Direct DNA sequencing of ‘*Candidatus Liberibacter asiaticus*’ from *Diaphorina citri*, the Asian citrus psyllid, and its implications for citrus greening disease management

Steven Adam Higgins^{1*}, Marina Mann², and Michelle Heck^{1,2}

¹Emerging Pests and Pathogens Research Unit, United States Department of Agriculture
Agricultural Research Service (USDA ARS), Ithaca, NY 14853

²Plant Pathology and Plant Microbe Biology Department, Cornell University, Ithaca, NY 14853

Send correspondence to:

Michelle Heck

michelle.cilia@usda.gov

Abstract

The Asian citrus psyllid, *Diaphorina citri* is an invasive insect¹ and a vector of 'Candidatus Liberibacter asiaticus' (CLas), a bacterium whose growth in *Citrus* species results in citrus greening disease^{2,3}. Methods to enrich and sequence CLas from *D. citri* often rely on biased genome amplification⁴ and nevertheless contain significant quantities of host DNA^{5,6}. To overcome these hurdles, we developed a simple pre-treatment DNase and filtration (hereafter PDF) protocol to directly sequence CLas and the complete, primarily uncultivable, microbiome from *D. citri* adults. The PDF protocol yielded CLas abundances upwards of 60% and enabled detection of 156 genetic variants in these strains compared to progenitor strains in Florida, which included prophage encoding regions with key functions in CLas pathogenesis, putative antibiotic resistance loci, and a single secreted effector. These variants suggest laboratory propagation of CLas may result in different phenotypic trajectories among laboratories, and may confound CLas physiology or therapeutic design and evaluation if these differences remain undocumented. Finally, we obtained genetic signatures affiliated with *Citrus* nuclear and organellar genomes, entomopathogenic fungal mitochondria, and commensal bacteria from laboratory-reared and field-collected *D. citri* adults. Hence, the PDF protocol can inform agricultural management strategies related to pathogen evolution⁷, insect microbiome surveillance⁸, antibiotic resistance screening⁹, and gut content analysis¹⁰.

Introduction

The abundance of host DNA in nucleic acid extracts is a significant barrier to cost-effective sequence analysis of host-associated microbiomes¹¹. In addition, the non-uniform abundance and spatial organization of a host's microbial populations significantly impedes direct sequence

analysis of host-associated microbiomes. Sequencing effort is therefore directed towards host DNA and limits the cost-effectiveness and utility of shotgun sequence data in microbiome analysis¹². To help overcome these issues, a combination of pre- and post-DNA extraction methods have been devised to limit host DNA in nucleic acid extracts¹¹. These methods include differential lysis of host tissues followed by exonuclease treatment¹¹, host DNA depletion via methylation status followed by multiple displacement amplification^{13–15}, use of targeted hybridization probes for selective enrichment of microbial DNA^{16,17}, and cell enrichment by cultivation in alternative hosts¹⁸. While useful, these methods are relatively costly, laborious, and can restrict analysis to a single taxon (e.g., probe hybridization) or introduce biases (e.g., using multiple displacement amplification) that limit the types of downstream analyses afforded by shotgun sequence data⁴.

One host-microbe system that illustrates the impediment of host DNA in microbiome analyses is that of citrus greening disease pathosystem^{6,19}. Citrus trees worldwide are threatened by this rapidly spreading disease, which over a period of years significantly reduces citrus fruit yield, marketability, and eventually results in tree decline and death^{2,3,20}. Citrus greening disease is estimated to cost some U.S. citrus growers \$1 billion annually and has reduced Florida citrus production by over 70% since the early 2000s^{21,22}. In the US, the disease is attributed to the growth of the obligately biotrophic, gram-negative bacterium ‘*Candidatus Liberibacter asiaticus*’ (CLas) within citrus trees, which induces phloem-plugging and results in the characteristic citrus greening pathology of aborted ripening, blotchy leaf mottle, and tree root decline^{2,20,22}. CLas is proposed to be transmitted by its insect vector, *Diaphorina citri* Kuwayama (Hemiptera: Liviidae) the Asian citrus psyllid in a circulative propagative manner which involves extensive interaction with and replication the vector’s tissues^{23,24}. *D. citri* acquire CLas during the

nymphal stage and become competent to transmit as adults when CLas titer reaches high levels in the insect's salivary glands²⁴. Like the majority of phloem-feeding hemipteran insects, *D. citri* harbor three known bacterial symbionts whose functions in their biology have been inferred from genome sequencing: *Wolbachia pipientis* strain wDi²⁵, 'Candidatus Proffittella armatura' a symbiont with predicted nutritional and defensive functions^{26,27} and 'Candidatus Carsonella rudii' a nutritional symbiont^{25,28}. These latter two symbionts are essential for *D. citri* survival but their role in CLas transmission is speculative²⁹. *D. citri* have spread CLas throughout citrus growing regions in the U.S. since its initial discovery in Florida in 2005². While we report on the impacts of citrus greening to the US citrus industry for simplicity, we would like to emphasize that this disease presents a significant threat to citrus production globally^{3,30}. Particularly alarming is that a handful of countries globally produce a majority of the world's citrus crops^{3,30}, and that some at-risk countries possess large swaths of habitat favoring the spread of citrus greening disease^{30,31}.

Historically, the cultivation of microbial plant pathogens has facilitated their experimentation and comparative sequence analysis, or design of therapeutics to control their spread, while currently uncultivable *Ca. Liberibacter* pathogens (like CLas) are difficult to study due to their obligately biotrophic lifestyle³². Since attempts to isolate and continuously culture CLas have been unsuccessful³³, high-throughput sequencing has become an invaluable tool for hypothesis generation and performing metabolic, comparative genomics, and evolutionary analyses of this pathogenic bacterium^{2,34,35}. Nevertheless, many hurdles to efficient and direct sequencing of CLas populations from both citrus and *D. citri* exist^{6,19}. For example, due to the overabundance of host DNA in nucleic acid extracts from citrus and psyllid samples, multiple enrichment strategies that include deep sequencing, multiple displacement amplification of host-depleted

DNA, and antibody- or RNA-probe capture hybridization are employed to facilitate sequence analysis of CLas and other *Ca. Liberibacter* populations^{16,19,33,36,37}.

Robust and facile sequencing of CLas and other pathogenic, host-associated microbial populations is necessary to keep pace with the rapid development and application of therapeutics to prevent their spread³⁸. While a variety of antimicrobials have been applied to prevent the spread of citrus greening disease^{21,38}, we are currently unable to perform cost-effective and rapid sequencing of CLas populations to screen for genetic resistances to therapeutics if and when they arise. Furthermore, methods to identify and compare CLas populations have been limited to a few dozen samples that have been sequenced worldwide³⁴. This small sample size limits the identification and development of targeted therapeutics and hampers our ability to contrast standing genetic variation detected in laboratory-propagated CLas populations with those actively spreading in agricultural settings. Thus, we developed a simple and efficient pre-treatment DNase and filtration (PDF) protocol to sequence CLas and other bacterial populations directly from individual members of their insect host, *D. citri*. We compare and contrast our method with existing sequencing methods used to sequence CLas populations and present results of our genomic analyses enabled by the PDF protocol on both lab-reared and field-collected *D. citri* sampled from a minimally-managed citrus grove in Florida. Specifically, we report on replication rate estimates and genetic polymorphisms detected among microbial populations within individual *D. citri* adults and discuss their implications for citrus greening disease management and insect microbiome research. The PDF protocol enables simultaneous analysis of all microbial community members in a single *D. citri* adult, from symbionts to commensals, that will greatly inform diverse research fields, such as therapeutics development and application

⁹, holobiont theory ³⁹, and ecological and evolutionary investigations of *D. citri* and other agriculturally relevant pests ^{7,8}.

Methods

Rearing and collection of laboratory-reared *D. citri* adults. Colonies of the Asian citrus psyllid, *Diaphorina citri* Kuwayama (Hemiptera: Liviidae), were reared in controlled growth chambers on CLas-infected *Citrus medica* (citron) plants under 14:10 h light:dark cycle at 28 °C. Citrus plants were regularly monitored by lab technicians and maintained by pruning and watering on a weekly basis or as needed. Individual *D. citri* adults from CLas-infected citron colonies were bulk-collected using a sterile 25 ml serological pipette under vacuum into a sterile 50 ml falcon tube. The falcon tube containing *D. citri* adults was then transferred onto ice to immobilize the insects prior to handling and sorting. Immobilized *D. citri* were then transferred into a plastic weigh boat on ice and individual *D. citri* were placed into sterile 2 ml microcentrifuge tubes using a 70% (v/v) ethanol-wiped, fine-tipped paintbrush and kept on ice until processing for DNA extraction.

Field collection of *D. citri* adults. With permission from a local citrus grower, *D. citri* adults were collected from *Citrus sinensis* (Valencia orange) trees in a minimally managed citrus grove located in St. Lucie county, FL, in early October of 2018. This grove did not utilize pesticides and natural enemies of *D. citri* were frequently observed (e.g., spiders and parasitoid wasps) ⁴⁰. Individual *D. citri* were collected from citrus leaf surfaces using mouth aspiration into sterile borosilicate glass collection vials from a few dozen citrus trees. The individual *D. citri* were then transferred into 95% ethanol (v/v) and kept stored on ice. In the lab, the ethanol was decanted

and the *D. citri* adults were air dried on Kimwipes until storage at -80 °C prior to DNA extraction.

