coordination and access to HLB relevant research, empowering citrus growers, scientists, and policymakers to make informed decisions.

*United States Department of Agriculture (USDA)
**Agricultural Research Service (ARS)
***National Institute of Food and Agriculture (NIFA)

METHODS

1. INFORMATION SOURCE:

Titles and summaries of HLB USDA-funded grants (2010-present) were obtained from USDA Data Gateway, USAspending.gov, and USDA multiagency initiative webpages. After ensuring data consistency, titles from ARS, NIFA, ECDRE and MAC were analyzed (n=573). Due to persistent data gaps, only project summaries from ECDRE-funded grants (n=72) were analyzed.

2. DATA ANALYSIS:

A word frequency analysis was applied to the grant titles, while grant summaries were grouped based on word frequencies using agglomerative hierarchical clustering. The process involved the following steps:

1

Estimating the TF-IDF:
Term frequency-inverse document frequency
(weighting factor)
$$\text{tf-idf}(t,d) = \text{tf}(t,d) * (\text{idf}(d,D)), \text{ where } D = \{d_1, d_2, \dots, d_n\}, d = \text{grants}$$
$$\text{tf}(t,d) = \frac{\# \text{ of occurrences of term } t \text{ in document } d}{\max \# \text{ of term occurrences in } d}$$
$$\text{Idf}(t,D) = \ln \left(\frac{\# \text{ of documents}}{\# \text{ of documents containing word } t} \right)$$

2

Calculating distance

3

Clustering
(Ward's method)

Analyses were conducted in R 4.4.1 (R core team 2024), using “tm” and “wordcloud” libraries for frequency calculations (Feinerer and Hornik, 2024) and “textminerR 3.0.5” for clustering (Jones, 2024).

RESULTS

SECTION 1. Word frequency patterns in HLB-focused USDA research grant titles.



Fig. 1. Frequently occurring words in the titles of 573 HLB USDA-funded research grants. Word frequencies are based on occurrence and correlation within the dataset. A. Top ten most frequent words. B. Most frequent 200 words. Word size and color are scaled proportionally, with greater prominence indicating higher frequency.

