Article

Citrus genomic resources unravel putative genetic determinants of Huanglongbing pathogenicity

A. Chromosome
B. Gene density
C. Density of TE
D. Density of miRNA
E. Density of snRNA
F. Density of rRNA
G. Density of tRNA
H. Synteny between chromosomes

Yuxia Gao, Jin Xu,
Zhilong Li, ...,
Danelle Seymour,
Balian Zhong,
Nian Wang

zhongbalian@gnnu.edu.cn
(B.Z.)
nianwang@ufl.edu (N.W.)

Highlights

Generated phased diploid
genome assembly of
Citrus sinensis Newhall for
ASE analysis

Citrus pan-genome
contains 50,442 genes
including 13,301 core
genes

Sequenced 91 citrus
accessions and conducted
GWAS using 447 citrus
accessions

Identification of candidate
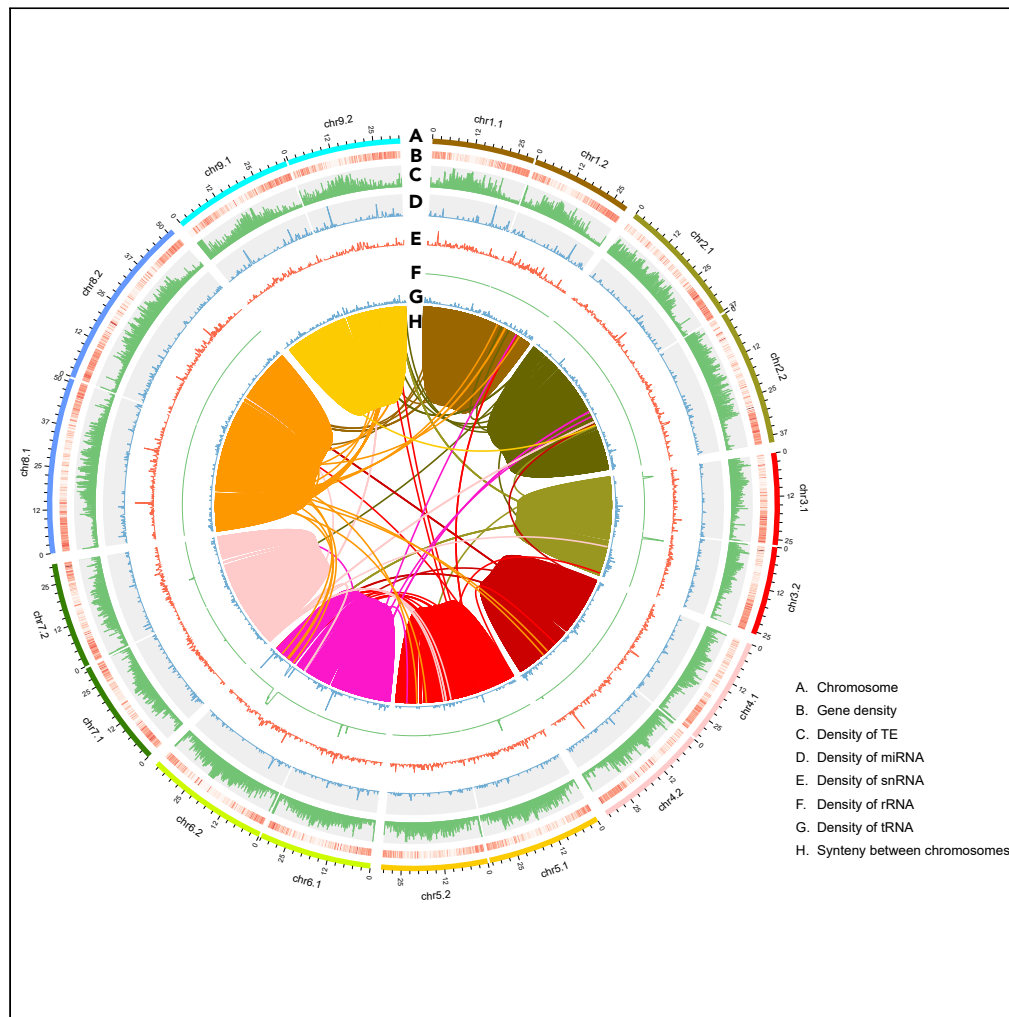genes for HLB resistance,
tolerance, or susceptibility

Article

Citrus genomic resources unravel putative genetic determinants of Huanglongbing pathogenicity