Pre-treatment DNase and filtration (PDF) protocol and sequencing. To each tube containing an individual *D. citri* adult, we added 100 µl of master mix containing 8 U DNase I (TURBO DNA-free kit, catalog no. AM1907, Thermo Fisher Scientific, Waltham, MA, USA), 1X DNase I buffer, 2X ProteoGuard EDTA-free protease inhibitor cocktail (catalog no. 635673, Takara Bio USA, Inc., San Jose, CA, USA), and DNase and RNase-free RPI molecular grade water (catalog no. 248700-500.0, Neta Scientific, Inc., Hainesport, NJ, USA). Psyllids were then manually homogenized within each tube for 30 sec using an autoclaved Bel-Art polypropylene microcentrifuge pestle (catalog no. 199230001, SP Industries, Inc., South Wayne, NJ, USA). Following homogenization, samples were briefly centrifuged and gently flicked with a finger to mix the tube contents, followed by incubation in a water bath at 32 °C for 2 h. Tubes were gently mixed by hand agitation during incubation at 30 min intervals. Following this incubation, samples were placed on ice and 100 µl of sterile 1X TE buffer (10 mM Tris HCl, 1 mM EDTA, pH 8.0) was added. The total mixture was then transferred to autoclaved, 1.5 ml plastic microcentrifuge tubes containing a 5 µm pore size PVDF filter cartridge (catalog no. UFC30SV00, MilliporeSigma, Burlington, MA, USA) followed by centrifugation at 8,000 x g for 3 min. The supernatant was carefully removed by pipetting and the resulting cell pellet was placed on ice. DNA was then extracted from the cell pellet by addition of 600 µl of buffer RLT (catalog no. 79216, Qiagen Sciences, Inc., Germantown, MD, USA) followed by gentle pipetting to reconstitute the cell pellet. The mixture was then transferred to a sterile 2 ml microcentrifuge tube containing pre-autoclaved and air-dried 0.1 mm zirconia-silicate beads (catalog no.

11079101Z, Bio Spec Products, Inc., Bartlesville, OK, USA) and homogenized using a mixer mill 400 (Retsch GmbH, Haan, Germany) at 25 Hz for 1 min. A 450 μ l aliquot of the homogenate was then transferred to a sterile 2 ml EconoSpin mini spin column (catalog no. 1910-250, Epoch Life Science, Inc., Sugar Land, TX, USA) containing 450 μ l of 70% (v/v) molecular grade ethanol (catalog no. 3916EA, Decon Laboratories, Inc., King of Prussia, PA, USA) and gently pipetted up and down to mix. The samples were then centrifuged at 8,000 rpm for 1 min, the flow-through discarded, and an additional 700 μ l of 75% (v/v) molecular grade ethanol was added to the column followed by centrifugation at 8,000 rpm for 30 sec and the flow-through discarded. This process was repeated an additional two times, followed by a 2 min centrifugation of the empty column at 12,000 rpm to remove residual ethanol. The DNA was eluted off the column using 25 μ l of 0.1X TE buffer and DNA concentration and quality evaluated using a NanoDrop 2000 instrument (Thermo Fisher Scientific, Inc.). DNA samples were provided to the [Microbial Genome Sequencing Center](#) (Pittsburgh, PA, USA) for 2 x 151 bp Illumina sequencing at 200 Mbp or 1 Gbp depth offerings. Of note, the DNA concentrations range of these samples were very low (~0.1 – 0.6 ng/ μ l). In earlier iterations of this workflow, preparation of sequencing libraries were unsuccessful using DNA samples not stored or shipped on ice. Any DNA derived from this method should therefore be kept on ice when handled for in-house sequencing or shipped on ice or dry ice to sequencing facilities to ensure stability of the minute DNA quantities present in the processed samples. In addition to the PDF protocol, *D. citri* adults were placed into a sterile 2 ml microcentrifuge tube containing two sterile 3.2 mm steel beads and were snap frozen in liquid nitrogen. Next, samples were cryoground on the Retsch mixer mill 400, their DNA extracted, and samples submitted for sequencing as described

above. We refer to this last step as the “Direct protocol” and compare and contrast results from both the PDF and Direct protocols in the results section (Supplementary Fig. 1).

Quantitative Polymerase Chain Reaction (qPCR) analysis. Prior to sequencing, DNA samples were analyzed by qPCR using gene specific PCR primers to monitor CLas enrichment relative to DNA from its *D. citri* host. The PCR primers targeted either the CLas 16S rRNA gene (5'-TCGAGCGCGTATGCGAATAC-3' and 5'- GCGTTATCCCGTAGAAAAAGGTAG -3' forward and reverse primers, respectively) or the *D. citri* wingless (*wg*) gene (5'- GCTCTCAAAGATCGGTTTGACGG -3' and 5'-GCTGCCACGAACGTTACCTTC-3' forward and reverse primers, respectively). PCR amplification was performed using 2X PowerSYBR green master mix (catalog no. 4367695, Thermo Fisher Scientific, Inc.). Briefly, the qPCR reaction was composed of 5 µl of 2X PowerSYBR green master mix, 0.25 µl each of the forward and reverse primers (10 µM stock), 0.125 µl of molecular grade bovine serum albumin (20,000 ng/µl stock, catalog no. B9000S, New England Biolabs, Inc., Ipswich, MA, USA), 1 µl of DNA, and 3.375 µl of molecular grade water. The qPCR amplification was performed on an Applied Biosystems QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher Scientific, Inc.). Thermocycler conditions consisted of 2 min incubation at 50 °C followed by 10 min incubation at 95 °C and 40 cycles at 95 °C for 15 s and 55 °C for 1 min followed by denaturation at 95 °C for 15 sec and melt curve analysis from 60 °C (1 min) to 95 °C (15 sec) at a rate 0.05 °C/sec. The qPCR C_q values were converted to CLas 16S rDNA gene copies using duplicate standard curves comprised of eight, ten-fold serial dilutions of a known concentration of pUC57 plasmid (2,710 bp length) containing a CLas 16S rDNA gene fragment (427 bp length) prior to comparison with *D. citri* wingless genes. A similar methodology was used to calculate *D. citri* *wg* gene copies,

except that genomic DNA was extracted as previously described⁴¹ from a pool of *D. citri* adults (n = 10) followed by six ten-fold serial dilutions to construct an *D. citri* genomic DNA standard curve. An estimated *D. citri* genome size of 485,705,802 bp was used to calculate *wg* gene copies for a given quantity of DNA in the dilution series. The range of CLas 16S rDNA gene copies in the standard curve dilution series were 8.4×10^2 to 8.4×10^9 , and the assay possessed a slope of -3.13, y-intercept of 39.70, R^2 of 0.986, and efficiency of 108.7%. The range of *wg* gene copies in the standard curve dilution series were 2.5×10^{-1} to 2.5×10^5 , and the *wg* qPCR assay possessed a slope of -2.98, y-intercept of 37.96, adjusted R^2 of 0.95, and efficiency of 118.37%.

Data acquisition and analysis. Raw read data sets of CLas genomes derived from citrus tissues were accessed from the National Center for Biotechnology and Information (NCBI) Sequence Read Archive (SRA)⁴² using accession numbers SRR016816, SRR7050882, SRR8177702, SRR9641236, and SRR9673120. These Illumina reads represent CLas genomic data from an unknown CLas strain identified in *Citrus sinensis* in Florida, USA, strain TX1712 collected from *C. sinensis* in Texas, USA⁴³, strain YNJS7C identified in *C. sinensis* from Yunnan province, China⁴⁴, strain AHCA17 identified in *C. maxima* (pummelo) in California, USA⁴⁵, and strain TaiYZ2 collected from *C. maxima* (pummelo) in Songkhla province, Thailand⁴⁶, respectively.

The latest *D. citri* (version 3) and *C. medica* (version 1, HZAU) and *C. sinensis* genome (version 2, HZAU) assemblies were downloaded from citrusgreening.org and citrusgenomedb.org, respectively. All raw fastq sequence files were processed with FaQCs (v2.10)⁴⁷ to trim (phred score cutoff ≥ 30) and remove low quality and short (<50 bp length) Illumina reads. Individual samples and a total combined sample set were assembled or co-assembled, respectively, using the SPAdes assembler (v3.14.0)⁴⁸ with k-mer sizes of 33, 55, and

77 for individual samples and 77, 99, and 127 for the co-assembly (all raw reads from individual samples pooled prior to assembly). Trimmed Illumina reads were aligned to assembled scaffolds using bowtie2 (v2.3.4)⁴⁹ and coverage calculated using samtools (v1.9)⁵⁰ mpileup and coverage modules with default settings. All scaffolds were aligned to the nt database using BLAST+ (v2.11.0)⁵¹ and taxonomic assignments inferred using BASTA (v 1.4.1)⁵² “sequence” module with settings -i 80 -b True -p 90. Blobtools (v1.1.1)⁵³ was used to visualize GC%, coverage, and taxonomic assignments of assembled scaffolds (data not shown), and inspired the plots used in portions of the results and Supplemental Information. Automatic binning of assembled scaffolds into coherent operational taxonomic units was performed using LAST (version 1145)⁵⁴ alignments and BASTA taxonomic assignments, followed by bin consistency analysis using the CheckM (v1.1.3) tool’s “lineage_wf” module to assess the bin completeness and consistency metrics⁵⁵. QAST (v5.1.0rc1, settings -m 100 -f -rna-finding -b)⁵⁶ was used to compare assembled genome bins of *D. citri* symbionts with high quality, complete reference symbiont genomes and report assembly quality statistics.