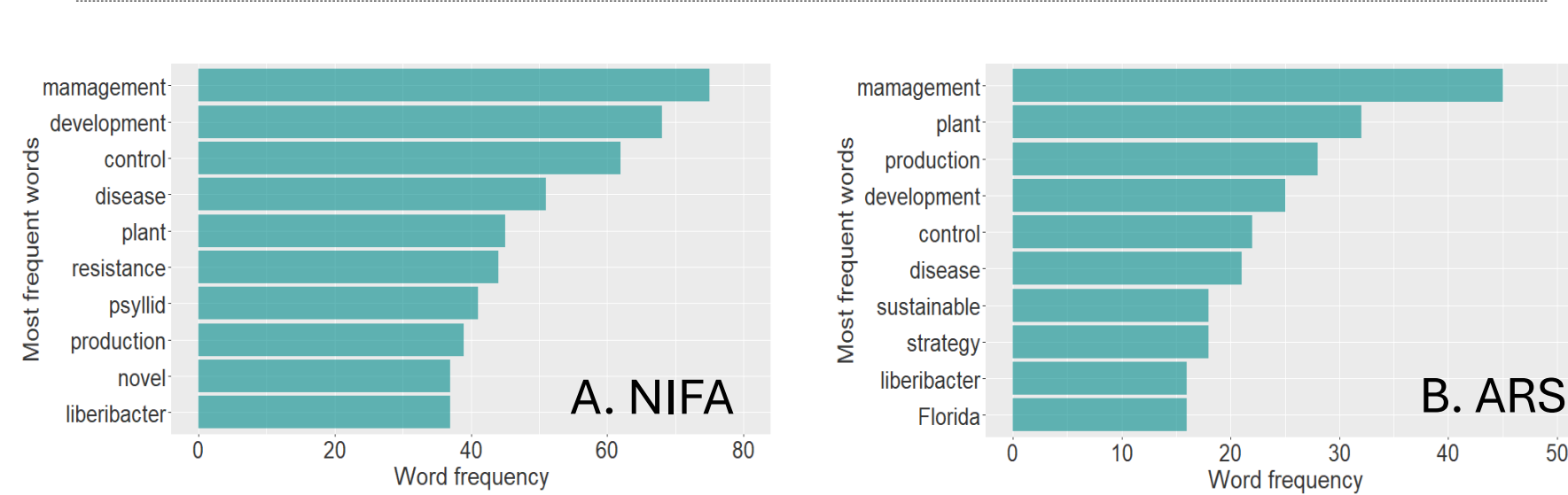


Fig. 2. Top ten most frequent words in the titles of HLB research grants funded by the USDA agencies. A. NIFA (n=269) and B. ARS (n=258). Word frequencies are based on occurrence and correlation within the dataset.

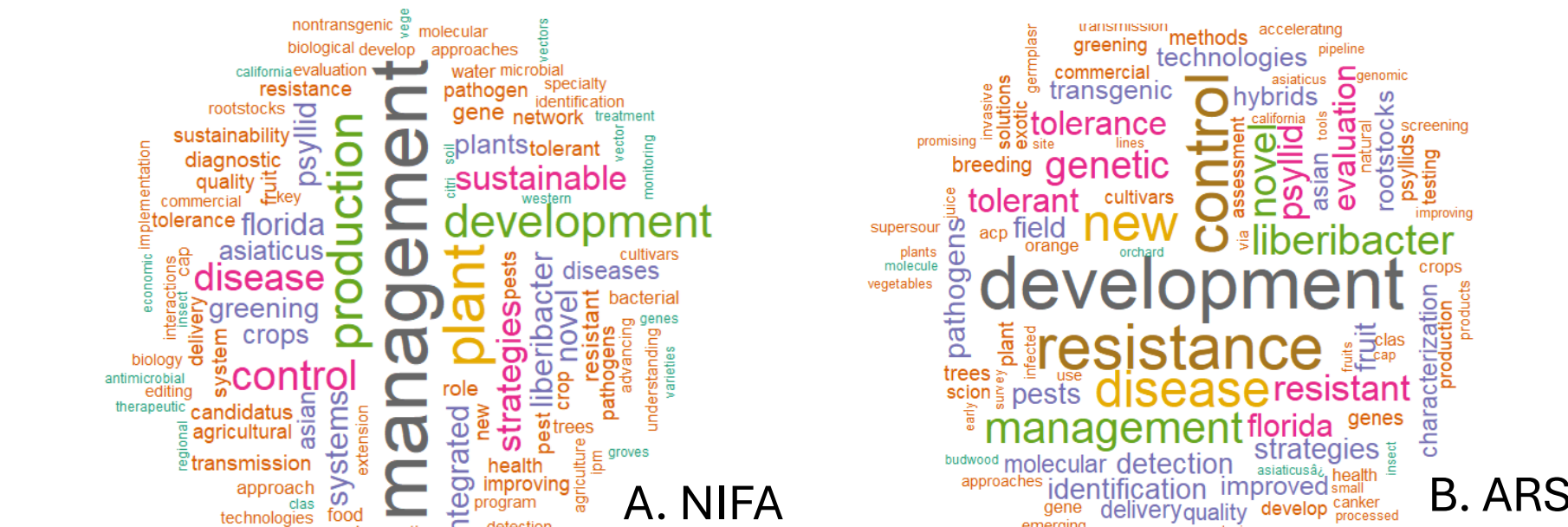


Fig. 3. Top 100 most frequent words in the titles of HLB research grants funded by USDA agencies. A. NIFA (n=269) and B. ARS (n=258). Word frequencies are based on occurrence and correlation within the dataset. Word size and color are scaled proportionally, with greater prominence indicating higher frequency.