Yuxia Gao,[1,4] Jin Xu,[2,4] Zhilong Li,[1,4] Yunzeng Zhang,[2] Nadia Riera,[2] Zhiwei Xiong,[1] Zhigang Ouyang,[1] Xinjun Liu,[1] Zhanjun Lu,[1] Danelle Seymour,[3] Balian Zhong,[1,*] and Nian Wang[2,5,*]

SUMMARY

Citrus HLB caused by *Candidatus* Liberibacter asiaticus is a pathogen-triggered immune disease. Here, we identified putative genetic determinants of HLB pathogenicity by integrating citrus genomic resources to characterize the pan-genome of accessions that differ in their response to HLB. Genome-wide association mapping and analysis of allele-specific expression between susceptible, tolerant, and resistant accessions further refined candidates underlying the response to HLB. We first developed a phased diploid assembly of *Citrus sinensis* 'Newhall' genome and produced resequencing data for 91 citrus accessions that differ in their response to HLB. These data were combined with previous resequencing data from 356 accessions for genome-wide association mapping of the HLB response. Genes determinants for HLB pathogenicity were associated with host immune response, ROS production, and antioxidants. Overall, this study has provided a significant resource of citrus genomic data and identified candidate genes to be further explored to understand the genetic determinants of HLB pathogenicity.

INTRODUCTION

Citrus is one of top three fruit crops worldwide and is an important source of vitamin C in human diet.[1] However, citrus production faces many challenges including diseases, drought, flood, and freezes. Among them, citrus Huanglongbing (HLB, also known as greening) caused by *Candidatus* Liberibacter spp. presents an unprecedented challenge and has spread to most citrus growing regions such as China, Brazil, and USA.[2,3] *Ca.* L. asiaticus (CLas) is the most prevalent HLB pathogen. Almost all commercial citrus varieties are susceptible to HLB with few exceptions, such as Sugar Belle Mandarin (*Citrus reticulata*), Persian lime (*Citrus latifolia*), and US-897 (*C. reticulata* Blanco ✕ *Poncirus trifoliata* L. Raf.) that show tolerance against HLB.[4–7] In addition, multiple citrus relatives such as *Microcitrus australis*, *Eremocitrus glauca*, *Swinglea glutinosa*, *M. warburgiana*, and *M. papuana*, have shown resistance against HLB.[8,9]

HLB has been suggested to be a pathogen-triggered immune disease.[10] CLas is vectored mainly by Asian citrus psyllid (*Diaphorina citri*) and after transmission begins to colonize the phloem where it initiates a systemic and chronic immune response including reactive oxygen species (ROS) production, subsequent cell death of phloem tissues, and eventual HLB symptom development. Consistent with this model, CLas causes fewer changes in the expression of immunity genes in the tolerant rootstock variety US-897 than the susceptible variety 'Cleopatra' mandarin.[7] In addition, CLas stimulates significantly higher levels of ROS production in both the HLB susceptible Mexican lime and the HLB-tolerant Persian lime, with the latter demonstrating higher antioxidants, suggesting that high ROS levels are tolerated in Persian lime.[6] However, the genetic determinants responsible for resistance, tolerance, and susceptibility of different citrus genotypes against HLB remain unknown.

With HLB being a pathogen-triggered immune disease, it has been proposed that CLas recognition and downstream immune signaling, ROS production, and antioxidants may be responsible for variation in the response to HLB in citrus (resistance, tolerance, and susceptibility).[10] Antimicrobial peptides have also been associated with resistance to HLB.[11]

[1]National Navel Orange Engineering Research Center, Gannan Normal University, Ganzhou, Jiangxi, China

[2]Citrus Research and Education Center, Department of Microbiology and Cell Science, IFAS, University of Florida, Lake Alfred, FL, USA

[3]University of California, Riverside, CA, USA

[4]These authors contributed equally

[5]Lead contact

*Correspondence:
zhongbalian@gnnu.edu.cn (B.Z.),
nianwang@ufl.edu (N.W.)
https://doi.org/10.1016/j.isci.2023.106024

Bacterial pathogens trigger immune response via recognition by either membrane-localized pattern recognition receptors (PRRs) and intracellular nucleotide-binding domain leucine-rich repeat receptors (NLRs),[12,13] leading to PAMP-triggered immunity (PTI) and effector-triggered immunity (ETI), respectively. PTI and ETI share multiple downstream immune responses including the influx of Ca^{2+}, ROS burst, activation of mitogen-activated protein kinase (MAPK) cascades, defense gene induction, and biosynthesis of defense phytohormones.[12,13] Recent studies showed that both PTI and ETI are needed to mount a robust immune response, as they synergistically enhance each other.[14–20] In citrus there are a large number of genes involved in these two signaling pathways. For example, citrus immune signaling cascades include approximately 925 pattern recognition receptors (PRRs), 703 nucleotide-binding, leucine-rich domains (NLRs), 45 calcium-dependent protein kinases (CDPKs or CPKs), 100 mitogen-activated protein kinase (MAPKs), 119 cytoplasmic receptor-like kinases (CRLK), as well as 137 pathogenesis-related (PR) genes, 19 ROS genes, and 436 antioxidant enzyme genes.