The iRep tool⁵⁷ with default settings was used to estimate the index of replication (average, instantaneous population replication rate) against representative reference genomes for CLas, *Wolbachia pipientis* strain wDi, ‘*Candidatus Proffella armatura*’ (hereafter Prof), and ‘*Candidatus Carsonella ruddii*’ (hereafter Car) populations detected within individual *D. citri* adults processed in the present study. Snippy (v4.6.0)⁵⁸ was used with parameters –mincov 10 –minfrac 9 and –minqual 100 to detect, enumerate, and compare genetic polymorphisms among *D. citri* symbionts and reference symbiont genomes. Gubbins (v2.4.1)⁵⁹ was used with settings -tree_builder raxml --raxml_model GTRGAMMA and additional default settings to flag and generate a recombination-free single nucleotide polymorphism (SNP) alignment among whole

CLas genomes assembled in the present study and publicly available reference genomes, which was supplied to IQ-TREE (v1.6.1)⁶⁰. IQ-TREE was run with settings -m K3Pu+F -bb 1000 -bnni to generate a SNP phylogeny. Some CLas genomes (both publicly available and in the present study) were highly fragmented and/or missing >25% of the genome and were not included in the phylogenetic analysis of SNPs. Statistics were performed in R 4.1.2⁶¹. Visualization of SNP data and phylogeny was performed using the R packages BioCircos v0.3.4⁶² and ggtree⁶³.

Data availability. Raw Illumina sequence data were deposited to NCBI Sequence Read Archive under BioProject no. PRJNA779156. The assembled data, their taxonomic assignments, SNP alignment and phylogeny, and all data inputs required to reproduce the figures presented in the manuscript are accessible in the figshare repository <https://doi.org/10.6084/m9.figshare.c.5810090.v1>.

Code availability. All scripts used to process data and generate Fig.s can be found in the figshare repository <https://doi.org/10.6084/m9.figshare.c.5810090.v1>.

Results

We introduced a variety of methods, including deep sequencing, multiple displacement amplification, methylation-based DNA depletion, and probe capture hybridization, that are available to enrich and sequence CLas from both citrus and *D. citri* samples. While limited comparisons using reported total read counts or the number of read alignments from the literature were possible (Supplementary Table 1, Fig. 1a), access to raw sequence data (i.e., Illumina

reads) for most published CLas genomes were unavailable on public sequence repositories, such as the SRA⁴². Therefore, we compared sequence data produced using the PDF protocol, which uses pre-extraction DNase treatment and filtration to remove host DNA, with data produced by direct DNA extraction from CLas-infected *D. citri* adults (Direct protocol) (Supplementary Fig. 1). We reasoned that robust comparisons of CLas-infected *D. citri* adults with variable CLas titers could be made using the PDF and Direct protocols whereas the same analysis was not feasible with other published methods since publicly available raw data sets were lacking. Regardless, improvements on baseline host contamination facilitated by the PDF protocol were made possible by comparison to the Direct protocol.

To determine the effect of the PDF protocol on CLas titers relative to the insect host *D. citri* DNA, we quantified CLas 16S rRNA and *D. citri* wingless (*wg*) gene copies in DNA samples generated using the PDF protocol and compared these values with the Direct protocol. The values of CLas 16S rRNA gene copies for the PDF and Direct protocols ranged from 1.1×10^5 to 3.2×10^7 and 1.4×10^3 to 1.6×10^7 , respectively, whereas *D. citri* *wg* gene copies ranged between 8.7×10^2 to 2.4×10^3 and 1.3×10^5 to 1.2×10^6 , respectively. Thus, *D. citri* genomic DNA (which encodes the *wg* gene) was reduced by up to three orders of magnitude using the PDF protocol compared to the Direct protocol (Supplementary Fig. 2).

To assess the impact of host DNA removal on sequence coverage of CLas populations, we sequenced DNA samples prepared using the PDF and Direct protocols that possessed a range of CLas 16S rRNA gene copies for Illumina shotgun sequencing. The reduction in *D. citri* *wg* gene copies we detected also corresponded to a significant reduction in the average coverage of the *D. citri* genome (mean coverage of 0.3 and 0.9 for the PDF and Direct protocols, respectively; $t_{\text{welch}} = 8.6$, $df = 33$, $p = 6.7 \times 10^{-10}$) and a concomitant increase in sequencing effort afforded to CLas

populations (Supplementary Fig. 2). In fact, when we compared the PDF protocol with existing protocols for sequencing CLas genomes available in the literature (Supplementary Table 1), the PDF protocol provided comparable coverage of CLas populations at orders of magnitude lower sequencing depth (Fig. 1a). Even for samples with high CLas titers (i.e., low Cq values observed for CLas 16S rRNA gene qPCR assays), average genome coverages of CLas populations in DNA samples generated using the PDF protocol were on average 35-fold greater (mean coverage of 43.2 and 1.24 for PDF and Direct protocols, respectively; $t_{\text{welch}} = -3.37$, $df = 31$, $p = 0.002$) than directly sequencing DNA extracted from an individual *D. citri* adult (“Direct protocol”, Fig. 1b). In addition to CLas, the PDF protocol significantly increased coverage of *Wolbachia* symbiont populations within *D. citri* adults by more than 7-fold compared to the Direct protocol ($t_{\text{welch}} = -11.8$, $df = 33$, $p = 2.7 \times 10^{-13}$) (Fig. 1b). While average coverage of *Wolbachia* symbiont populations approached 20X in the Direct protocol, they increased to 140X using the PDF protocol, greatly enhancing sequencing effort towards additional symbionts present within *D. citri* adults. Conversely, average coverage of the *D. citri* symbionts Car and Prof were significantly reduced by 6- and 3-fold, respectively, when comparing the PDF and Direct protocols, respectively. Nevertheless, average coverage of Car and Prof populations were 5X and 39X, respectively, using the PDF protocol which is sufficient for downstream genome assembly or comparative genomic analyses.

Aside from genome coverage estimation, we performed *de novo* genome assembly of individual and pooled samples (i.e., a coassembly) and examined the taxonomic identity of each contig by alignment to high quality reference genome sequences or NCBI’s nt database in concert with the BASTA tool for taxonomic annotation⁵². Assembly statistics of *de novo* genome assemblies for all *D. citri* symbionts analyzed in the present study can be found in

Supplemental Information (Supplementary Table 2). Of note, we assembled draft bacterial genomes from laboratory-reared *D. citri* adults assigned to the genus *Paracoccus* and *Brevundimonas* at 85 and 96% completeness, respectively, and with less than 5% contamination as assessed via the CheckM tool⁵⁵. Due to the uneven abundances of these taxa among *D. citri* individuals and their low coverage (Fig. 1b), their genomes were highly fragmented with low N50 values of 3,480 and 8,650 nt for *Paracoccus* and *Brevundimonas*, respectively, which are the sequence lengths at which sequences this long or longer compose at least 50% of the genome length. All assembled microbiome sequence data reported for laboratory-reared and field-collected *D. citri* adults (see below) can be accessed in the ‘Data availability’ section.

Next, we utilized the PDF protocol to sequence, assemble, and examine the microbiomes of field-collected *D. citri* adults (n = 6) collected from a citrus grove in Florida, USA in November, 2018. Unfortunately, CLas titers in these samples were quite low (CLas 16S rRNA gene Cq values > 30), and consequently any assembled CLas sequences detected in these samples were short in length and possessed low coverage (mean coverage $\pm$ s.d. = 0.74 ± 1.18). However, the coassembly of the six field-collected *D. citri* adults revealed large contigs assigned to the *D. citri* symbionts Car and Prof in each *D. citri* adult from the FL citrus grove (Supplementary Fig. 3a). Rather unexpectedly, we also detected 10,830 contigs (> 2,500 nt length) in the coassembly that were assigned to the *C. sinensis* genome, and in particular, large, high coverage contigs aligning to chloroplast and mitochondrial sequence elements of this same species (Supplementary Fig. 3b). Unlike our laboratory-reared specimens, in which we observed a significant sequence signal associated with the *D. citri* genome, only six contigs (>2,500 nt length) were assigned to the *D. citri* genome in field-collected adults, the longest of which was an ~15 kb contig affiliated with the *D. citri* mitochondrial genome (Supplementary Fig. 3a). We also detected large numbers of

low coverage contigs assigned to commensal bacterial populations including members of the genera, *Escherichia*, *Flavobacterium*, *Pseudomonas*, and *Xanthomonas*. Of note, only contigs assigned to the genus *Pseudomonas* achieved appreciably high coverage (~20X) in one of the citrus grove-collected *D. citri* adults we sequenced (panel SPK8; Supplementary Fig. 3a). Finally, we observed high coverage contigs, some of which were large (~15 – 30 kb), that were assigned to fungal mitochondrial sequences from well-known entomopathogenic fungal genera *Entomophthora*, *Capillidium*, *Neoconidiobolus* (Supplementary Fig. 3a). These sequences achieved appreciable coverage in only two of the six field-collected *D. citri* adults we sequenced (panels SPK1 and SPK4; Supplementary Fig. 3a), suggesting the PDF protocol may enable identification of *in situ* populations of entomopathogenic fungi potentially infecting *D. citri* in citrus groves.

