

The protein interactome of the citrus Huanglongbing pathogen *Candidatus Liberibacter asiaticus*

Received: 4 July 2023

Erica W. Carter^{1,2}, Orlene Guerra Peraza¹ & Nian Wang^{1,3} 

Accepted: 15 November 2023

Published online: 29 November 2023


 Check for updates

The bacterium *Candidatus Liberibacter asiaticus* (CLAs) causes citrus Huanglongbing disease. Our understanding of the pathogenicity and biology of this microorganism remains limited because CLAs has not yet been cultivated in artificial media. Its genome is relatively small and encodes approximately 1136 proteins, of which 415 have unknown functions. Here, we use a high-throughput yeast-two-hybrid (Y2H) screen to identify interactions between CLAs proteins, thus providing insights into their potential functions. We identify 4245 interactions between 542 proteins, after screening 916 bait and 936 prey proteins. The false positive rate of the Y2H assay is estimated to be 2.9%. Pull-down assays for nine protein-protein interactions (PPIs) likely involved in flagellar function support the robustness of the Y2H results. The average number of PPIs per node in the CLAs interactome is 15.6, which is higher than the numbers previously reported for interactomes of free-living bacteria, suggesting that CLAs genome reduction has been accompanied by increased protein multi-functionality. We propose potential functions for 171 uncharacterized proteins, based on the PPI results, guilt-by-association analyses, and comparison with data from other bacterial species. We identify 40 hub-node proteins, including quinone oxidoreductase and LysR, which are known to protect other bacteria against oxidative stress and might be important for CLAs survival in the phloem. We expect our PPI database to facilitate research on CLAs biology and pathogenicity mechanisms.

Citrus Huanglongbing (HLB) is one of the most destructive plant diseases worldwide. HLB has devastated the citrus industry in many regions, including Florida, US, and no effective and economic disease control measures are available for these HLB-endemic citrus production regions^{1,2}. HLB is caused by *Candidatus Liberibacter asiaticus* (CLAs), *Ca. L. americanus*, and *Ca. L. africanus*, with CLAs being the most prevalent³. *Ca. Liberibacter* species are endophytic bacteria that colonize the phloem tissues of citrus plants and are transmitted by the Asian citrus psyllid *Diaphorina citri*. The environments inside the phloem tissue and psyllids are nutrient-rich. Consequently,

Ca. Liberibacter species have undergone reductive evolution and have considerably smaller genomes (approximately 1.2 MB) than their culturable, free-living relatives, such as *Agrobacterium*^{4,5}. Cultivation of HLB pathogens in artificial media has not been achieved despite extensive effort⁶, which prevents its genetic manipulation and impedes the understanding of their biology and pathogenicity mechanism.

The CLAs str. Psy62 (NC_012985.3) genome encodes approximately 1136 genes, 1027 of which are non-redundant proteins. Sixty percent (612 ORFs) of these coding regions have predicted functions based on sequence similarities with characterized proteins from

¹Citrus Research and Education Center, Institute of Food and Agricultural Sciences, University of Florida, Lake Alfred, FL, USA. ²Department of Plant Pathology, Institute of Food and Agricultural Sciences, University of Florida, Lake Alfred, FL, USA. ³Department of Microbiology and Cell Science, Institute of Food and Agricultural Sciences, University of Florida, Lake Alfred, FL, US.  e-mail: nianwang@ufl.edu

culturable microorganisms. However, 415 CLas proteins have no known functions; these were defined by Duan et al. as 69 with general function only, 46 had no known function, and 300 without a COG family⁵. Investigating protein functions, particularly those with unknown roles, is critical for elucidating the biology and pathogenicity of CLas and developing effective strategies to combat this notorious citrus disease. Protein–protein interactions (PPIs) are fundamental to all cellular processes and machinery. PPIs modify enzyme and protein activities, catalyze metabolic reactions, and activate signaling pathways⁷. Exploring PPIs has also been used to investigate biological pathways, especially interactions between proteins of known and unknown function. These PPIs provide informative clues for pathway and function prediction^{8,9}. Furthermore, targeting PPIs can be used to develop novel antimicrobials¹⁰. The Yeast Two-hybrid system (Y2H) is the most commonly used approach to investigate interactomes and has been elegantly applied to various organisms, such as animals including *Drosophila melanogaster*¹¹ and *Caenorhabditis elegans*^{12,13}; plants including *Arabidopsis thaliana*¹⁴; fungi including *Saccharomyces cerevisiae*⁸; viruses including potyviruses Soybean mosaic virus¹⁵; and bacteria including *Streptococcus pneumoniae*¹⁶, *Treponema pallidum*¹⁷, *Campylobacter jejuni*¹⁸, *Synechocystis sp.*¹⁹, *Mycobacterium tuberculosis*²⁰, *Mesorhizobium loti*²¹, *Escherichia coli*²², *Bacillus subtilis*²³, and *Helicobacter pylori*²⁴. Affinity Purification Mass Spectrometry (AP-MS) has also been used to study the *Mycoplasma pneumoniae* interactome^{7,25}. However, AP-MS detects PPIs in the native organism, revealing protein complex topology and not necessarily binary topology, whereas Y2H bypasses this limitation by identifying the binary PPIs, which has been used to elucidate the binary PPIs found in AP-MS studies²². Interactome studies (Supplementary Table 1) have led to significant progress in inferring the function of unknown proteins, elucidating protein organization of a cell, identifying novel functions of known proteins, and, in some cases, are used to find

interaction points between pathogenic organisms and their eukaryotic hosts^{26–32}.

Here, we conducted a genome-wide Y2H interactome study for CLAs, identifying over 4000 PPIs and proposing potential functions for 171 uncharacterized proteins.

Results

High-throughput Y2H screening

We conducted Y2H assays to identify the CLAs interactome. We used full-length open reading frames (ORFs) for yeast vector construction to increase the possibility of identifying biologically relevant interactions. We successfully amplified 974 CLas ORFs (Supplementary Data 1) and cloned 942 ORFs into Y2H expression vectors (Supplementary Fig. 1A). In total, 926 “bait” proteins in pGEO_BD, which was modified from pGBKT7_BD (Supplementary Fig. 1B), and 950 “prey” proteins in pGAD_T7. After the removal of autoactivating constructs (Supplementary Fig. 2), 916 “bait” and 936 “prey” constructs remained suitable and utilized in the screening, representing 91% of genes encoded by CLAs. We constructed the CLAs binary interactome using a three-phase array screening method (Fig. 1A). The first two pooled array phases removed non-interactors. Proteins interacting on the most stringent media, QX (-trp, -leu, -ade, -his, +XαGal) and QXA (-trp, -leu, -ade, -his, +XαGal, +AbA) in the first two screening phases were used in the pairwise array in the third screening phase. The final (third phase) screening between 617 baits and 466 preys resulted in 4245 interactions between 542 proteins (Supplementary Data 2), covering 52.8% of CLas genes, denoted as the CLAs_{whole} network (Fig. 1B). Among the 542 proteins, 371 had known functions, whereas 171 were uncharacterized proteins (Supplementary Data 3; Tables 1 and 2). The pairwise screening enabled us to identify all potential interactions from the mini-pooled arrays and detect the interacting partners. In addition, pairwise rescreening eliminated probable promiscuous

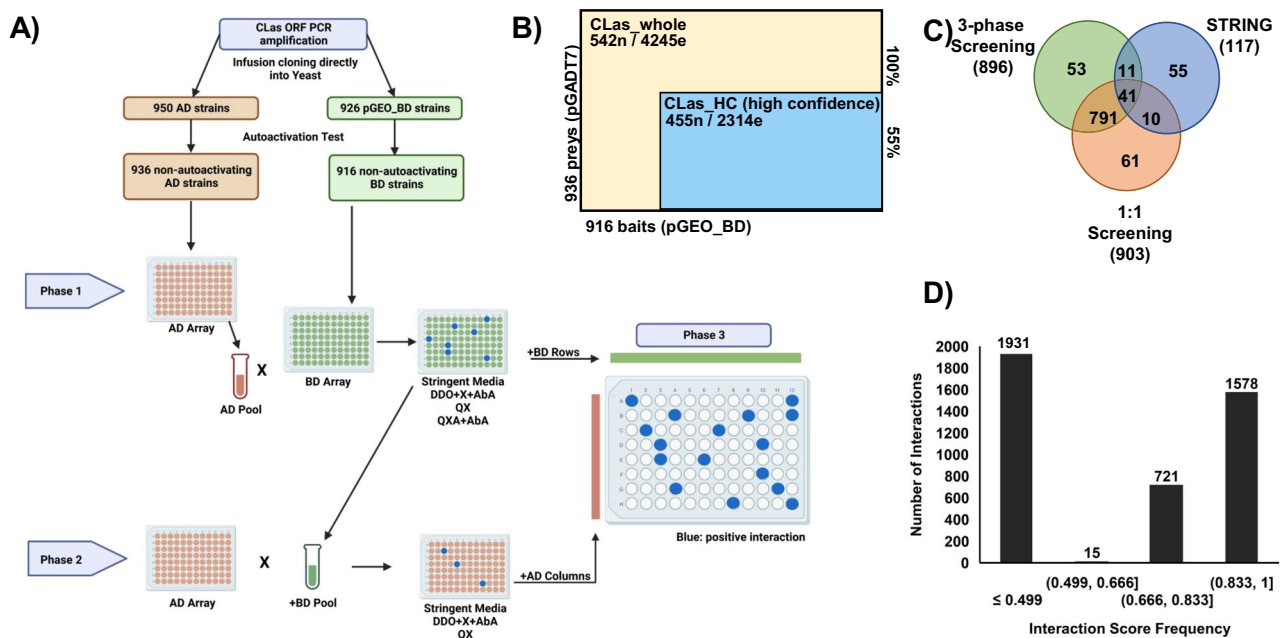


Fig. 1 | The CLas Y2H interactome screening, validation, and confidence scores.