We hypothesized that the genetic determinants of HLB pathogenicity, including candidates in these immune signaling pathways, can be identified through the integration of existing and newly produced citrus genomic resources to facilitate pan-genome analysis, a genome-wide association study (GWAS) for HLB response, and allelic-specific expression analyses of citrus accessions that differ in HLB resistance, tolerance and susceptibility. The genomes of multiple citrus genotypes and relatives have been assembled including *C. sinensis*,[21] *Atalantia buxifolia*, *Citrus medica*, *Citrus ichangensis*,[22] *Fortunella hindsii*,[23] *Citrus clementina*,[1] *C. reticulata*,[24] *Citrus grandis*, and *P. trifoliata*.[25] However, chromosome-level phased genomes were not available except *C. sinensis* cv. Valencia,[26] limiting exploration of haplotype-specific differences in gene content. In addition, thousands of accessions of citrus cultivars and relatives are available worldwide, representing a treasure to be mined.

In this study, we developed a haplotype-resolved genome assembly of *C. sinensis* 'Newhall', which was used for identification of genes that are specific to HLB resistant accessions. In total, pan-genome analysis was performed using genome assemblies from both susceptible and tolerant/resistant citrus cultivars. In addition, 26 citrus accessions (HLB resistant, tolerant, or susceptible) that have high quality genome sequencing data were mined for small indels linked to HLB response. The phased chromosome-level genome of *C. sinensis* was also used to investigate the difference in HLB susceptibility and tolerance of Valencia sweet orange and Sugar Belle mandarin LB9-9, respectively. We have also sequenced 91 citrus accessions, which, together with 356 previously sequenced citrus accessions, were used for GWAS analysis of genetic determinants responsible for HLB pathogenicity. Overall, this study has developed a significant resource of citrus genomic information and identified candidate genes to be further explored to understand the genetic determinants of HLB pathogenicity and to generate HLB resistant/tolerant citrus varieties.

RESULTS

Phased diploid genome assembly of *C. sinensis* 'Newhall'

We have selected *C. sinensis* 'Newhall' for sequencing because of its interesting features including maturation in the winter, whereas most other sweet orange cultivars mature before the summer, a second "twin" fruit opposite its stem (navel), and seedlessness. As a side note, it is worth mentioning that this project was started in 2019, which has suffered many hurdles caused by COVID19 similar to others. In total, 54 gigabases (Gb) Illumina paired-end short read data, 35 Gb PacBio HIFI read data, and 45 Gb Hi-C data were produced (Table S1). The genome of Newhall navel orange was *de novo* assembled using both HIFI and Illumina reads. The *de novo* assembly length was 685.27 Mb, including 1,624 contigs with N50 value of 12.5 Mb. Those contigs were further assembled into 1,013 scaffolds with N50 value of 32.8 M (Table S2) using Hi-C scaffolding methods. The final assembly contained 618 Mb in 18 chromosomes, which were assigned into primary or secondary haplotypes. Only 67 Mb sequences were unassigned to a scaffold (Figure 1, Tables S3 and S4). Syntenic blocks between homologous chromosomes indicated that the haplotypes were phased correctly (Figure S1)