In addition to shotgun sequence analysis and assembly of microbial genomes from lab-reared *D. citri* adults, we estimated replication rates for all four symbiont populations detected within individual *D. citri* sequenced using the PDF and Direct protocols. The genomic region predicted to contain the origin of replication (*oriC*) in bacteria can be distinguished by both GC skew (Fig. 2a, green line), a peak in average genome coverage (Fig. 2a, gray line), or conserved genetic elements associated with the *oriC*, usually *dnaA*, a gene involved in replication at the *oriC*, but whose genomic location differs in the alphaproteobacterial symbionts investigated herein and other loci are localized there instead (e.g., *hemE*, Fig. 2a)^{57,64}. Using the iRep tool, which performs regression analysis of ordered sequence read coverage over complete or draft microbial genome assemblies to estimate the index of replication (iRep), we were able to calculate an instantaneous measure of the average population replication rate⁵⁷ for the *D. citri* symbionts (Fig. 2b). The iRep values for the four symbionts were ordered Car > Prof > CLas > *Wolbachia*

(Fig. 2c), but they differed between the PDF and Direct protocols. These differences were largely attributed to differential coverage between symbiont populations when comparing both protocols (Fig. 2d). For example, low sequence coverage of *CLas* and *Wolbachia* populations tended to result in higher iRep estimates in the Direct protocol data, whereas symbionts Car and Prof possessed higher coverage and lower iRep estimates with data produced by the Direct protocol. When comparing samples with coverage >20X, iRep values for Car and Prof populations in the PDF protocol data (1.52 ± 0.03 and 1.53 ± 0.11 , respectively) approached an asymptote comparable to iRep estimates of Car and Prof populations in the Direct protocol data (1.67 ± 0.17 and 1.36 ± 0.09 , respectively), but were not identical (Fig 2c, d). In contrast, *Wolbachia* populations achieved sufficient coverage (~20X) for genome analysis in Direct protocol data, but their iRep estimates between the PDF and Direct protocol differed (1.33 ± 0.05 and 1.11 ± 0.02 , respectively). To include samples from the Direct protocol, we chose to present unfiltered iRep values since coverage of some taxa, especially for the Direct protocol samples, failed to satisfy cutoff criteria required by the iRep tool to report an iRep estimate⁵⁷. For example, Car populations did not satisfy the iRep tool's requirement that a minimum of 98% of genome windows possess 5X average coverage, even though the percentage of windows at this coverage were >93% in 39 out of 42 (93%) Car populations sequenced we sequenced. We also calculated iRep values for the handful of publicly available raw sequencing data available for *CLas* strains sequenced from *Citrus sinensis* and *Citrus maxima* DNA extracts, which ranged from 1.32 to 1.64 (Supplementary Table 3). However, average genome coverage of these *CLas* strains (see Methods for details) were between only 5 and 18X, which may inflate iRep estimates compared to *CLas* strains herein whose iRep estimates plateaued with increasing average genome coverage (Fig. 2d, PDF protocol panels). Overall, we observed that sequence data generated by the PDF

protocol provides ample sequence coverage required to estimate stable iRep values for CLas (and other *D. citri* symbionts), especially when CLas titers in DNA extracts are sufficiently high (CLas 16S rRNA gene copies $\geq 1 \times 10^5$) (Supplementary Fig. 2). All iRep estimates and associated metrics for symbionts in lab-reared *D. citri* and CLas genomes isolated from *Citrus* spp. can be found in Supplementary Table 3.

Our next aim was to investigate the relationship of and genetic variation between CLas strains sequenced with the PDF protocol and publicly available CLas strains. Phylogenetic analysis of putatively non-recombinogenic SNPs identified with Gubbins⁵⁹ (see Methods for details) revealed a clear monophyletic signal among CLas strains sequenced with the PDF protocol (Fig. 3). The CLas strains sequenced from lab-reared *D. citri* adults in this study were observed to be descended from CLas strains (psy62 and JRPAMB1) derived from *D. citri* colonies maintained at USDA facilities in FL⁵ (unpublished strain JRPAMB1 NCBI BioSample accession SAMN11842353; Fig. 3). In the 7 years since arriving in NY from FL rearing facilities, our CLas strains accumulated as many as 156 (78 ± 56 , mean $\pm$ s.d.) genetic polymorphisms compared to reference CLas strain psy62 from FL (Fig. 4). In contrast, CLas sequence data produced using the Direct protocol in this study did not reveal detectable (min SNP coverage = 10) genetic polymorphisms in CLas, even in samples with high CLas titers (Fig. 1b, right panel). Of note, the PDF protocol enabled detection of genetic variants in two CLas genes predicted to be involved in antibiotic resistance by the ProGenomes database⁶⁵, a putative cysteine desulfurase and N-acetyltransferase (Fig. 4, purple text). CLas strains 20E and 22E possessed a missense mutation in the cysteine desulfurase *sufS* gene, at position 1,066,382, resulting in a base change of A to G and an associated amino acid sequence change of glutamine to arginine (amino acid 37 of CLas strain psy62 protein accession WP_015824969.1). We also

detected a deletion mutation in the gene encoding a putative N-acetyltransferase (WP_102134467.1), in which a CA to C mutation at position 1,158,964 resulted in a frameshift mutation after amino acid position 36 in the encoded amino acid sequence. Interestingly, the gene (CLIBASIA_05345) encoding the putative N-acetyltransferase is located within a recently predicted prophage region of the *CLas* genome⁶⁶, and may constitute an important mobile element transferable among *CLas* strains if the prophage becomes lytic or is excised from the *CLas* genome⁶⁷.

In addition to potential antibiotic resistance, *CLas* encodes secreted protein effectors which may enable *CLas* to establish or persist during infection of its citrus or *D. citri* hosts^{34,68}. No variants were detected in loci encoding the secreted peroxidase SC2_gp095 (CLIBASIA_00485), the putative glutathione-dependent peroxidase SC2_gp100 (CLIBASIA_05595), or the Sec-translocon dependent effector SDE1 (CLIBASIA_05315). However, we did detect an insertion mutation in CLIBASIA_04530, encoding a putative Sec-dependent effector, which was shown to be predominantly expressed by *CLas* in citrus tissues³⁴ (Supplementary Table 4). While the exact functional consequences of the genetic polymorphisms detected between *CLas* strains in our lab and ancestral *CLas* strains in FL are unknown, the PDF protocol enabled robust detection and documentation of these mutations between *CLas* strains. Documentation of *CLas* genetic variants among laboratories should be a priority to ensure reproducible experimental outcomes among laboratories that maintain *CLas* strains via citrus grafting or acquisition and transmission of *CLas* by *D. citri*.

As previously noted, *CLas* strains also harbor distinct lytic and non-lytic prophage encoding regions which confer a variety of functions related to *CLas* pathogenicity and countering host immune responses^{33,66–68}. In the three predicted prophage elements FP1, FP2, and a recently

discovered Type IV prophage element ⁶⁶, we detected a total of 69, 11, and one polymorphisms in predicted coding regions, respectively, and only five polymorphisms in non-coding elements of the FP1 prophage region and none in the FP2 and the Type IV prophage element (Fig. 4). This was in contrast to non-prophage portions of the CLas genome, where only 33 polymorphisms were detected in coding regions compared to 52 in non-coding elements. The prophage variants occurred in phage primases, terminases, and four of the CLas strains we sequenced possessed an 8 bp insertion in the bacteriophage repressor protein c1, resulting in a frameshift mutation which could alter repression of the prophage lytic state within the strains this mutation was observed (Supplementary Table 4) ⁶⁹.

Aside from CLas, the Prof symbiont was observed to possess the next greatest number of genetic polymorphisms when compared to the latest Prof reference genome from the citrusgreening.org database when using either protocol (Fig. 4). The location of genetic polymorphisms detected in Prof using either protocol generally agreed, although coverage was usually higher with the Direct protocol in Car and Prof populations and only a few unique polymorphisms were detected with each protocol (Fig. 4). Notably, a majority of the variants detected were frameshift or non-synonymous polymorphisms within genetic machinery involved in polyketide synthesis (Supplementary Table 4), which could indicate alteration to the defensive compounds (i.e., diaphorin) provided to *D. citri* by the Prof symbiont ²⁷ and may be attributed to continuous laboratory propagation of *D. citri*. Genetic variants detected within Car, the nutritional symbiont of *D. citri* ^{27,28}, were localized to genes encoding hypothetical proteins, RNA polymerase subunit beta', DNA polymerase III subunit alpha, and the tRNA gene for phenylalanine (Supplementary Table 4). The genomes of *Wolbachia pipentis* strain wDi possessed the fewest number of polymorphisms (n = 6) compared to the *W. pipentis* strain wDi

reference genome, with highest coverage of polymorphisms detected in samples generated using the PDF protocol (Fig. 4). However, all variants detected were frameshift or missense mutations in key genes putatively involved cell wall remodeling, conjugation, and cell regulation (Supplementary Table 4). We did not detect genetic variants in *Wolbachia* loci encoding proteins with a demonstrated interaction with CLas⁷⁰. All genetic variants detected in *D. citri* symbionts can be accessed in Supplementary Table 4.