A The three-phase 96-well plate screening pipeline for constructing the CLAs_{whole} network. Phase one and two systematically screened pools and individual yeast constructs to identify potential interacting protein partners. All interacting partners were identified in the third phase in a pairwise screening; phase 1 and 2 interacting proteins were screened pairwise against all other phase 1 and 2 interacting proteins to produce the binary data. This figure was generated using BioRender. **B** CLas Y2H three-phase screening results. n node, e edge, 100% indicates:

CLas_{whole} network; 55% indicates: CLas_{HC} (high confidence) network portion of the CLAs_{whole} network. **C** Venn diagram showing the overlap of interactions detected in the three-phase screening, an independent pairwise screening, and predicted interactions for 163 CLas proteins; predicted interactions were downloaded from the STRING consortium. **D** Interaction confidence score frequency across the CLAs_{whole} network. Interactions with a confidence score ≥ 0.5 are considered high confidence; denoted as the CLAs_{HC} network.

proteins from the dataset by observing PPIs resulting from promiscuous proteins.

CLas interactome coverage and confidence scores

The robustness of the Y2H datasets is imperative for utilizing the PPI information^{33–35}. We reduced false positives by pairwise rescreening and stringent selection of the CLas PPI network in the three-phase screening. To evaluate the sensitivity and reproducibility of the three-phase method, we subjected a random set of 326 yeast constructs (163 proteins in each vector) to an independent pairwise Y2H screening. We observed 832 reproducible interactions between the three-phase and pairwise screening, representing 92% of the interactions detected between those 163 proteins in the CLas_{whole} network.

We further employed known PPIs in the STRING-db to evaluate the sensitivity of the two screening assays. Limited information is available for CLas in the STRING database, with no experimentally determined interactions reported for CLas specifically. However, there were 474 and 369 experimentally determined interactions between homologous, interolog proteins (iPPIs) and interacting protein family interologs (iCOGs) at confidence cutoffs ≥ 0.4 and ≥ 0.9 , respectively (Table 1; Supplementary Table 2)^{36,37}. Our results showed that both Y2H screening methods had comparative sensitivity in detecting predicted CLas interactions based on known PPIs of homolog proteins. The three-phase screening detected 52 whereas the pairwise approach identified 51 of the 117 predicted interactions among these 163 proteins with 41 detected by both methods (Fig. 1C, Supplementary Data 4).

Using the interactions detected in the two Y2H screening methods, we calculated the false positive rate for the high-throughput three-phase Y2H screening as 2.9% and the false negative rate as 55.6%. Thus, our three-phase Y2H is stringent against false positives in identifying putative PPIs of CLas. However, though the high throughput three-phase screening is efficient, it also brings a significant false negative rate of 55.6%. Because of the highly reproducible PPIs (92%) between the three-phase and pairwise screenings (Fig. 1C), it is probable that the false negative rate is an intrinsic limitation of the Y2H assay as previously reported for other binary PPI assays^{14,38,39}. Nevertheless, the low false positive rate rendered our three-phase Y2H data highly robust to analyze the CLas interactome and infer the protein function of hypothetical proteins, further supported by the CLas interactome specificity of 97.1%, where very few non-interactors were detected.

To evaluate the robustness of individual PPIs, we calculated interaction confidence scores for the CLas_{whole} network PPIs using iPPIs as the positive reference set. The interologs were obtained by

querying the STRING-db for experimentally determined PPIs from *Agrobacterium radiobacter*, *B. subtilis*, *C. jejuni*, *E. coli*, *H. pylori*, *L. crescens*, *M. genitalium*, *M. loti*, *M. pneumoniae*, *S. cerevisiae*, *S. meliloti*, *T. pallidum*, and iCOGs. There were 635 iPPIs between 208 protein pairs in the CLas_{whole} network (Supplementary Data 5 and 6). We used 100 PPIs corresponding to interologs with the highest colony scores to represent the positive reference set for logistic regression analysis and assign interaction confidence scores to the CLas_{whole} PPIs. In total, 2314 interactions between 455 proteins demonstrated high confidence scores (≥ 0.5 , CLAS_{HC}), representing 55% of the total interactions detected in the CLas_{whole} network and 44% of the CLas ORFs (Fig. 1B, D, Supplementary Fig. 3, Supplementary Data 7).

To further verify our Y2H PPIs, we conducted pull-down assays for nine PPIs focusing on flagellar proteins because of the crucial roles of CLas motility in the HLB pathosystem⁴⁰. Consistent with the Y2H result, nine tested PPIs were confirmed (Fig. 2), supporting the high quality of the three-phase Y2H screening.

Network topology

Protein network topology provides valuable information for protein functions^{41,42}. We visualized and analyzed the topology of the CLas Y2H interactome (Fig. 3A–C)^{43,44}. The degree distribution topology of CLas_{whole} had a scale-free organization, typical of a scale-free PPI network (Fig. 3D)²⁸. The average node connectivity was 15.664 (average node degree of 7.8) in the CLas_{whole} network; CLas_{HC} has an average node connectivity of 10.171 (average node degree of 5.1). The CLas network had many node neighbors, consistent with the notion that protein function complexity increases with genome reduction (Fig. 3F)^{7,45}. The CLas_{HC} network maintained a degree distribution and topological coefficient distribution like that of the CLas_{whole} network (Fig. 3D, G). However, the clustering coefficient and betweenness centrality distribution of CLas_{HC} were less dense than the CLas_{whole} with fewer crosstalk between nodes (Fig. 3E, F)⁶.

Highly connected nodes (hubs) are more likely to be responsible for maintaining the overall connectivity of the network. They are more likely to be essential genes or involved in critical PPIs than non-hub nodes⁴⁶. We identified 40 hub nodes in the CLas_{whole} network. Using betweenness centrality and degree distribution, we manually curated 27 hub nodes (Fig. 3D, E)⁴⁶. We validated the hubs using the Maximal Clique Centrality (MCC) and MCODE algorithms^{47,48}. MCODE clustering confirmed 26 of our 27 manually curated nodes as hubs with scores ≥ 4 . The MCC clustering algorithm confirmed 14 hub nodes defined by manual curation and MCODE clustering and suggested 13 additional hub proteins (Fig. 3B, Supplementary Data 8). Interestingly, eight of the hub proteins were reported to be essential for *L. crescens* [CLIBASIA_RS05300: terminase; CLIBASIA_RS05115: hypothetical protein; CLIBASIA_RS01595: DUF1036 domain-containing protein; CLIBASIA_RS01230: electron transfer flavoprotein subunit alpha/FixB family protein; CLIBASIA_RS01735: DNA gyrase inhibitor YacG; CLIBASIA_RS01125: riboflavin synthase; CLIBASIA_RS03885: PilZ domain-containing protein; CLIBASIA_RS04735: DUF59 domain-containing

Table 1 | Number of annotated and uncharacterized proteins in the CLas_{whole} and CLas_{HC} networks

Network	Annotated	Uncharacterized	Total
CLas _{whole}	371	171	542
CLas _{HC}	295	160	455

Table 2 | Interactions between annotated and uncharacterized proteins

Interaction type	CLas _{whole} Node	CLas _{whole} edge	CLas _{HC} node	CLas _{HC} edge
Annotated - annotated	339	1491	246	692
Annotated - hypothetical	502	1976	388	1106
Hypothetical - hypothetical	148	570	123	388
Interologs (annotated - annotated)	128	171	96	106
Interologs (annotated - hypothetical)	47	30	30	18
Interologs (hypothetical - hypothetical)	12	7	7	4
Total	542	4245	455	2314

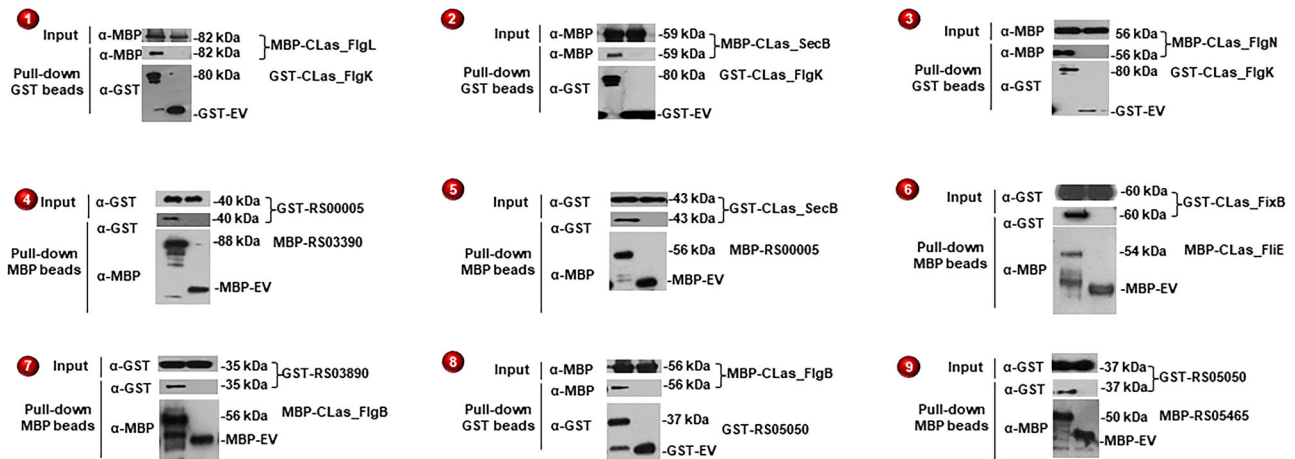


Fig. 2 | Confirmation of the PPIs identified in Y2H using pull-down assays. Both GST and MBP pull-down assays were conducted, dependent on protein pairs. Interactions confirmed *in vitro* are numbered for illustration in Fig. 5. EV- empty vector. CLIBASIA_ was removed from the CLas protein access ID in 4, 5, 7, 8, and 9. Each experiment was repeated independently at least twice with similar results. Source data are provided as a Source data file.