DISCUSSION

1. The analysis revealed a strong focus on theoretical management strategies, but many of them will only be potentially implemented in the long-term.
2. Among long-term strategies, topics are equitably addressed from all agencies together, but NIFA emphasizes integrated management, production, and sustainability, whereas ARS prioritizes innovation, resistance, and tolerance. Multiagency initiatives reflect regional needs.
3. The examination of ECDRE summaries reveals that efforts have primarily focused on controlling the Asian citrus psyllid and understanding host-pathogen responses to address the HLB crisis.
4. Previously, the REON team identified four priority areas for database design (see Fig. 6), highlighting the importance of short-term solutions like antimicrobial strategies and nutritional programs. These areas have not been a funding focus of USDA programs as of this analysis.

CONCLUSIONS

1. Significant efforts have been made to find future solutions for combating the disease, but near-term solutions have been more elusive.
2. To make any research findings useful to growers, resulting information must be presented in a way that empowers stakeholders to make informed decisions.
3. Text mining techniques are valuable for identifying research gaps and developing strategies that align with the needs of the HLB community, guiding the development of our database.

REFERENCES

United States Department of Agriculture (USDA), National Agricultural Statistics Service (NASS). 2023. Citrus Fruits 2022-2023 Summary. Available at: Citrus Fruits 2023 Summary 09/03/2023 (ccqc.org).

SECTION 2. Word frequency and cluster analyses of ECDRE grant summaries.

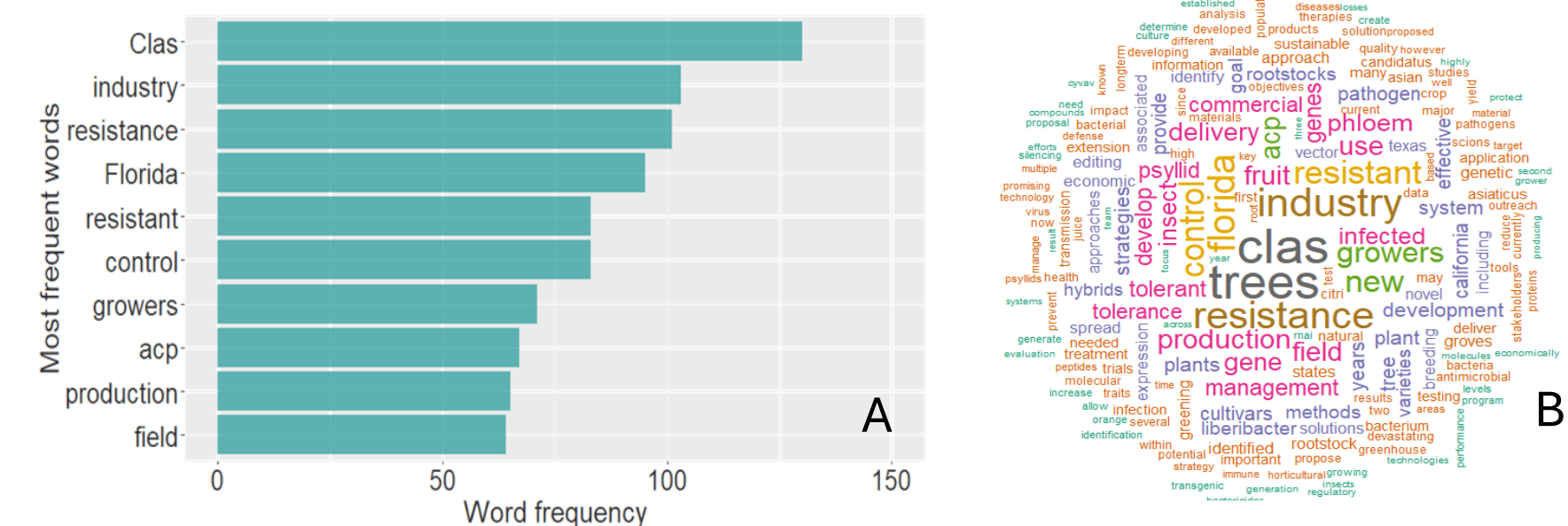


Fig. 4. Frequently occurring words in the summaries of 72 HLB ECDRE-funded research grants. Word frequencies are based on occurrence and correlation within the dataset. A. Top ten most frequent words. B. Most frequent 200 words. Word size and color are scaled proportionally, with greater prominence indicating higher frequency.