Discussion

Methods that enable or enhance sequencing of insect-associated microbiomes can inform solutions to longstanding challenges in diverse fields such as phytopathology, insect biology, and holobiont theory^{39,71}. For example, sequence-based analyses of microbiome composition and function can generate hypotheses related to the microbiome's influence on insect fitness or the number and types of ecosystem services an insect and its microbiome (i.e., the holobiont) can provide³⁹. Furthermore, sequenced-based investigations of genetic modules underlying microbially-mediated pesticide resistance may facilitate new strategies to circumvent pesticide resistance, and may uncover novel resistance phenotypes of use to beneficial insects^{39,72}. Yet the high quantities of host DNA detected within insect microbiomes can hamper detection sensitivity of both microbial diversity and genetic polymorphisms present therein¹². Further, many insect-associated microorganisms of great importance are obligate or facultative symbionts that are as yet uncultivable^{73,74}, and therefore cultivation-independent methods to predict functions of uncultivable taxa are vital research tools^{35,71}. Thus, new approaches are needed to circumvent the limitations inherent to sequence-based analyses of insect microbiomes. The PDF protocol described herein enables cost-effective, efficient, and direct access to insect-associated

microorganisms at orders of magnitude lower cost and total sequencing effort (Fig. 1a, Supplementary Table 1). Hence, the PDF protocol provides more microbiome data and facilitates investigations of microbe-microbe interactions or evolution at the molecular level.

Aside from the operational benefits afforded by the PDF protocol, it also enabled a comprehensive assessment of the microbiota present within and on *D. citri* adults. The PDF protocol not only enabled genetic analysis of both obligate (Car and Prof) and facultative (CLas and *Wolbachia*) symbionts, it also captured a variety of commensal bacterial and eukaryotic sequences in the *D. citri* adults examined (Fig. 1, Supplementary Fig. 3). The eukaryotic sequences we detected in field-collected *D. citri* are particularly intriguing as they represent genetic signatures of *Citrus* organelles retained due to *D. citri*'s phloem-feeding lifestyle. Thus, the shotgun sequence data provided by the PDF protocol may contain genetic signature's consistent with an insect's diet that could complement or supplant existing PCR-based approaches for gut content analysis in psyllids and other insects¹⁰. The latter is particularly promising since PCR amplicons have limited value in taxonomic assignment due to their short length, whereas citrus chloroplast and mitochondrial contigs detected in PDF protocol data are tens to hundreds of kilobases in length, which will greatly assist with taxonomic identification of citrus or other plant species fed on by *D. citri*. Of note, stark differences in the composition of genetic signatures detected between the laboratory-reared and field-collected *D. citri* were observed. For example, laboratory-reared *D. citri* possessed much higher quantities of the *D. citri* genome than field-collected *D. citri* (Fig. 1b; except for a high coverage mitochondrial contig in field-collected *D. citri*, Supplementary Fig. 3a), whereas the field-collected *D. citri* data mostly contained contigs assigned to the *Citrus sinensis* genome (Supplementary Fig. 3b). The exact reasons for these differences are unknown, but they may be due to methodological

differences in the capture and storage of laboratory-reared versus field-collected *D. citri*. For example, field-collected *D. citri* were preserved in 95% (v/v) ethanol when field-collected and then were subsequently frozen, whereas laboratory-reared *D. citri* were collected and frozen immediately. Thus, the cells of field-collected *D. citri* stored in ethanol prior to freezing may facilitate cell lysis, releasing more *D. citri* DNA that can be degraded during the DNase I treatment in the PDF protocol. The rigid wall of plant cells found in field-collected *D. citri* may be less affected by ethanol storage, and their cells were probably intact during DNase I treatment. This could explain why genetic signatures of citrus were higher in field-collected rather than laboratory-reared *D. citri*, if host cell lysis was less effective without ethanol storage or a freeze-thaw step. Nonetheless, these storage differences need to be explored in future iterations of the PDF protocol and may enhance digestion of *D. citri* or other insect DNA by DNase treatment.

In addition to genetic signatures indicative of the *D. citri* diet, we also detected mitochondrial signatures assigned to putatively entomopathogenic fungi, which suggests some members of the field-collected *D. citri* we studied were actively infected at the time of sampling (Supplementary Fig. 3). Interestingly, the mitochondrial signatures of the entomopathogenic fungi we detected are distinct from fungal genera previously reported to infect *D. citri*¹, and suggests additional, unexplored fungal taxa may be utilized as biocontrol agents for *D. citri* populations in the field. Fungi previously reported to infect *D. citri*, including members of the genera *Isaria*, *Hirsutellum*, *Lecanicillium*, and *Beauveria*, are all classified within the order Hypocreales, whereas the mitochondrial signatures we detected in the field-collected psyllids were all classified within the order Entomophthorales, members of which have been reported to infect *D. citri* in Mexico^{75,76} but whose potential as biocontrol agents have not been explored.

Despite the modest number of field-collected psyllids processed using the PDF protocol ($n = 6$), these abundant fungal signatures were also detectable, albeit at low coverage, within an additional 8 out of 17 (47%) *D. citri* adults that were collected from the same citrus grove, but were processed and sequenced using the Direct protocol (see Data availability to access raw data for the $n = 17$ *D. citri* individuals reported). These findings suggest an epizootic infection was occurring in the *D. citri* populations sampled from Spyke's citrus grove in Florida at the time of collection. Therefore, the PDF protocol can facilitate the identification of native or potentially novel fungal pathogens of invasive insects, such as *D. citri*, and may assist the development of biocontrol strategies to prevent their spread⁷⁷.

Aside from *D. citri* microbiome composition, the sequence data afforded by our PDF protocol also enabled robust estimation of the replication rate of each symbiont present within the *D. citri* microbiome (Fig. 2). The same analysis is unlikely to be compatible with alternative bacterial enrichment protocols, such as multiple displacement amplification or probe capture hybridization methods. This is due to biased genome amplification and the variable capture efficiency among probes or loss of relative abundance among captured sequence elements for each method, respectively. However, probe hybridization is capable of enriching and sequencing CLas DNA from citrus tissues¹⁶, which is currently unavailable with the PDF protocol described herein but is under active development. The latter two methods are the most widely used to sequence CLas genomes from *D. citri* adults or citrus samples, but both require greater time and financial investment and do not provide a comprehensive analysis of the *D. citri* microbiome. Nevertheless, the PDF protocol also has limitations, including reduced sequence coverage of obligate symbionts (Car and Prof) within *D. citri* compared to direct extraction and sequencing of DNA from *D. citri* adults using the Direct protocol. Despite this limitation, coverage of Car

and Prof populations within *D. citri* adults enabled robust iRep estimates on par with those calculated from the Direct protocol (Fig. 2d). The estimation of CLas replication rates in *D. citri* adults is of interest to citrus greening disease research since actively replicating CLas populations are more likely to be spread by their *D. citri* vector⁷⁸. However, whether CLas actively replicates or only accumulates within its insect vector via feeding has been a matter of debate^{24,79}. According to iRep values recorded in the present study, approximately 20 to 30% of the CLas population with >25X coverage in the *D. citri* adults examined were actively replicating (Fig. 2). These values were much lower than Car and Prof populations, which at equivalent coverage (25X or greater) suggest that 50 to 75% of these symbiont populations are actively replicating, consistent with the key nutritional and defensive roles of these bacteria within *D. citri*^{26–28}. Considering the high prevalence of the *Wolbachia* symbiont within *D. citri*^{80,81}, we were surprised by the low iRep values estimated for this symbiont which suggests only 10% of the *Wolbachia* populations examined were actively replicating (Fig. 2). It's possible that the unique mode of transovarial transmission by *Wolbachia*, competition with other symbionts, or inhibition by *D. citri*'s immune system might limit its replication within *D. citri*^{80,81}.

Despite its uncultivable nature, CLas is capable of propagation in laboratory settings using young citrus trees and laboratory-reared *D. citri*^{82–84}. While this indirect method of CLas cultivation has afforded valuable insights from therapeutics development to the response of *D. citri*'s immune system to CLas infection^{85–87}, continual propagation of bacteria in a laboratory can lead to bottleneck effects and introduce genetic variation not represented in founding or wild populations of the same bacteria^{88,89}. This fact is immediately apparent when comparing SNPs between reference CLas strain psy62 and the CLas strains sequenced in our laboratory (Fig. 3, 4). The great number of SNPs detected between CLas strain psy62 (a FL CLas strain) and our

CLas strains in NY are concerning for several reasons. First, since our CLas strains derive from CLas strains maintained in FL greenhouses and *D. citri* rearing facilities, the number of differences suggest that our CLas strains have significantly diverged from strains in FL (average of ~22 mutations/yr), which may result in inconsistent responses by therapeutics or other treatments being explored between facilities. To our knowledge, only a dozen CLas strains have been sequenced from commercial citrus groves or field-collected *D. citri* populations^{14,34,90–92}. Rather, genome sequences of most CLas strains are derived from laboratory-reared *D. citri* infected with CLas or infected citrus scions grafted onto healthy citrus rootstock (Supplementary Table 1). This suggests our understanding of the existing genomic diversity of CLas strains in commercial groves or urban settings is poor. Furthermore, our understanding of genetic variation within and among laboratory CLas strains is vital to ensure reproducible results between institutions studying citrus greening disease and developing novel therapeutics to prevent CLas spread. Research conducted using institute-specific CLas strains may obfuscate or stymie investigations of CLas physiology or therapeutics development without dedicated methods to monitor and account for genetic variation among lab-propagated CLas strains. Our results and others⁹³ also support the idea that future studies take into account a role for CLas strain genetic diversity and propensity for mutation in fundamental studies on plant infection and psyllid transmission⁷⁸. Although hypothetical and difficult in practice, the development of approaches to inoculate plants and insects using standardized stocks of a diversity of CLas strains is warranted, as well as their storage and retrieval as viable stocks from a dedicated facility to promote reproducibility of results and accessibility of strains among institutions.