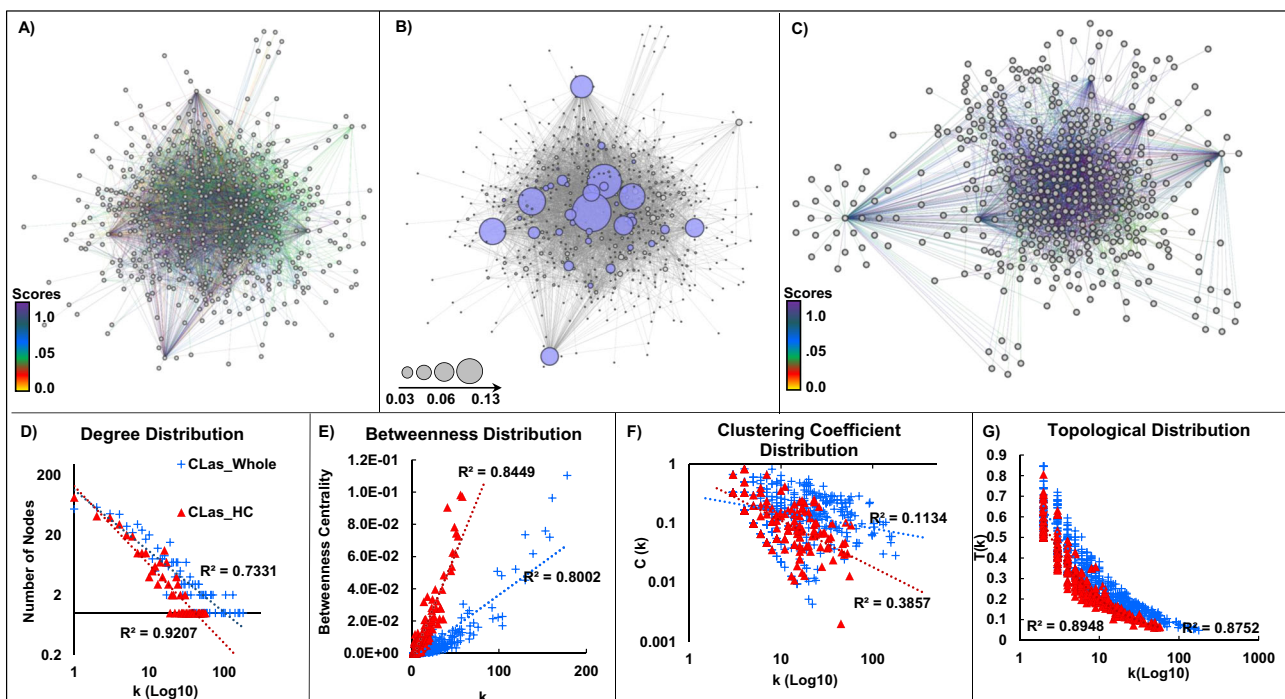


Fig. 3 | The CLas Y2H network topology. **A–C** Graphic representation. n node indicating a protein in the network, e edge indicating the connection between the interacting proteins. **A** CLas_whole network (542n, 4245e). **B** hub nodes in the CLas_whole network are colored blue; node size is proportional to its betweenness centrality value. **C** CLas_HC network; PPI confidence scores ≥ 0.5 . **D–G** Network topology of the CLas_whole and CLas_HC networks. **D** Degree distribution. **E** Betweenness centrality distribution. **F** Clustering coefficient distribution. **G** Topological coefficient distribution.

protein]⁴⁹, and 24 are known to be essential in other bacteria (Supplementary Data 8)⁵⁰.

Inferring functions of CLas hypothetical proteins

CLas contains 415 uncharacterized proteins, including proteins annotated as hypothetical proteins, unknown function, or not in a COG^{5,51}. The CLas_HC network contained 1485 high-confidence interactions involving 160 uncharacterized proteins. The features of these proteins include: 15 have signal peptides that may direct them to the cell membrane or outside the cell; 13 are secreted by a nonclassical pathway that does not involve signal peptides as

previously described by Du et al.⁵²; 27 have transmembrane domains that span the cell membrane; and 107 have no assigned COG that indicates their function⁵³. One hundred forty-nine uncharacterized proteins interacted with at least one annotated protein, and 105 interacted with two or more proteins with known functions (Supplementary Data 3).

To infer the functions of those uncharacterized proteins based on their interacting partners, we employed four guilt-by-association (GBA) methods: 1. Binary interactions and iPPIs, 2. Operon associations, 3. Protein family associations, and 4. Meta-interactome analysis (Fig. 4)^{54–56}.

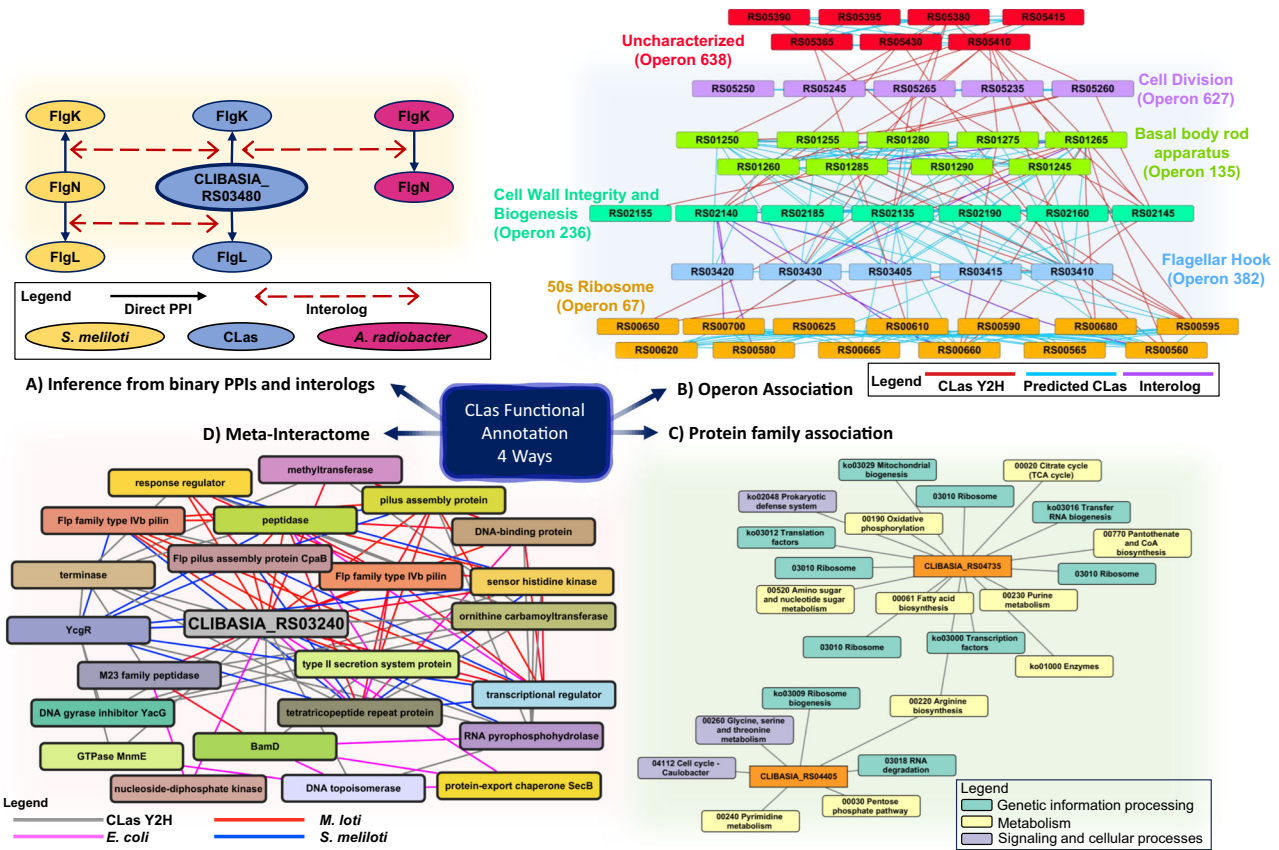


Fig. 4 | Four approaches were used to infer CLas hypothetical protein functional annotations. **A** Hypothetical protein annotations were implied by interologs and binary interactions. These validity of the interactions were confirmed by in vitro assays and tertiary protein structure homology. **B** Inter-operon interactions highlight associations between motility related genes, respiration, membrane integrity, ribosomal proteins, and uncharacterized genes and operons. Operon associations, network and properties are in Supplementary Data 9. CLIBASIA_ was removed from the loci ID for simplicity. **C** Guilt by association was used for CLas

proteins interacting with ≥ 2 proteins in the same protein family or KEGG pathway. CLIBASIA_RS04405 has a functional association with enzymatic proteins, ribosome and RNA biogenesis, and exosome, cytoskeleton proteins. The binary interactions between CLIBASIA_RS04735 and ribosome, transcription/translation proteins implies a functional association with cell response and growth. **D** The sub network meta-interactome my2H₃₂₄₀ for CLIBASIA_RS03240 inferred ortholog associations with protein export, secretion, kinase, pilus, and peptidase proteins.