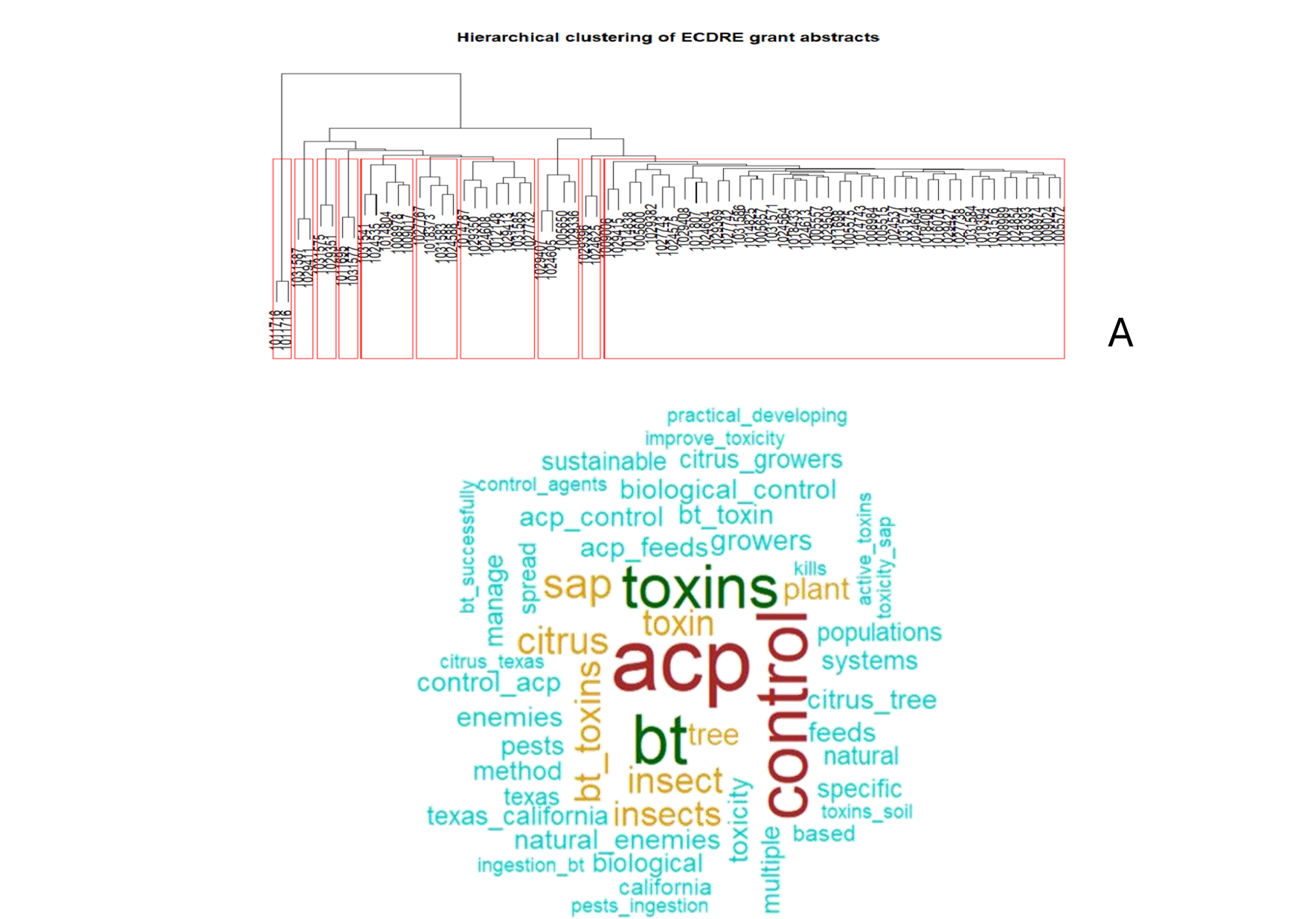


Fig. 5. Trends in research from 72 HLB ECDRE-funded grants. A. Dendrogram of key terms. Words are grouped into ten clusters based on frequency and similarity. Numbers indicate USDA grant identification codes. B. Word relationships within clusters. Fifty (50) key terms are displayed, with font size, color, and prominence scaled according to their TF-IDF values. More significant words and clusters shown in larger fonts (e.g. “acp” and “control”) are the most frequent words and are part of the largest cluster in the dendrogram).

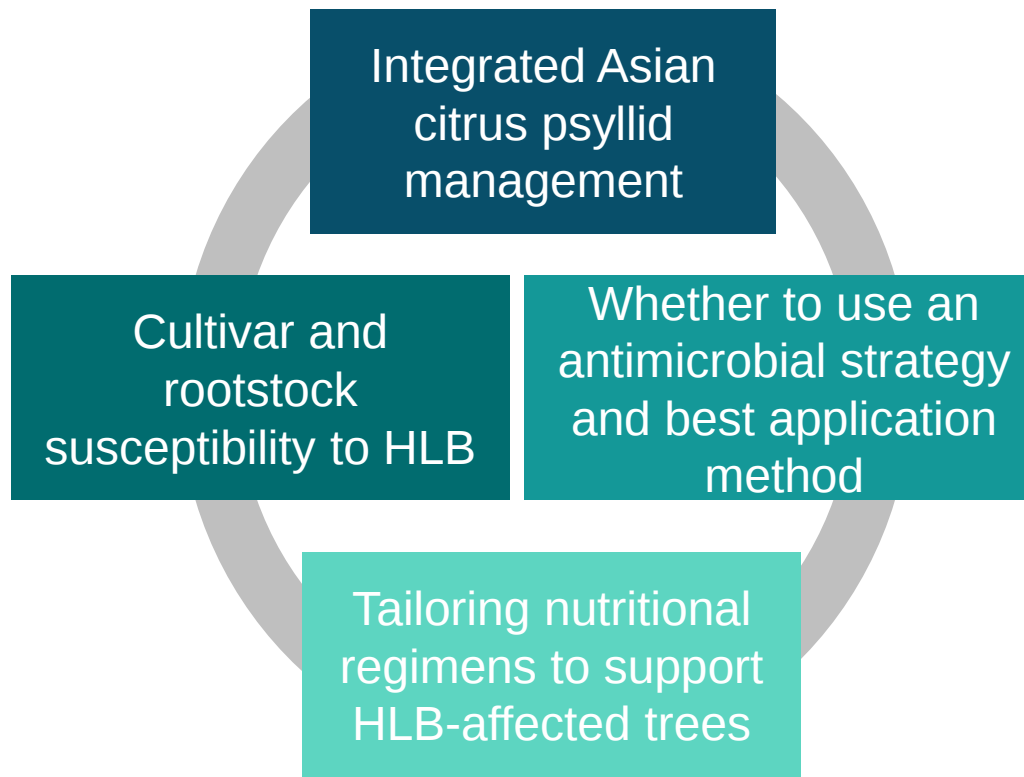


Fig. 6. Four areas identified by REON for the development of HLB web-based tools

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the USDA for its support in gathering information, the citrus extension agents, and the citrus growers in Florida, Texas, and California for their valuable insights and timely advice.