The biodiversity of insects is vastly underexplored and we have only yet to scratch the surface in our understanding of the relationship between insects and their microbial partners⁹⁴.

Furthermore, there exists a paucity of methods to explore the host-microbe interface at the molecular level in insects and these methods are desperately needed for basic and applied entomological studies of insect vectors with medical and agricultural importance^{95,96}. We hope the PDF protocol and any future iterations thereof fulfills these needs and can enable advances in insect biology and microbiology to the benefit of all.

Acknowledgements

SAH is grateful to Dr. Ashley M. Frank for assistance with revising the manuscript. SAH would also like to thank Dr. Hannah J. MacLeod and Stephanie Preising for their helpful scientific discussions regarding the development of the PDF protocol. All authors are grateful to Dr. Stacy DeBlasio, Dr. David Igwe, Luke Thompson, and Julie Blaha for their tireless efforts maintaining and provisioning the citrus colonies and CLas-infected *D. citri* adults utilized in this work. This research used resources provided by the SCINet project of the USDA Agricultural Research Service, ARS project number 0500-00093-001-00-D. We gratefully acknowledge the financial support of the California Citrus Research Board grant # 5300-196, the USDA ARS Administrator's Class of 2021 Research Associate Program, a USDA NIFA predoctoral fellowship to MM and the USDA Agricultural Research Service CRIS Project 8062-22410-007-000-D for funding.

Competing interests

The authors declare no competing financial interests related to this research.

References

1. Grafton-Cardwell, E. E., Stelinski, L. L. & Stansly, P. A. Biology and management of asian citrus psyllid, vector of the huanglongbing pathogens. *Annu. Rev. Entomol.* **58**, 413–432 (2013).
2. Wang, N. & Trivedi, P. Citrus huanglongbing: A newly relevant disease presents unprecedented challenges. *Phytopathology* **103**, 652–665 (2013).
3. Ghosh, D. K., Motghare, M. & Gowda, S. Citrus greening: Overview of the most severe disease of citrus. *Adv. Agric. Res. Technol. J.* **2**, 83–100 (2018).
4. Marine, R. *et al.* Caught in the middle with multiple displacement amplification: The myth of pooling for avoiding multiple displacement amplification bias in a metagenome. *Microbiome* **2**, 1–8 (2014).
5. 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).
6. Tyler, H. L., Roesch, L. F. W., Gowda, S., Dawson, W. O. & Triplett, E. W. Confirmation of the sequence of ‘*Candidatus liberibacter asiaticus*’ and assessment of microbial diversity in huanglongbing-infected citrus phloem using a metagenomic approach. *Mol. Plant-Microbe Interact.* **22**, 1624–1634 (2009).
7. Burdon, J. J. & Thrall, P. H. Pathogen evolution across the agro-ecological interface: implications for disease management. *Evol. Appl.* **1**, 57–65 (2008).
8. Badial, A. B. *et al.* Nanopore sequencing as a surveillance tool for plant pathogens in plant and insect tissues. *Plant Dis.* **102**, 1648–1652 (2018).
9. Sundin, G. W. & Wang, N. Antibiotic resistance in plant-pathogenic bacteria. *Annu. Rev.*

- 643 *Phytopathol.* **56**, 161–180 (2018).
- 644 10. Cooper, W. R. *et al.* Host and non-host ‘whistle stops’ for psyllids: Molecular gut content
645 analysis by high-throughput sequencing reveals landscape-level movements of psylloidea
646 (Hemiptera). *Environ. Entomol.* **48**, 554–566 (2019).
- 647 11. Marotz, C. A. *et al.* Improving saliva shotgun metagenomics by chemical host DNA
648 depletion. *Microbiome* **6**, 1–9 (2018).
- 649 12. Pereira-Marques, J. *et al.* Impact of host DNA and sequencing depth on the taxonomic
650 resolution of whole metagenome sequencing for microbiome analysis. *Front. Microbiol.*
651 **10**, 1–9 (2019).
- 652 13. Zheng, Z., Deng, X. & Chen, J. Whole-genome sequence of “*Candidatus Liberibacter*
653 asiaticus” from Guangdong, China. *Genome Announc.* **2**, 1–2 (2014).
- 654 14. Liu, K. *et al.* Genome sequence resources of two ‘*Candidatus Liberibacter asiaticus*’
655 strains from Pakistan. *Plant Dis.* **104**, 2048–2050 (2020).
- 656 15. Wu, F. *et al.* Draft genome sequence of ‘*Candidatus Liberibacter asiaticus*’ from
657 *Diaphorina citri* in Guangdong, China. *Genome Announc.* **3**, 1–2 (2015).
- 658 16. Cai, W., Nunziata, S., Rascoe, J. & Stulberg, M. J. SureSelect targeted enrichment, a new
659 cost effective method for the whole genome sequencing of ‘*Candidatus Liberibacter*
660 asiaticus’. *Sci. Rep.* **9**, 1–8 (2019).
- 661 17. Lefoulon, E. *et al.* Large enriched fragment targeted sequencing (LEFT-SEQ) applied to
662 capture of *Wolbachia* genomes. *Sci. Rep.* **9**, 1–10 (2019).
- 663 18. Li, T., Zhang, L., Deng, Y., Deng, X. & Zheng, Z. Establishment of a *Cuscuta campestris*-
664 mediated enrichment system for genomic and transcriptomic analyses of ‘*Candidatus*
665 *Liberibacter asiaticus*’. *Microb. Biotechnol.* **14**, 737–751 (2021).

- 666 19. Duan, Y. *et al.* Complete genome sequence of citrus huanglongbing bacterium,
667 'Candidatus Liberibacter asiaticus' obtained through metagenomics. *Mol. Plant-Microbe*
668 *Interact.* **22**, 1011–1020 (2009).
- 669 20. Johnson, E. G., Wu, J., Bright, D. B. & Graham, J. H. Association of 'Candidatus
670 Liberibacter asiaticus' root infection, but not phloem plugging with root loss on
671 huanglongbing-affected trees prior to appearance of foliar symptoms. *Plant Pathol.* **63**,
672 290–298 (2014).
- 673 21. Li, S., Wu, F., Duan, Y., Singerman, A. & Guan, Z. Citrus greening: Management
674 strategies and their economic impact. *HortScience* **55**, 604–612 (2020).
- 675 22. Dala-Paula, B. M. *et al.* Effect of huanglongbing or greening disease on orange juice
676 quality, a review. *Front. Plant Sci.* **9**, 1–19 (2019).
- 677 23. Inoue, H. *et al.* Enhanced proliferation and efficient transmission of 'Candidatus
678 Liberibacter asiaticus' by adult *Diaphorina citri* after acquisition feeding in the nymphal
679 stage. *Ann. Appl. Biol.* **155**, 29–36 (2009).
- 680 24. Ammar, E. D., Ramos, J. E., Hall, D. G., Dawson, W. O. & Shatters, R. G. Acquisition,
681 replication and inoculation of 'Candidatus Liberibacter asiaticus' following various
682 acquisition periods on huanglongbing-infected citrus by nymphs and adults of the asian
683 citrus psyllid. *PLoS One* **11**, (2016).
- 684 25. Saha, S. *et al.* Survey of endosymbionts in the *Diaphorina citri* metagenome and assembly
685 of a *Wolbachia* wDi draft genome. *PLoS One* **7**, (2012).
- 686 26. Nakabachi, A. *et al.* Defensive bacteriome symbiont with a drastically reduced genome.
687 *Curr. Biol.* **23**, 1478–1484 (2013).
- 688 27. Nakabachi, A., Piel, J., Malenovsky, I. & Hirose, Y. Comparative genomics underlines

multiple roles of *Proffotella*, an obligate symbiont of psyllids: Providing toxins, vitamins, and carotenoids. *Genome Biol. Evol.* **12**, 1975–1987 (2020).

28. Nakabachi, A. *et al.* The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* **314**, 267 (2006).

29. Ukuda-Hosokawa, R. *et al.* Infection density dynamics of the citrus greening bacterium ‘*Candidatus liberibacter asiaticus*’ in field populations of the psyllid *Diaphorina citri* and its relevance to the efficiency of pathogen transmission to citrus plants. *Appl. Environ. Microbiol.* **81**, 3728–3736 (2015).

30. Wang, N. The citrus Huanglongbing crisis and potential solutions. *Mol. Plant* **12**, 607–609 (2019).

31. Ajene, I. J. *et al.* Habitat suitability and distribution potential of *Liberibacter* species (“*Candidatus Liberibacter asiaticus*” and “*Candidatus Liberibacter africanus*”) associated with citrus greening disease. *Divers. Distrib.* **26**, 575–588 (2020).