Inferring CLas protein functions from binary interactions

iPPIs are conserved interactions between proteins whose homologs interact in other organisms¹³. As aforementioned, we identified 635 iPPIs between 208 protein pairs in the CLas_{whole} network (Supplementary Data 5 and 6). Among them, 34 were uncharacterized proteins. We have assigned functions to four proteins based on interacting partners and similarities between tertiary (3D) protein structures (Supplementary Fig. 4). CLIBASIA_RS03480 interacted with FlgK_{Las} and FlgL_{Las} in the interactome; the interactions between these proteins are well documented and are iPPIs found in *S. meliloti* and *A. radiobacter* (Fig. 4A)⁵⁷. The protein structure alignment of CLIBASIA_RS03480 in Phyre2 found it shares 31% structure similarity (confidence score of 83.2) with the *Bradyrhizobium* sp. FlgN protein (Uniprot: A0A1I3GD01 (<https://www.uniprot.org/uniprotkb/A0A1I3GD01/entry>)) whereas they share 17.89% sequence identity. CLIBASIA_RS03480 was reannotated as FlgN_{Las}, an FlgK and FlgL specific flagellar switch chaperone because it interacted with FlgK_{Las} and FlgL_{Las} proteins, involved in flagellar hook formation in the CLas network and its structure homology with the *Bradyrhizobium* sp. FlgN protein structure template (Fig. 4A, Supplementary Fig. 4A). Similarly, we reannotated CLIBASIA_RS03395 as FliK_{Las}, a flagellar hook-length regulator, because of its interactions with FlgD_{Las}, FlhB_{Las}, and FlIP_{Las} (Supplementary Fig. 4B), three proteins involved in flagellar motor assembly^{38–60}. In 2020, MotD from *S. meliloti* was characterized as a flagellar-hook-length regulator and reassigned as FliK; MotD in other α -proteobacteria was reassigned

as FliK dependent on an average AA sequence identity of $\geq 45\%$ in its C-terminus. CLIBASIA_RS03395, a CLas protein annotated as a chemotaxis protein, shares iPPIs in *S. meliloti*, and the C-terminus protein structure alignment has a 68.83% identity and confidence score of 90.2 compared to the *S. meliloti* FliK protein (Uniprot: F7X8H0 (<https://www.uniprot.org/uniprotkb/F7X8H0/entry>)). Additionally, CLas hypothetical proteins CLIBASIA_RS03285 and CLIBASIA_RS03885 were annotated as ToIA_{Las} and YcgR_{Las} respectively (Supplementary Fig. 4C, D). CLIBASIA_RS03285 interacts with TolB in the CLas interactome. The interactions between these two proteins, ToIA and TolB, are required for a functional Tol-Pal system in bacteria to maintain cell membrane integrity. Using Phyre2, we found that CLIBASIA_RS03285 has a 45.63% protein structure identity and a confidence score of 100 when aligned with the *E. coli* ToIA (Uniprot: P19934 (<https://www.uniprot.org/uniprotkb/P19934/entry>)) C-terminal domain albeit they only share 23.36% sequence identity. CLIBASIA_RS03885, a PilZ domain-containing protein, is likely a YcgR flagellar brake protein. The protein structure alignment between CLIBASIA_RS03885 and *E. coli* YcgR (Uniprot: P76010 (<https://www.uniprot.org/uniprotkb/P76010/entry>)) has 45.63% identity and confidence score of 98, while they share 25.93% sequence identity.

Inter- and Intra-operon binary interaction topology

Next, we took advantage of the fact that bacterial genes of related functions are transcribed together under a single promoter as

operons^{61,62}. We reasoned that operons with known proteins interacting with the operons containing uncharacterized proteins provide functional associations/clues for those groups of uncharacterized proteins. For this purpose, we only considered inter- and intra-operon interactions where two or more proteins from each operon interacted together for further analysis (Supplementary Data 9). We examined how the PPIs of other organisms relate to the CLAs Y2H network, specifically inter- and intra-operon PPIs (Fig. 4B and Supplementary Fig. 5). We found that CLAs had more inter-operon PPIs than intra-operon interactions (Supplementary Data 9 and 10) compared to interactions between homologous proteins. These inter-operon interactions provided evidence of functional association of three CLAs operons (operon 638, operon 603, and operon 643), which consist of 21 uncharacterized proteins. Specifically, our interactome data demonstrated that CLAs operon 638 has a functional association with flagellar assembly, stress response-related proteins, 50 s ribosomal proteins, cell division, and cell wall integrity proteins (Fig. 4B). In addition, CLAs operon 603 has a functional association with translation, FeS cluster assembly, and succinate dehydrogenase. CLAs operon 643 has a functional association with metabolism, enzymes, and translation; like operon 603, operon 643 also shares an association with FeS cluster assembly and succinate dehydrogenase (Supplementary Fig. 5).

Protein family associations

We evaluated PPIs between hypothetical proteins and CLAs proteins within known protein families or functions from the Kyoto Encyclopedia of Genes and Genomes (KEGG)⁶³. Sixty-one CLAs proteins with unknown functions interacted with two or more proteins from the same protein family or function in the CLAs_HC Y2H network. These interactors were primarily involved in metabolism and genetic information processing, as shown in examples of protein association networks for CLIBASIA_RS04735 and CLIBASIA_RS04405, which primarily interacted with translation, ribosomal, and enzyme proteins, indicating their functional relationship with metabolism and genetic information processing proteins (Fig. 4C). Operon associations for eleven GBA reannotated CLAs proteins confirmed the KEGG protein associations, increasing confidence in their functional assignment. (Supplementary Data 11, 12; Table 2, Supplementary Table 2).

CLAs meta-interactome

As a final sweep of the Y2H data, we looked for associations in a meta-interactome¹⁶. We generated a meta-interactome (mY2H) using experimentally confirmed PPIs for *A. radiobacter*, *B. subtilis*, *C. jejuni*, *E. coli*, *H. pylori*, *L. crescens*, *M. genitalium*, *M. loti*, *M. pneumoniae*, *S. cerevisiae*, *S. meliloti*, and *T. pallidum*. The CLAs mY2H has 12,780 edges between its 542 proteins; 416 CLAs proteins have orthologs with organisms with available interactome maps (Fig. 4D, Supplementary Fig. 6). *M. loti* is more closely related to CLAs than other organisms with available Y2H interactome maps, such as *E. coli* and *H. pylori*^{9,21,22}. Consequently, most interactions (not including the CLAs_whole Y2H PPIs) in the mY2H were PPIs in *M. loti*, totaling 2296 edges. Of these mY2H PPIs, 635 iPPIs between 208 protein pairs were conserved in the CLAs_whole, whereas 4037 were not (Supplementary Data 13). We enriched the CLAs_whole by adding first neighbor proteins from the ortholog networks for CLAs uncharacterized proteins. We infer the functions of 26 uncharacterized CLAs proteins based on their interactions with known proteins (Supplementary Fig. 6; Supplementary Data 12; Table 2, Supplementary Table 2). For example, CLIBASIA_RS03240 (Fig. 4D) is a hypothetical protein interacting with the outer membrane protein BamD, DNA topoisomerases, and a GTPase protein in the CLAs Y2H. Its meta-interactome was enriched for interactions among protein export, secretion, kinase, pilus, and peptidase proteins. We found CLIBASIA_RS03240 has a high-confidence structure similarity (score of 100) and 59% structure alignment, with 43%

sequence identity to the *Bacillus halodurans* [TaxId:86665] PDB 2nly, a divergent polysaccharide deacetylase (Pfam PF04748), which is consistent with its recent reannotation in the NCBI database. Similarly, CLIBASIA_RS02190, a DUF1009 domain-containing protein, has a high-confidence structure similarity to the lipid A biosynthetic pathway protein LpxI (confidence score of 100, 98% structure alignment, and 31% sequence alignment) from *Caulobacter vibrioides*, and associates with lipid biosynthesis and cell growth-related proteins in the Y2H interactome. Lastly, CLIBASIA_RS01535 is a hypothetical protein that has a high-confidence structure similarity to the GcrA cell cycle regulatory protein (5Z7I) from *Caulobacter crescentus* with a confidence of 100, 96% structure alignment, and 64% sequence identity. It has also been reannotated as such in the NCBI database.

Novel interactions identified for CLAs proteins with known functions

The CLAs_whole network identified 4245 interactions between 542 proteins (Fig. 1B). Among the 371 CLAs proteins with known functions (Tables 1, 2), we identified 1654 interactions, whereas their homologs had 171 unique experimentally verified interactions in previous studies^{9,16-19,21-24,64}. This study identified 1483 novel PPIs that have not been previously reported for proteins with known functions (Supplementary Data 14).

Y2H PPIs related to flagellar proteins

Flagellar motility might play critical roles in CLAs infection of different organs of psyllids and movement in phloem tissues. Importantly, active movement of CLAs was observed in the phloem tissue against the flow of phloem sap^{65,66}. We paid close attention to flagella-related proteins. Our Y2H screening detected 399 interactions among known flagellar proteins in our Y2H screening (Fig. 5; Supplementary Data 15; Supplementary Tables 3, 4). In addition, nine interactions were confirmed with pull-down assays (Fig. 2). Of those PPIs confirmed, two were putative secreted proteins, CLIBASIA_RS05050⁵³ and CLIBASIA_RS05465⁵², interacting with each other and FlgB_{Las}, a flagellar basal-body protein⁶⁷. Another secreted protein, CLIBASIA_RS03890, also interacted with FlgB_{Las}. These interactions imply that these proteins may play a role in flagellar formation and functions, which could affect the pathogenicity or survival of CLAs⁶⁸. FlgN is a known flagellar export chaperone of FlgK and FlgL, two hook-filament junction proteins⁵⁷. We also identified SecB, a chaperone protein dedicated to translocating proteins across the cytoplasmic membrane, as an interactor of FlgK. We validated both interactions, FlgK_{Las}-FlgN_{Las} and FlgK_{Las}-SecB_{Las}, consistent with the involvement of the Sec pathway in flagellar synthesis⁶⁹. Moreover, SecB_{Las} interacted with CLIBASIA_RS0005, a protein with a T-SNARE-like domain. T-SNARE proteins mediate membrane fusion in eukaryotic cells. Pathogens can exploit host SNARE proteins to enter host cells by displaying SNARE-like domains on their surface membrane⁶⁹. It remains to be determined whether CLIBASIA_RS0005 is involved in CLAs entering cells or glands of psyllids⁷⁰.