The quality of assembly was supported by the *k*-mer analysis, which indicated that the haploid genome size was 350.15 Mb, approximately half of the assembled diploid genome size (Tables S2 and S5). In addition, BUSCO (95.1% of completeness), CEGMA (97.74% of completeness) assessments, and genomic sequencing coverage (99.97%) indicated high integrity of the phased diploid sweet orange genome assembly (Tables S6–S8). Synteny blocks were identified and revealed an overall synteny between the
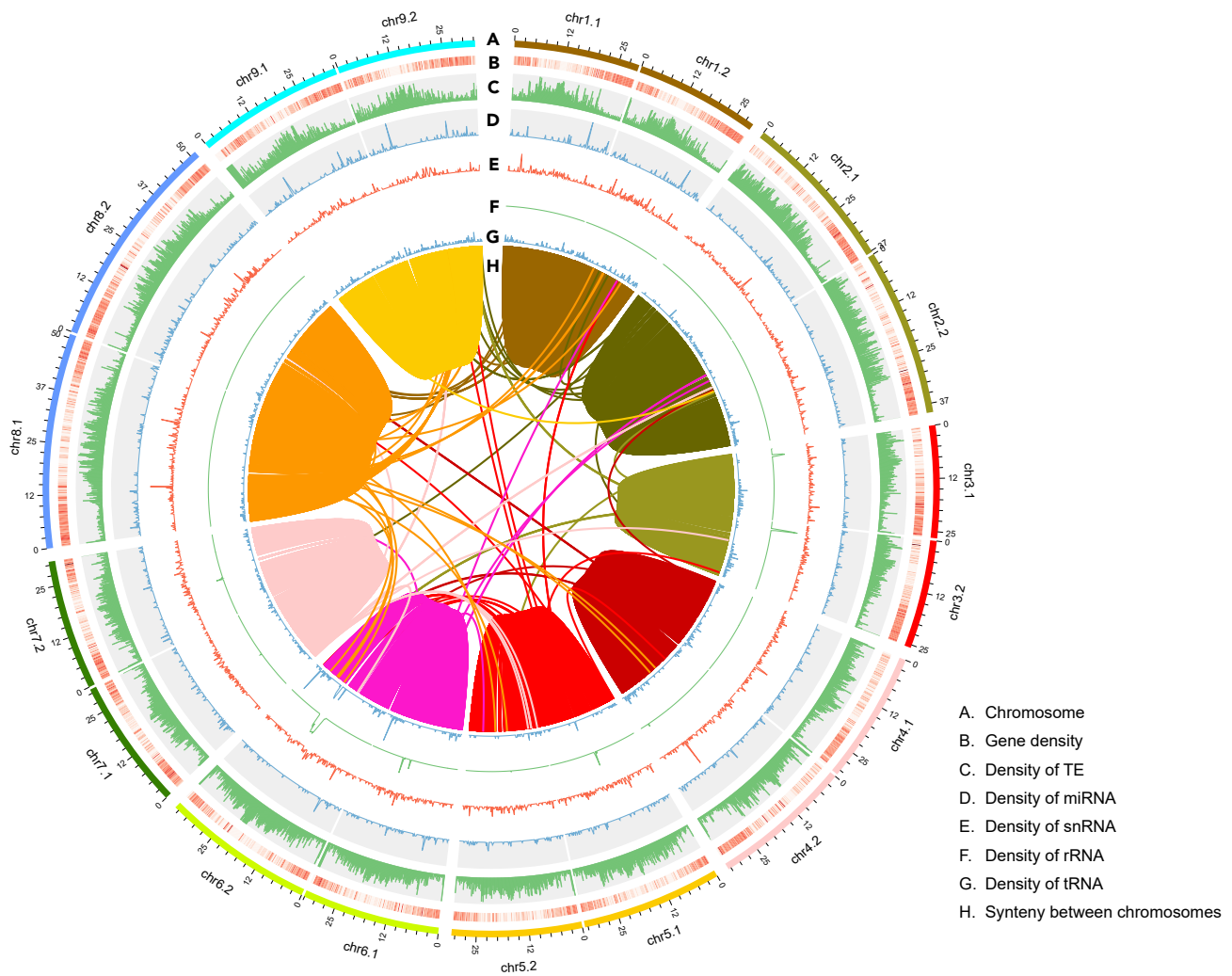
Figure 1. Genomic features of the *Citrus sinensis* Osbeck cv. Newhall

(A–G) The circle graph depicts the genomic characteristics, including chromosome length (A), gene density (B), density of transposable element (C), density of miRNA (D), density of snRNA (E), density of rRNA (F), density of tRNA (G), and the synteny between homologous chromosomes (H).

homoeologous chromosomes in Newhall navel orange (Figures 1, S1 and S2). Relatively low similarity values (37.41%–46.82%) were observed between the homoeologous chromosomes in Newhall navel orange (Table S9).