32. Haapalainen, M. Biology and epidemics of *Candidatus Liberibacter* species, psyllid-transmitted plant-pathogenic bacteria. *Ann. Appl. Biol.* **165**, 172–198 (2014).

33. Merfa, M. V. *et al.* Progress and obstacles in culturing ‘*Candidatus Liberibacter asiaticus*’, the bacterium associated with huanglongbing. *Phytopathology* **109**, 1092–1101 (2019).

34. Thapa, S. P. *et al.* Genome-wide analyses of *Liberibacter* species provides insights into evolution, phylogenetic relationships, and virulence factors. *Mol. Plant Pathol.* **21**, 716–731 (2020).

35. Coyle, J. F., Lorca, G. L. & Gonzalez, C. F. Understanding the physiology of *Liberibacter asiaticus*: An overview of the demonstrated molecular mechanisms. *J. Mol. Microbiol.*

- 712 *Biotechnol.* **28**, 116–127 (2019).
- 713 36. Katoh, H., Miyata, S. I., Inoue, H. & Iwanami, T. Unique features of a Japanese
714 ‘*Candidatus Liberibacter asiaticus*’ strain revealed by whole genome sequencing. *PLoS*
715 *One* **9**, (2014).
- 716 37. Lin, H. *et al.* The complete genome sequence of ‘*Candidatus Liberibacter solanacearum*’,
717 the bacterium associated with potato Zebra Chip disease. *PLoS One* **6**, (2011).
- 718 38. Blaustein, R. A., Lorca, G. L. & Teplitski, M. Challenges for managing *Candidatus*
719 *Liberibacter* spp. (Huanglongbing disease pathogen): Current control measures and future
720 directions. *Phytopathology* **108**, 424–435 (2018).
- 721 39. Simon, J. C., Marchesi, J. R., Mougél, C. & Selosse, M. A. Host-microbiota interactions:
722 From holobiont theory to analysis. *Microbiome* **7**, 1–5 (2019).
- 723 40. Kondo, T. *et al.* A checklist of natural enemies of *Diaphorina citri* Kuwayama
724 (Hemiptera: Liviidae) in the department of Valle del Cauca, Colombia and the world.
725 *Insecta mundi* **0457**, 1–14 (2015).
- 726 41. Igwe, D. O., Higgins, S. A. & Cilia, M. An excised leaf assay to measure acquisition of
727 “*Candidatus Liberibacter asiaticus*” by psyllids associated with citrus Huanglongbing
728 disease. *Phytopathology* (2021) doi:10.1094/PHYTO-03-21-0124-SC.
- 729 42. Leinonen, R., Sugawara, H. & Shumway, M. The sequence read archive. *Nucleic Acids*
730 *Res.* 1–3 (2010) doi:10.1093/nar/gkq1019.
- 731 43. Cai, W., Yan, Z., Rascoe, J. & Stulberg, M. J. Draft whole-genome sequence of
732 ‘*Candidatus Liberibacter asiaticus*’ strain TX1712 from citrus in Texas. *Genome*
733 *Announc.* **6**, 1–2 (2018).
- 734 44. Chen, Y., Li, T., Zheng, Z., Xu, M. & Deng, X. Draft whole-genome sequence of a

- 735 “*Candidatus Liberibacter asiaticus*” strain from Yunnan, China. *Microbiol. Resour.*
736 *Announc.* **8**, e01413-18 (2019).
- 737 45. Cai, W. *et al.* Genome resource for the huanglongbing causal agent ‘*Candidatus*
738 *liberibacter asiaticus*’ strain AHCA17 from citrus root tissue in California, USA. *Plant*
739 *Dis.* **104**, 627–629 (2020).
- 740 46. Li, T. *et al.* Genome sequence resource of ‘*Candidatus liberibacter asiaticus*’ from
741 Thailand. *Plant Dis.* **104**, 624–626 (2020).
- 742 47. Lo, C. & Chain, P. S. G. Rapid evaluation and quality control of next generation
743 sequencing data with FaQCs. *BMC Bioinformatics* **15**, 1–8 (2014).
- 744 48. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to
745 single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
- 746 49. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*
747 **9**, 357–359 (2012).
- 748 50. Li, H. *et al.* The sequence alignment/map format and samtools. *Bioinformatics* **25**, 2078–
749 2079 (2009).
- 750 51. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 1–9
751 (2009).
- 752 52. Kahlke, T. & Ralph, P. J. BASTA – Taxonomic classification of sequences and sequence
753 bins using last common ancestor estimations. *Methods Ecol. Evol.* **10**, 100–103 (2019).
- 754 53. Laetsch, D. R., Blaxter, M. L. & Leggett, R. M. BlobTools: Interrogation of genome
755 assemblies [version 1; peer review: 2 approved with reservations]. *F1000Research* **6**, 1–
756 18 (2020).
- 757 54. Kielbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame

genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).

55. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

56. Mikheenko, A., Saveliev, V. & Gurevich, A. MetaQUAST: Evaluation of metagenome assemblies. *Bioinformatics* **32**, 1088–1090 (2016).

57. Brown, C. T., Olm, M. R., Thomas, B. C. & Banfield, J. F. Measurement of bacterial replication rates in microbial communities. *Nat. Biotechnol.* **34**, 1256–1263 (2016).

58. Seeman, T. snippy: fast bacterial variant calling from NGS reads. (2015).

59. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).

60. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

61. Team, R. D. C. & R Core Team. R: A language and environment for statistical computing. (2018).

62. Chen, X. & Cui, Y. BioCircos: Interactive circular visualization of genomic data using ‘htmlwidgets’ and ‘BioCircos.js’. (2019).

63. Yu, G., Smith, D., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).

64. Ioannidis, P. *et al.* New criteria for selecting the origin of DNA replication in *Wolbachia* and closely related bacteria. *BMC Genomics* **8**, (2007).

- 781 65. Mende, D. R. *et al.* ProGenomes: A resource for consistent functional and taxonomic
782 annotations of prokaryotic genomes. *Nucleic Acids Res.* **45**, 529–534 (2017).
- 783 66. Dominguez-Mirazo, M., Jin, R. & Weitz, J. S. Functional and comparative genomic
784 analysis of integrated prophage-like sequences in “*Candidatus Liberibacter asiaticus*”.
785 *mSphere* **4**, 1–13 (2019).
- 786 67. Fleites, L. A., Jain, M., Zhang, S. & Gabriel, D. W. ‘*Candidatus Liberibacter asiaticus*’
787 prophage late genes may limit host range and culturability. *Appl. Environ. Microbiol.* **80**,
788 6023–6030 (2014).
- 789 68. Jain, M., Fleites, L. A. & Gabriel, D. W. Prophage-encoded peroxidase in ‘*Candidatus*
790 *Liberibacter asiaticus*’ is a secreted effector that suppresses plant defenses. *Mol. Plant-*
791 *Microbe Interact.* **28**, 1330–1337 (2015).
- 792 69. Miller, R. V. & Kokjohn, T. A. Cloning and characterization of the cI repressor of
793 *Pseudomonas aeruginosa* bacteriophage D3: A functional analog of phage lambda cI
794 protein. *J. Bacteriol.* **169**, 1847–1852 (1987).
- 795 70. Jain, M., Fleites, L. A. & Gabriel, D. W. A small *Wolbachia* protein directly represses
796 phage lytic cycle genes in ‘*Candidatus Liberibacter asiaticus*’ within psyllids. *mSphere* **2**,
797 1–12 (2017).
- 798 71. Guyomar, C. *et al.* Multi-scale characterization of symbiont diversity in the pea aphid
799 complex through metagenomic approaches. *Microbiome* **6**, 1–21 (2018).
- 800 72. Gressel, J. Microbiome facilitated pest resistance: potential problems and uses. *Pest*
801 *Manag. Sci.* **74**, 511–515 (2018).
- 802 73. Moran, N. A., McCutcheon, J. P. & Nakabachi, A. Genomics and evolution of heritable
803 bacterial symbionts. *Annu. Rev. Genet.* **42**, 165–190 (2008).

- 804 74. Hosokawa, T. *et al.* Obligate bacterial mutualists evolving from environmental bacteria in
805 natural insect populations. *Nat. Microbiol.* **1**, 1–7 (2016).
- 806 75. Berlanga-Padilla, A. M. *et al.* Entomopathogenic fungi associated to *Diaphorina citri*
807 (Hemiptera: Liviidae) in Colima, Mexico. *Rev. Mex. Biodivers.* **89**, 986–1001 (2018).
- 808 76. Guizar-Guzman, L. & Sanchez-Peña, S. R. Infection by *Entomophthora* sensu stricto
809 (Entomophthoromycota: Entomophthorales) in *Diaphorina citri* (Hemiptera: Liviidae) in
810 Veracruz, Mexico. *Florida Entomol.* **96**, 624–627 (2013).
- 811 77. Clifton, E. H., Castrillo, L. A., Gryganskyi, A. & Hajek, A. E. A pair of native fungal
812 pathogens drives decline of a new invasive herbivore. *Proc. Natl. Acad. Sci. U. S. A.* **116**,
813 9178–9180 (2019).
- 814 78. Ammar, E. D., Hall, D. G., Hosseinzadeh, S. & Heck, M. The quest for a non-vector
815 psyllid: Natural variation in acquisition and transmission of the huanglongbing pathogen
816 ‘*Candidatus Liberibacter asiaticus*’ by Asian citrus psyllid isofemale lines. *PLoS One* **13**,
817 1–23 (2018).
- 818 79. Pelz-Stelinski, K. S., Brlansky, R. H., Ebert, T. A. & Rogers, M. E. Transmission
819 parameters for ‘*Candidatus Liberibacter asiaticus*’ by Asian Citrus Psyllid (Hemiptera:
820 Psyllidae). *J. Econ. Entomol.* **103**, 1531–1541 (2010).
- 821 80. Ren, S. L. *et al.* Localization and dynamics of *Wolbachia* infection in Asian citrus psyllid
822 *Diaphorina citri*, the insect vector of the causal pathogens of Huanglongbing. *Microbiol.*
823 *Open* **7**, 1–11 (2018).
- 824 81. Kruse, A. *et al.* Combining’omics and microscopy to visualize interactions between the
825 Asian citrus psyllid vector and the Huanglongbing pathogen ‘*Candidatus Liberibacter*
826 *asiaticus*’ in the insect gut. *PLoS One* **12**, 1–28 (2017).