Discussion

We have constructed a large-scale protein-protein interaction map of CLAs using a high-throughput Y2H approach. In total, we have identified 4245 PPIs among CLAs proteins. We have employed a three-phase screening approach to improve the robustness of the Y2H. We estimated that our high-throughput Y2H data has a false positive rate of 2.9%, consistent with previously reported false positive rates of 0.5%³⁸, less than 5%¹⁴, and 6.5%⁷¹. The pull-down assay supported the robustness of our Y2H with the verification of nine PPIs found in the Y2H network. Among the 1662 interactions between proteins with known functions of CLAs, only 171 have been observed in other systems, indicating that CLAs proteins might have evolved new functions to adapt to the reduced genome size. CLAs, *T. pallidum*, and

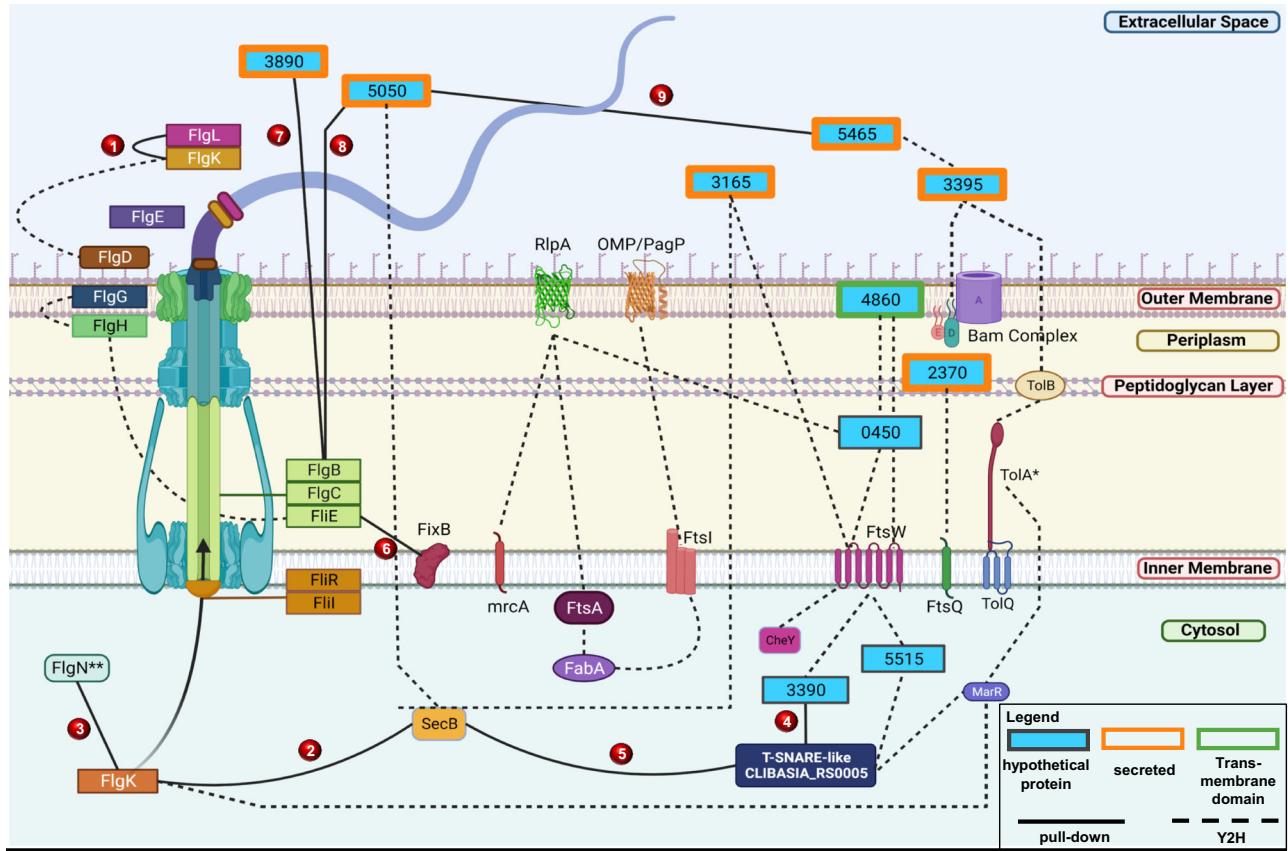


Fig. 5 | Integrated view of CLas proteins related to flagellar biosynthesis and membrane proteins, their interacting partners identified in Y2H. Subcellular localization representation of CLas hypothetical and annotated protein interactions. The source of the interactions and hypothetical proteins are denoted by dashed and solid lines. Hypothetical proteins with orange borders are

experimentally confirmed secreted proteins as described in Prasad et al. and Du et al. Interactions confirmed with pull-down experiments were numbered from 1–9 in red circles. The hypothetical proteins were named using their CLIBASIA_RS#. **Here CLIBASIA_RS03840 is represented as FlgN. This figure was generated using BioRender.

M. pneumoniae have a greater number of interactions per node in their PPI networks than free-living bacteria with larger genomes, such as *E. coli*, *M. loti*, *H. pylori*, *C. jejuni*, and *B. subtilis*. An increase in node neighbors seems to be a common phenomenon for bacteria that have undergone reductive evolution⁴⁵.

We identified 40 putative hub nodes in the network⁴⁷. Hub proteins might act as moonlighting proteins and have multiple functions as suggested by their multiple interacting partners⁷². It remains to be determined whether intrinsically disordered proteins (IDPs) and regions (IDRs) are involved in the multitasking of moonlighting proteins⁷³. Hub proteins are suggested to be essential to the organism or involved in critical PPIs^{23,41,74}. On the contrary, such findings were partially debunked in a study by Yu et al.⁷⁵. Three hundred fourteen genes are required for *L. crescens* growth in pure culture, and the CLas genome encodes 238 homologs to these *L. crescens* genes⁴⁹. However, CLas is missing 76 essential *L. crescens* genes, suggesting that CLas has different, and possibly more, physiological needs for growth than *L. crescens*. Hub proteins may provide insight into culturing CLas since we may be able to identify critical proteins necessary for its growth in culture. Eight of the hub proteins were reported to be required for culturing *L. crescens*⁴⁹, and 24 are known to be essential in other bacteria⁴⁰.

Furthermore, ten of the 40 hub genes are significantly up- or down-regulated in either citrus or the ACP vector, suggesting they play crucial roles for CLas survival in the host or vector⁷⁶. This differential expression could allude to lifestyle switches in the bacterium during the host-vector transition and may be exploited for antibiotic or other targeted strategies for HLB management. Several CLas hub proteins

have antioxidant capabilities, protecting the bacteria from oxidative stress damage, including quinone oxidoreductase 1⁷⁷ and LysR-type transcriptional regulator⁷⁸. Quinone oxidoreductase 1 acts as a cytoplasmic antioxidant, whereas LysR is a key circuit component in regulating microbial stress responses and is required for bacterial tolerance to H₂O₂ in vitro⁷⁸. The protective mechanism of CLas is consistent with the model that HLB is a pathogen-triggered immune disease; CLas triggers systemic and chronic production of reactive oxygen species (ROS), leading to phloem tissue cell death and HLB symptoms^{79–81}. Thus, those proteins will protect CLas from the high ROS levels in the phloem tissues. Another hub protein, CtrA, regulates bacterial cells, such as *Caulobacter crescentus*, to generate cells of different morphologies, with and without flagella⁸². CLas was also reported to be polymorphic in shape⁴⁰. CtrA is a master cell cycle regulator, negatively regulating the septum-inhibiting proteins, altering the location FtsZ from the middle of the cell to a polar end of the replicating cell. This alteration causes the cell to split without symmetry, producing elongated polymorphic cells and what are known as mini cells. CtrA may be involved in the non-flagellated and polymorphic phenotypes of CLas cells in phloem tissues. Nineteen of the 40 hub proteins are uncharacterized, requiring further investigation into their role in CLas biology. However, some hub nodes were given functional associations; for example, CLIBASIA_RS00015 associates with metabolism, DNA replication, and signaling proteins, and CLIBASIA_RS05475 associates with metabolism, enzymes, translation, FeS cluster assembly, and succinate dehydrogenase proteins; these associations were determined from the KEGG protein family guilt by association and operon association analysis. For example, Phyre2

predicted that CLIBASIA_RS00015 is an ATP-dependent deoxyribonuclease subunit. The prediction for CLIBASIA_RS00015 is consistent with the interactome network functional associations assigned as a metabolism protein interacting with DNA replication since ATP-dependent deoxyribonuclease subunits play a role in DNA recombination.

Two hundred and eight binary PPIs, of the 4245 PPIs detected in this study, had been reported in other organisms^{8,9,16–23,64}. On the other hand, 4037 novel interactions were identified in this study, expanding the PPI database. We identified two thousand five hundred eighty-three interactions involving 171 uncharacterized CLAs proteins, providing clues for their functions. For instance, we have assigned protein functions to CLIBASIA_RS03480 as FlgN_{Las}, CLIBASIA_RS03395 as FliK_{Las}, CLIBASIA_RS03285 as TolA_{Las}, and CLIBASIA_RS03885 as YcgR_{Las}. CLIBASIA_RS04860 and CLIBASIA_RS02370 might play a role in cell division since they interact with FtsW_{Las} and FtsQ_{Las}, two essential components of the divisome complex³³. These proteins interact with a hub protein CLIBASIA_RS00450, CLIBASIA_RS05515, and RlpA (septal ring lytic transglycosylase, CLIBASIA_RS04130). RlpA is known to be involved in cell separation and rod shape of bacteria⁸⁴. Thus, even though the functions of CLIBASIA_RS04860, CLIBASIA_RS02370, CLIBASIA_RS00450, and CLIBASIA_RS05515 are unknown, they are probably involved in cell division or cell wall formation, providing hints for further characterization.

We have identified 37 uncharacterized proteins that interact with multiple proteins of an operon, and 61 uncharacterized proteins interacted with two or more proteins from the same KEGG family or function, providing solid indications regarding their roles associated with the known functions of the operons or protein families compared to single interactions. Interestingly, CLAs seems to encode most flagellar genes even though flagella have not been observed in CLAs *in planta*⁴⁰. Though flagella have been observed for CLAs in the psyllid vector, supported by recent plant vs. psyllid transcriptome data showing higher levels of flagellar gene expression in the psyllid versus expression in the plant, The extensive PPIs involving flagellar proteins suggest they are functional and consistent with the active movement of CLAs against the flow of phloem sap⁶⁶.

Overall, we have generated the interactome map for CLAs, which has provided insights regarding the biology and pathogenicity of CLAs and the putative functions of uncharacterized proteins. This interactome will be a valuable resource and provide clues to further characterize unknown proteins' functions. The PPIs identified have the potential to be used as targets for the development of novel antimicrobials to control HLB.

Methods

Yeast two-hybrid vector construction

Full-length CLAs ORFs were PCR amplified using genomic DNA extracted from CLAs-positive citrus leaf tissues. CLAs protein information including Uniprot IDs and annotations are included in Supplementary Data 16. The gel-purified amplicons were fused with the activation (AD) and binding domain (BD) Y2H expression vectors (pGADT7 and pGEO_BD, respectively) using infusion recombination cloning (Clontech). We modified the Gal4 Matchmaker Gold Yeast Two-hybrid system (clontech) pGBK_T7 (Clontech) expression vector to harbor the pGADT7_AD vector's multiple cloning site (MCS) between RE sites NdeI and BamHI (Supplementary Fig. 1B) and was named pGEO_BD. This modification simplified the cloning of all CLAs ORFs into both vectors between RE sites EcoRI and XmaI using one infusion primer set per gene to reduce costs and labor. We directly transformed these AD and BD infusion reactions (~20 µl per 50 µl reaction) into yeast strains Y187 (AD, prey clones; MATa; G1::lacZ, M1::MEL1) and Y2H Gold (BD, bait clones; MATα; G1::HIS3, G2::ADE2, M1::AUR1-C, M1::MEL1; G1, G2, and M1 are Gal4 driven promoters)

following Yeastmaker Yeast Transformation protocol (Takara). We verified transformants with colony PCR and sequencing a random set of ~200 constructs.

Autoactivation test

Autoactivation of the reporter genes by individual ORFs was assessed by mating the transformants with the opposite corresponding mating partner empty vector and plating on selective media double dropout (DDO)/X/AbA (-Leu, -Trp, +XαGal, +Aureobasidin A). Clones that autonomously activated reporter genes were removed from the 96-well plates and high throughput screening.

Yeast two-hybrid three-phase screening

A three-phase 96-well mating scheme was adapted from a matrix approach³³. Briefly, all transformants were cultured individually in 96-well plates in 1 ml of the appropriate minimal liquid media (AD: -Leu; BD: -Trp) at 30 °C until saturated (~36–48 h). For the first mating phase (three phases in total), all individual 1 ml AD cultures were pooled together and centrifuged at 700 g for 5 min. The pellet was resuspended in 25 ml SD/-Leu liquid media with 25% glycerol. Aliquots (1 ml) of the pool were stored at -80 °C. The BD cultures were left in a 96-well format for the phase 1 mating.

Mating procedure

Phase 1. Each AD pool was mated to individual pGEO_BD clones in a 96-well format. The matings were spotted on YPDA agar plates by spotting three µl of the AD pool aliquot and three µl of the individual BD clone on top of one another and incubated at 30 °C for ~48 h, or until the colonies were ~1 mm in diameter. The 1 mm colonies were transferred using a pin replicator to 96-well plates with DDO (-Leu, -Trp) liquid medium to select diploids harboring AD and BD plasmids and incubated at 30 °C and 210 rpm on a rotary shaker. This step reduces the background on the selection plates. After 48 h, 5 µl of the mating cultures were plated on DDO (-Leu, -Trp) agar plates as a mating control, and selective media to determine interaction and interaction intensity: D/X/A (-Leu, -Trp, +XαGal, +Aureobasidin A), Q/X (-Leu, -Trp, -His, -Ade + XαGal), and Q/X/A (-Leu, -Trp, -His, -Ade +XαGal, +Aureobasidin A). All subsequent matings and screenings were carried out in the same manner. Phase 2. Positive BD clones from the phase 1 mating were pooled using the previously described AD pooling method. These BD pools (containing all positive BD interactors from Phase 1) were mated back to the AD plate that comprised the pool of Phase 1. This is to determine which of the AD clones within the pool were the interacting partners to the positive BD interactors from phase 1 (Fig. 1A). Phase 3. The final phase determined the interacting partners from the previous two screenings and allowed for a pairwise rescreening of all interacting constructs. All positive AD and BD yeast constructs interacting in the first two phase screenings were organized into a row (BD) and column (AD) format in 96-well plates. Each row of a 96-well plate is a single BD construct (For example, row A: BD construct #1, row B: BD construct #2, row C: BD construct #3... to row H: BD construct #8) and each column is a single AD construct (For example column 1: AD construct #1, column 2: AD construct #2, column 3: AD construct #3... to column 12: AD construct #12) which were mated pairwise. The row and column format allowed all phase one and two proteins to interact to determine the interacting partners and the interaction specificity. This method helped determine what interactions are more likely or accurate since interactions must be repeatable pairwise and allowed us to weed out sticky interactors.

Analysis of the CLAs Y2H network topology

We visualized, analyzed, and utilized the network developed in this study with Cytoscape network software⁴³. Network topologies, such as betweenness and closeness centrality, degree, and clustering

coefficient, were determined using the Network Analyzer in Cytoscape⁴⁴. We used this information to select hub nodes in the network for further analysis and to understand general CLAs interactome network organization found in the Y2H screening. To validate the CLAs_{whole} network topology, NetworkRandomizer version 3 was used to compare the CLAs_{whole} network to a randomly computed network. We compared the difference between the real network (CLAs_{whole}) and its most similar random network for each centrality⁸⁵. Hub nodes were determined using manual curation, MCODE clustering, and MCC algorithm^{46,86,87}.

Scoring interactions and assigning interaction confidence values

PPIs were scored on a scale from one to three for each selection media. These colony scores were used with iPPIs, PPIs within the same operon, and bait out-degree and prey in-degree to create an additive score for the logistic regression formula to assign interaction confidence scores. The logistic regression training set of 100 iPPIs was used as true positives and 100 PPIs with the lowest combined Y2H score and highest average shortest path link value with randomized bait and prey node degree values as true negatives. We calculated the interaction score (s) using the Eq. (1) four times using different variables for each regression. The variables were the number of times as a bait (b), number of times as a prey (p), and colony score (c). We calculated the interaction scores for each PPI by averaging four interaction scores, which were calculated independently based on the value of the independent variable (c), i.e., interolog score alone, Y2H repeatability, interaction intensity, and interolog score combined with Y2H repeatability together. We compared the scores within each group of interactions and found that they were all within one standard deviation of their group mean⁸⁸. This average score is termed interaction score in the datasets.

$$s = 1 / (1 + e^{-(178.67 - (-2.25484) * b - 0.4624 * p + 506.1149 * c)}) \quad (1)$$

s : interaction score; b : count of occurrence as a bait node; p : count of event as a prey node; c : colony score.

Conserved protein-protein interactions (iPPIs)

CLAs homologs in *A. radiobacter*, *B. subtilis*, *C. jejuni*, *E. coli*, *H. pylori*, *L. crescens*, *M. genitalium*, *M. loti*, *M. pneumoniae*, *S. cerevisiae*, *S. meliloti*, *T. pallidum*, and iCOGs were collected using IMG/MER and NCBI databases^{17,22,36,88,89}. Only protein sequences with an identity of >30% were considered for analysis. All available experimentally determined (i.e., Y2H, AP/MS, and CO-IP) ortholog PPIs and iCOG PPIs were collected from the STRING database for these homologs³⁶. The known and predicted interactions between CLAs proteins were also included in the data collected from STRING³⁶.