- 827 82. Skelley, L. H. & Hoy, M. A. A synchronous rearing method for the Asian citrus psyllid
828 and its parasitoids in quarantine. *Biol. Control* **29**, 14–23 (2004).
- 829 83. Hosseinzadeh, S. *et al.* Proteomic polyphenism in color morphotypes of *Diaphorina citri*,
830 insect vector of citrus greening disease. *J. Proteome Res.* **20**, 2851–2866 (2021).
- 831 84. Hilf, M. E. & Lewis, R. S. Transmission and propagation of ‘*Candidatus Liberibacter*
832 asiaticus’ by grafting with individual citrus leaves. *Phytopathology* **106**, 452–458 (2016).
- 833 85. Mann, M. *et al.* *Diaphorina citri* nymphs are resistant to morphological changes induced
834 by ‘*Candidatus Liberibacter asiaticus*’ in midgut epithelial cells. *Infect. Immun.* **86**,
835 e00889-17 (2018).
- 836 86. Kruse, A. *et al.* ‘*Candidatus Liberibacter asiaticus*’ minimally alters expression of
837 immunity and metabolism proteins in hemolymph of *Diaphorina citri*, the insect vector of
838 Huanglongbing. *J. Proteome Res.* **17**, 2995–3011 (2018).
- 839 87. Arp, A. P., Hunter, W. B. & Pelz-Stelinski, K. S. Annotation of the Asian citrus psyllid
840 genome reveals a reduced innate immune system. *Front. Physiol.* **7**, 1–18 (2016).
- 841 88. Elena, S. F. & Lenski, R. E. Microbial genetics: Evolution experiments with
842 microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* **4**, 457
843 (2003).
- 844 89. Marks, M. E. *et al.* The genetic basis of laboratory adaptation in *Caulobacter crescentus*.
845 *J. Bacteriol.* **192**, 3678–3688 (2010).
- 846 90. Zheng, Y., Guo, J., Deng, X. & Zheng, Z. Genome sequence resource of ‘*Candidatus*
847 *Liberibacter asiaticus*’ strain myan16 from myanmar. *Plant Dis.* **105**, 1162–1164 (2021).
- 848 91. Kunta, M. *et al.* Draft whole-genome sequence of ‘*Candidatus Liberibacter asiaticus*’
849 strain TX2351 isolated from Asian citrus psyllids in Texas, USA. *Genome Announc.* **5**,

e00170-17 (2017).

92. Dai, Z. *et al.* Prophage diversity of ‘*Candidatus Liberibacter asiaticus*’ strains in California. *Phytopathology* **109**, 551–559 (2019).

93. Pitino, M. *et al.* The phloem-sap feeding mealybug (*Ferrisia virgata*) carries ‘*Candidatus Liberibacter asiaticus*’ populations that do not cause disease in host plants. *PLoS One* **9**, 1–7 (2014).

94. Gurung, K., Wertheim, B. & Falcao Salles, J. The microbiome of pest insects: it is not just bacteria. *Entomol. Exp. Appl.* **167**, 156–170 (2019).

95. French, E., Kaplan, I., Iyer-Pascuzzi, A., Nakatsu, C. H. & Enders, L. Emerging strategies for precision microbiome management in diverse agroecosystems. *Nat. Plants* **7**, 256–267 (2021).

96. Chevrette, M. G. *et al.* The antimicrobial potential of *Streptomyces* from insect microbiomes. *Nat. Commun.* **10**, 1–11 (2019).

97. Luo, H. & Gao, F. DoriC 10.0: An updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic Acids Res.* **47**, 74–77 (2019).

98. Silva, P. A. *et al.* Genome sequence resource of ‘*Candidatus Liberibacter asiaticus*’ strain 9PA from Brazil. *Plant Dis.* **105**, 199–201 (2021).

Figures

Fig. 1 Comparison of CLas and additional microbial community member abundances between (a) the PDF protocol and existing CLas sequencing protocols in the literature or (b) between the PDF and Direct protocols using laboratory-reared *Diaphorina citri* adults with equivalent CLas titers. In the top panel, the PDF protocol (orange circles) generates equivalent or greater amounts of CLas data with orders of magnitude less sequencing data than existing protocols (green circles), reducing sequencing costs and increasing useable microbiome data. (b) In addition to increasing CLas abundance relative to the host, the PDF protocol also improves detection and

analysis of *Wolbachia pipientis* strain wDi and other commensal bacteria (*Paracoccus* sp. and *Brevundimonas* sp.). For *D. citri* individuals with comparable CLas titers (assessed via qPCR; text inside bottom panels), the relative abundance of CLas and *Wolbachia* in these samples is an order of magnitude greater with the PDF protocol compared to directly sequencing DNA extracted from an individual *D. citri* adult (“Direct protocol” in bottom right panel). The size and color of points in (b) represent the length and taxonomic affiliations of contigs generated from the coassembly of sequence reads from all lab-reared *D. citri* individuals processed in the present study. Hence, taxa with fewer, larger circles represent larger, more contiguous assemblies. Plots in (b) were inspired by BlobTools⁵³.

Fig. 2 Index of replication (iRep) estimates for the four *D. citri* symbionts detected in sequence data sets generated from the PDF and Direct protocols. (a) The mean coverage (5 kb length windows, 10 nt sliding window) and GC skew across the CLas genome provide information about the replication origin (red line) and terminus (blue line). (b) The iRep tool⁵⁷ filters and sorts the coverage by rank along the CLas genome to derive a metric consistent with an average instantaneous replication rate within a bacterial population. (c) The iRep estimates were highest for ‘*Candidatus Carsonella armatura*’ in both the PDF and Direct protocols followed by

‘*Candidatus Proffittella armatura*’, CLas, and *Wolbachia pipientis* strain wDi, but were shown to be highly dependent upon genome coverage (d). Squares and triangles in (c) and (d) indicate whether iRep could robustly calculate the filtered index of replication or not, respectively. Lines in (d) are second order polynomial regressions fit by formula $iRep \sim \text{poly}(\log(\text{coverage}), 2)$ in R 4.1.2. Unfiltered index or replication values are shown to directly compare estimates from the PDF and Direct protocols. *oriC* = predicted origin or replication determined by the DoriC database ⁹⁷ for CLas strain psy62; *hemE* = gene encoding uroporphyrinogen decarboxylase involved in heme biosynthesis and often detected adjacent to *oriC* in members of the *Alphaproteobacteria* ⁶⁴; *dnaA* = encodes DnaA protein that binds to *dnaA* boxes to promote strand separation at the *oriC* region. All iRep data displayed in panels (c) and (d) be found in Supplementary Table 3.

Fig. 3 A midpoint-rooted, whole genome SNP phylogeny of the CLas strains derived from laboratory-reared *Diaphorina citri* adults using the PDF protocol (black tip labels) or *Citrus* spp. and *D. citri* tissues from public databases (colored tip labels). Shapes next to the tree tips indicate the DNA tissue source and colors of the shapes and tip labels represent the country of origin of the CLas genome (except black tip labels, which are US samples from this study). Values at nodes represent ultrafast bootstrap support reported by IQ-TREE⁶⁰ estimated using 1,000 bootstrap replicates. The scale bar represents the number of nucleotide substitutions per site. Strains with genome assemblies possessing median contig sizes of < 30 kb for more than half the

942 contigs in the assembly were excluded, except for strain 9PA, which was included due to its
943 unique geographic context in Brazil ⁹⁸.

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

Fig. 4 Coverage profiles of genetic polymorphisms detected within genomes of bacterial pathogens and symbionts of *Diaphorina citri*, the Asian citrus psyllid. Genetic polymorphisms identified in sequence reads generated using the PDF protocol (red points) versus the Direct protocol (black points) are compared. Teal-colored regions in the top left panel are predicted prophage-encoding regions of the CLas genome⁶⁶. Regions and labels in purple in the top left are two genes predicted to encode antibiotic resistance genes and observed to possess genetic

970 polymorphisms compared to CLas strain psy62. Four letter codes inside each ring are gene
 971 abbreviations associated with annotated loci in which polymorphisms were detected. If a four
 972 letter abbreviation is absent, the polymorphism was either located within a gene lacking an
 973 annotation or was located in a noncoding region of the genome. All polymorphisms detected by
 974 snippy⁵⁸, including their four letter annotations, can be found in Supplementary Table 4.

a **b**

[illegible]