Validating CLAs interactions in vivo by a pairwise Y2H screening

We randomly chose 163 ORF constructs (15.9% of the total CLAs ORFs and 30% of the nodes in the CLAs_{whole} network) to use in a pairwise Y2H screening. The interactions from this screening were compared for overlap between the three datasets: the pairwise screening, the three-phase Y2H PPI screening, and the known and predicted interactions between these proteins available from the STRING database. The rates of false positives, false negatives, specificity, and sensitivity within the three-phase screening were determined using the known interactions and pairwise screening. The false positive rate (FPR) was calculated by comparing the number of PPIs detected in the high throughput screening and not found in the pairwise rescreening using the $FPR = FP / (FP + TP)$. The false positive (FP) number is the total PPIs found in the CLAs_{whole} high throughput dataset not found in the STRING database as predicted for the random 163 proteins screened

pairwise. The true positive (TP) value represents the number of protein pairs that did not interact in either screening method. True negative (TN) rates ($TN = n(n-1)/2$; $n =$ PPIs found only in the high throughput screening) are difficult to determine because there is no definitive method for proving whether proteins do or do not interact in vivo. The TN value was calculated using predicted/iPPIs by comparing the high throughput results with the STRING expected PPIs and the 163 random protein pairs rescreened pairwise to increase the likelihood of finding true interactors for reference. The false negative rate (FNR) was calculated similarly ($FNR = FN / (TP + FN)$), where the false negative (FN) value is the number of PPIs detected pairwise but not in the high throughput screening.

The true positive (TP) value is the number of PPIs found in the high throughput and pairwise screenings. The FNR for the CLAs interactome shows the ratio of PPIs expected but not found in the CLAs_{whole}. Specificity and sensitivity were calculated using the above TP, TN, FN values and the false positive (FP) number, which is the number of PPIs detected in the pairwise but not in the high throughput dataset. Sensitivity is the proportion of true PPIs detected of all protein pairs that interact in the known dataset, represented by the formula: $Sensitivity = TP / (TP + FN)$. The specificity represents the proportion of non-interactors correctly excluded from the dataset, represented by the formula: $Specificity = TN / (FP + TN)$.

Confirming CLAs interactions using a pull-down assay

Interactions were validated using MBP/GST pull-down assay. The pGEX-4T-1 (GST) (GE Healthcare) and pMAL-c5x (MBP) (NEB) vectors were used to construct the *E. coli* expression vectors for GST-Glutathione or MBP-Amylose pull-down assay between multiple CLAs proteins. The GST/MBP expression vectors were constructed by amplifying the ORFs from the CLAs yeast expression vectors and inserted between RE sites: BamHI and Sall in the pGEX vector and XmnI and EcoRI in the pMAL vector using infusion cloning (Clontech). These newly generated vector constructs were transformed into *E. coli* Rosetta (DE3) electrocompetent cells, transformants were selected by plating on LB agar plates containing ampicillin (100 µg/ml) and chloramphenicol (30 µg/ml); all subsequent steps used these antibiotics at these concentrations. The transformants were cultured overnight in 3 ml liquid LB. The next day, the culture was diluted 1:100 in 25 ml LB and grown until the culture reached an OD₆₀₀ 0.2–0.3. The culture was induced with IPTG (1 µg/ml) overnight at 16 °C.

After induction, the resulting cell pellet was lysed using 5 µg/ml of pellet with B-PER™ Bacterial Protein Extraction Reagent (Thermo-Fisher) and incubated at room temperature for 20 min while rocking. The lysates were cleared by centrifugation at 4 °C for 20 min at 16,000 g. Supernatants were collected in a fresh tube, and total protein concentration was measured using Bradford assay. The protein expression and solubility were evaluated by both Coomassie stain and western blot (WB) of the GST and MBP fusion proteins using anti-GST (ab92, Abcam, 1:1000 dilution) and anti-MBP (E8032S, NEB, 1:10,000 dilution) monoclonal antibodies before proceeding. The GST-fusion protein and empty-GST supernatants were mixed with the MBP-fusion protein and incubated with glutathione or amylose agarose beads for 3–4 h at 4 °C with rotating. The glutathione or amylose agarose was then washed 5–10 times to remove unbound proteins before boiling and analysis by SDS-PAGE WB using anti-MBP (NEB) and anti-GST (Abcam) monoclonal antibodies followed by secondary antibody Goat Anti-Mouse IgG H&L (HRP) (ab205719, Abcam, 1:10,000). Protein-protein interactions were further confirmed by pull-down assays in Supplementary Table 5.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The datasets generated and/or analysed during the current study are available in the Supplementary Information, Supplementary Data and Source Data files. Source data are provided with this paper.

References

- Yuan, X. et al. Region-wide comprehensive implementation of roguing infected trees, tree replacement, and insecticide applications successfully controls citrus huanglongbing. *Phytopathology* **111**, 1361–1368 (2021).
- Singerman, A. & Rogers, M. E. The economic challenges of dealing with citrus greening: the case of Florida. *J. Integr. Pest Manag.* **11**, 3 (2020).
- Bové, J. M. Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. *J. Plant Pathol.*, 7–37 <https://www.jstor.org/stable/41998278> (2006).
- Wang, N. et al. The *Candidatus liberibacter*-host interface: insights into pathogenesis mechanisms and disease control. *Annu. Rev. Phytopathol.* **55**, 451–482 (2017).
- Duan, Y. et al. Complete genome sequence of citrus huanglongbing bacterium, '*Candidatus Liberibacter asiaticus*' obtained through metagenomics. *Mol. Plant Microbe Interact.* **22**, 1011–1020 (2009).
- Merfa, M. V. et al. Progress and obstacles in culturing '*Candidatus Liberibacter asiaticus*', the bacterium associated with huanglongbing. *Phytopathology* **109**, 1092–1101 (2019).
- Kühner, S. et al. Proteome organization in a genome-reduced bacterium. *Science* **326**, 1235–1240 (2009).
- Ito, T. et al. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* **98**, 4569–4574 (2001).
- Rain, J. C. et al. The protein-protein interaction map of *Helicobacter pylori*. *Nature* **409**, 211–215 (2001).
- Cossar, P. J., Lewis, P. J. & McCluskey, A. Protein-protein interactions as antibiotic targets: a medicinal chemistry perspective. *Med. Res. Rev.* **40**, 469–494 (2020).
- Giot, L. et al. A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736 (2003).
- Li, S. et al. A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543 (2004).
- Walhout, A. J. et al. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**, 116–122 (2000).
- Braun, P. et al. An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* **6**, 91–97 (2009).
- Lin, L. et al. Protein-protein interactions in two potyviruses using the yeast two-hybrid system. *Virus Res.* **142**, 36–40 (2009).
- Wuchty, S. et al. The protein interactome of *Streptococcus pneumoniae* and bacterial meta-interactomes improve function predictions. *mSystems* **2**, e00019–17 (2017).
- Titz, B. et al. The binary protein interactome of *Treponema pallidum* the syphilis spirochete. *PLoS One* **3**, e2292 (2008).
- Parrish, J. R. et al. A proteome-wide protein interaction map for *Campylobacter jejuni*. *Genome Biol.* **8**, R130 (2007).
- Sato, S. et al. A large-scale protein-protein interaction analysis in *Synechocystis* sp. PCC6803. *DNA Res.* **14**, 207–216 (2007).
- Wang, Y. et al. Global protein-protein interaction network in the human pathogen *Mycobacterium tuberculosis* H37Rv. *J. Proteome Res.* **9**, 6665–6677 (2010).
- Shimoda, Y. et al. A large scale analysis of protein-protein interactions in the nitrogen-fixing bacterium *Mesorhizobium loti*. *DNA Res.* **15**, 13–23 (2008).
- Rajagopala, S. V. et al. The binary protein-protein interaction landscape of *Escherichia coli*. *Nat. Biotechnol.* **32**, 285–290 (2014).
- Marchadier, E. et al. An expanded protein-protein interaction network in *Bacillus subtilis* reveals a group of hubs: exploration by an integrative approach. *Proteomics* **11**, 2981–2991 (2011).
- Häuser, R. et al. A second-generation protein-protein interaction network of *Helicobacter pylori*. *Mol. Cell Proteom.* **13**, 1318–1329 (2014).
- Meseguer, M. A. et al. *Mycoplasma pneumoniae*: a reduced-genome intracellular bacterial pathogen. *Infect Genet. Evol.* **3**, 47–55 (2003).
- Schwikowski, B., Uetz, P. & Fields, S. A network of protein-protein interactions in yeast. *Nat. Biotechnol.* **18**, 1257–1261 (2000).
- Galperin, M. Y. & Koonin, E. V. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Res.* **32**, 5452–5463 (2004).
- Barabási, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
- Typas, A. & Sourjik, V. Bacterial protein networks: properties and functions. *Nat. Rev. Microbiol.* **13**, 559–572 (2015).
- Yook, S. H., Oltvai, Z. N. & Barabási, A. L. Functional and topological characterization of protein interaction networks. *Proteomics* **4**, 928–942 (2004).
- Cao, F. Y. et al. A host-pathogen interactome uncovers phyto-pathogenic strategies to manipulate plant ABA responses. *Plant J.* **100**, 187–198 (2019).
- Crua Asensio, N., Muñoz Giner, E., de Groot, N. S. & Torrent Burgas, M. Centrality in the host-pathogen interactome is associated with pathogen fitness during infection. *Nat. Commun.* **8**, 14092 (2017).
- Zhong, J., Zhang, H., Stanyon, C. A., Tromp, G. & Finley, R. L. Jr. A strategy for constructing large protein interaction maps using the yeast two-hybrid system: regulated expression arrays and two-phase mating. *Genome Res.* **13**, 2691–2699 (2003).
- Goll, J. & Uetz, P. The elusive yeast interactome. *Genome Biol.* **7**, 223 (2006).
- Rajagopala, S. V. Mapping the protein-protein interactome networks using yeast two-hybrid screens. *Adv. Exp. Med. Biol.* **883**, 187–214 (2015).
- Szklarczyk, D. et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–d612 (2021).
- Galperin, M. Y. et al. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* **49**, D274–d281 (2021).
- Venkatesan, K. et al. An empirical framework for binary interactome mapping. *Nat. Methods* **6**, 83–90 (2009).
- Choi, S. G. et al. Maximizing binary interactome mapping with a minimal number of assays. *Nat. Commun.* **10**, 3907 (2019).
- Andrade, M. O. et al. The flagella of '*Candidatus Liberibacter asiaticus*' and its movement in planta. *Mol. Plant Pathol.* **21**, 109–123 (2020).
- Zotenko, E., Mestre, J., O'Leary, D. P. & Przytycka, T. M. Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. *PLoS Comput. Biol.* **4**, e1000140 (2008).
- Hart, G. T., Lee, I. & Marcotte, E. R. A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality. *BMC Bioinform.* **8**, 236 (2007).
- Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
- Assenov, Y., Ramírez, F., Schelhorn, S. E., Lengauer, T. & Albrecht, M. Computing topological parameters of biological networks. *Bioinformatics* **24**, 282–284 (2008).

45. Kelkar, Y. D. & Ochman, H. Genome reduction promotes increase in protein functional complexity in bacteria. *Genetics* **193**, 303–307 (2013).
46. Jeong, H., Mason, S. P., Barabási, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
47. Chin, C. H. et al. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **8**, S11 (2014).
48. Doyle, J. R. MCC—multiple correlation clustering. *Int. J. Man-Machine Stud.* **37**, 751–765 (1992).
49. Lai, K. K., Davis-Richardson, A. G., Dias, R. & Triplett, E. W. Identification of the genes required for the culture of *Liberibacter crescens*, the closest cultured relative of the liberibacter plant pathogens. *Front. Microbiol.* **7**, 547 (2016).
50. Zhang, R., Ou, H. Y. & Zhang, C. T. DEG: a database of essential genes. *Nucleic Acids Res.* **32**, D271–D272 (2004).
51. Consortium, G. O. Gene ontology consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).
52. Du, P. et al. “*Candidatus Liberibacter asiaticus*” secretes non-classically secreted proteins that suppress host hypersensitive cell death and induce expression of plant pathogenesis-related proteins. *Appl. Environ. Microbiol.* **87**, e00019–e00021 (2021).
53. Prasad, S., Xu, J., Zhang, Y. & Wang, N. SEC-translocon dependent extracytoplasmic proteins of *Candidatus Liberibacter asiaticus*. *Front. Microbiol.* **7**, 1989 (2016).
54. Nguyen, P. V., Srihari, S. & Leong, H. W. Identifying conserved protein complexes between species by constructing interolog networks. *BMC Bioinform.* **14**, S8 (2013).
55. Oliver, S. Guilt-by-association goes global. *Nature* **403**, 601–603 (2000).
56. Moreno-Hagelsieb, G. The power of operon rearrangements for predicting functional associations. *Comput. Struct. Biotechnol. J.* **13**, 402–406 (2015).
57. Aldridge, P., Karlinsey, J. & Hughes, K. T. The type III secretion chaperone FlgN regulates flagellar assembly via a negative feedback loop containing its chaperone substrates FlgK and FlgL. *Mol. Microbiol.* **49**, 1333–1345 (2003).
58. Eggenhofer, E., Rachel, R., Haslbeck, M. & Scharf, B. MotD of *Sinorhizobium meliloti* and related alpha-proteobacteria is the flagellar-hook-length regulator and therefore reassigned as FlkI. *J. Bacteriol.* **188**, 2144–2153 (2006).
59. Moriya, N., Minamino, T., Imada, K. & Namba, K. Genetic analysis of the bacterial hook-capping protein FlgD responsible for hook assembly. *Microbiology (Reading)* **157**, 1354–1362 (2011).
60. Hirano, T., Yamaguchi, S., Oosawa, K. & Aizawa, S. Roles of FlkI and FlhB in determination of flagellar hook length in *Salmonella typhimurium*. *J. Bacteriol.* **176**, 5439–5449 (1994).
61. Osbourn, A. E. & Field, B. Operons. *Cell Mol. Life Sci* **66**, 3755–3775 (2009).
62. Wang, N., Lu, S. E., Yang, Q. W., Sze, S. H. & Gross, D. C. Identification of the *syr-syp* box in the promoter regions of genes dedicated to syringomycin and syringopeptin production by *Pseudomonas syringae* pv. *syringae* B301D. *J. Bacteriol.* **188**, 160–168 (2006).
63. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
64. Uetz, P. et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
65. Pandey, S. S. & Wang, N. Targeted early detection of citrus huanglongbing causal agent ‘*Candidatus Liberibacter asiaticus*’ before symptom expression. *Phytopathology* **109**, 952–959 (2019).
66. Pandey, S. S., F, N. C. V. & Wang, N. Spatiotemporal dynamics of *Candidatus Liberibacter asiaticus* colonization inside citrus plant and Huanglongbing disease development. *Phytopathology* <https://doi.org/10.1094/phyto-09-20-0407-r> (2020).
67. Tan, J. et al. Structural basis of assembly and torque transmission of the bacterial flagellar motor. *Cell* **184**, 2665–2679.e2619 (2021).
68. Young, G. M., Schmiel, D. H. & Miller, V. L. A new pathway for the secretion of virulence factors by bacteria: the flagellar export apparatus functions as a protein-secretion system. *Proc. Natl Acad. Sci. USA* **96**, 6456–6461 (1999).
69. Tsigotaki, A., De Geyter, J., Šoštaric, N., Economou, A. & Karamanou, S. Protein export through the bacterial Sec pathway. *Nat. Rev. Microbiol.* **15**, 21–36 (2017).
70. Ghanim, M., Fattah-Hosseini, S., Levy, A. & Cilia, M. Morphological abnormalities and cell death in the Asian citrus psyllid (*Diaphorina citri*) midgut associated with *Candidatus Liberibacter asiaticus*. *Sci. Rep.* **6**, 33418 (2016).
71. Chen, Y.-C., Rajagopala, S. V., Stellberger, T. & Uetz, P. Exhaustive benchmarking of the yeast two-hybrid system. *Nat. Methods* **7**, 667–668 (2010).
72. Henderson, B. & Martin, A. Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. *Infect Immun.* **79**, 3476–3491 (2011).
73. Necci, M., Piovesan, D. & Tosatto, S. C. E. Critical assessment of protein intrinsic disorder prediction. *Nat. Methods* **18**, 472–481 (2021).
74. Gollapalli, P., G, T. S., H, M., Shetty, P. & N, S. K. Network topology analysis of essential genes interactome of *Helicobacter pylori* to explore novel therapeutic targets. *Microb Pathog* **158**, 105059 (2021).
75. Yu, H. et al. High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008).
76. De Francesco, A. et al. Transcriptome profiling of ‘*Candidatus Liberibacter asiaticus*’ in citrus and psyllids. *Phytopathology* **112**, 116–130 (2022).
77. Qiu, D. et al. NAD(P)H: quinone oxidoreductase 1 attenuates oxidative stress and apoptosis by regulating Sirt1 in diabetic nephropathy. *J. Transl. Med.* **20**, 44 (2022).
78. Reen, F. J., Haynes, J. M., Mooij, M. J. & O’Gara, F. A non-classical LysR-type transcriptional regulator PA2206 is required for an effective oxidative stress response in *Pseudomonas aeruginosa*. *PLoS One* **8**, e54479 (2013).
79. Ma, W. et al. Citrus Huanglongbing is a pathogen-triggered immune disease that can be mitigated with antioxidants and gibberellin. *Nat. Commun.* **13**, 529 (2022).
80. Gao, Y. et al. Citrus genomic resources unravel putative genetic determinants of Huanglongbing pathogenicity. *iScience* **26**, 106024 (2023).
81. Pandey, S. S., Xu, J., Achor, D. S., Li, J. & Wang, N. Microscopic and transcriptomic analyses of early events triggered by ‘*Candidatus Liberibacter asiaticus*’ in young flushes of huanglongbing-positive citrus trees’. *Phytopathology* **113**, 985–997 (2023).
82. Siam, R. & Marczyński, G. T. Cell cycle regulator phosphorylation stimulates two distinct modes of binding at a chromosome replication origin. *Embo j* **19**, 1138–1147 (2000).
83. Errington, J. & Wu, L. J. Cell cycle machinery in *Bacillus subtilis*. *Subcell Biochem.* **84**, 67–101 (2017).
84. Jorgenson, M. A., Chen, Y., Yahashiri, A., Popham, D. L. & Weiss, D. S. The bacterial septal ring protein RlpA is a lytic transglycosylase that contributes to rod shape and daughter cell separation in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **93**, 113–128 (2014).
85. Tosadori, G., Bestvina, I., Spoto, F., Laudanna, C. & Scardoni, G. Creating, generating and comparing random network models with NetworkRandomizer. *F1000Res.* **5**, 2524 (2016).
86. Morris, J. H. et al. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinform.* **12**, 436 (2011).
87. Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.* **4**, 2 (2003).

88. Markowitz, V. M. et al. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25**, 2271–2278 (2009).
89. Johnson, M. et al. NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9 (2008).

Acknowledgements

We thank Wang lab members for constructive suggestions and insightful discussions. This project was supported by funding from Florida Citrus Initiative Program, Citrus Research and Development Foundation, U.S. Department of Agriculture National Institute of Food and Agriculture grants 2022-70029-38471, 2021-67013-34588, 2018-70016-27412 and 2016-70016-24833, FDACS Specialty Crop Block Grant Program, and Hatch project [FLA-CRC-005979] (N. Wang).

Author contributions

N.W. and E.W.C. conceptualized and designed the experiments. E.W.C. and O.G.P. designed and constructed the pGEO_BD vector and oligonucleotides. E.W.C. performed the experiments. E.W.C. and N.W. wrote the manuscript with input from all co-authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-43648-7>.

Correspondence and requests for materials should be addressed to Nian Wang.

Peer review information *Nature Communications* thanks Michelle Heck, Peter Uetz and the other, anonymous, reviewer for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023