A total of 356 Mb (52%) of the assembled genome was masked and annotated as repeated sequences, of which 44.42% were long terminal repeat (LTR) retrotransposons, 1.74% were DNA transposable elements, 0.83% were long interspersed nuclear elements (LINE), and 0.06% were short interspersed nuclear elements (SINE) (Figure 1, Tables S2 and S10). To reduce false positives in gene prediction, we combined *de novo*, homology, and RNA-seq based approaches (Figure S3). Consequently, a total of 46,616 gene models were identified (Tables S2 and S11, Figure S3), among which 45,431 (97.46% of total predicted genes) were protein-coding genes (Table S12). The Newhall genome contained 22,916 genes on one chromosome set and 22,824 genes on the other chromosome set, and 876 genes unable to be assigned to either chromosome set. The annotated genes were classified into 26 COG categories including signal transduction, transcription, post-translational modification, protein turnover, and chaperones, carbohydrate transport and metabolism, translation, ribosomal structure, and biogenesis, intracellular trafficking, secretion, and vesicular transport, and secondary metabolites biosynthesis, transport, and catabolism (Table S13), with more than 42.6% of predicted genes as functionally unknown. In addition,
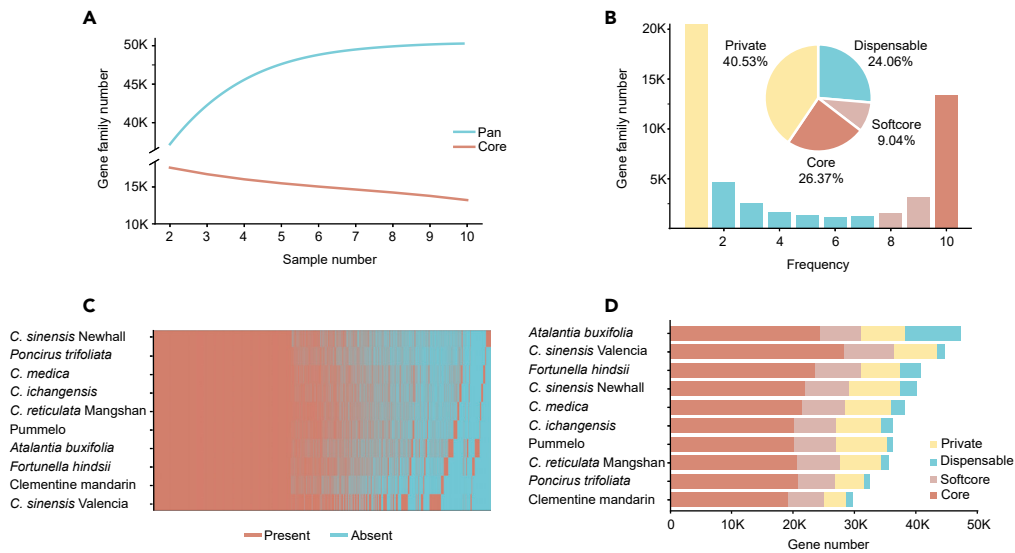
Figure 2. Pan genome of *Citrus* and relatives

(A) Rarefaction curve of detected genes in the pan and core genomes.

(B) The composition of the pan genome, including core, soft core, dispensable and private genes (pie plot). The histogram depicting the number of gene families of pan genome under different presence frequency in 26 accessions of *Citrus* and relatives.

(C) Presence and absence information of pan gene families in accessions of citrus and citrus relatives.

(D) Gene number of each composition for each individual genome. HKC: *Atalantia buxifolia*, CSV: 'Valencia' sweet orange, FOR: *Fortunella hindsii*, CSN: 'Newhall' sweet orange, XZ: *Citrus medica*, XJC: *Citrus ichangensis*, HWB: pummelo, CMS: *Citrus reticulata*'Mangshan', PON: *Poncirus trifoliata*, and CLM: Clementine mandarin.

11,221 non-coding RNA sequences were identified and annotated, including 931 micro RNAs (miRNAs), 1,200 transfer RNAs (tRNAs), 6,765 ribosomal RNA (rRNAs) and 2,325 small nuclear RNA (snRNA) (Figure 1, Table S14). We found most homologous gene pairs (13,257/15,423) have a *Ka/Ks* ratio <1, indicative of purifying selection (Figure S4). A total of 894 gene pairs with a *Ka/Ks* ratio >1 may have undergone positive selection, which included genes involved in metabolic and biosynthetic process, and response to stress and stimulus (Figure S5).

Pan-genome analyses of genes that are differentially present in HLB tolerant and susceptible genotypes

We explored the level of presence/absence sequence variation across the genus citrus and its relatives to dissect the genetic determinants of HLB pathogenicity. This was done using 9 previously assembled genomes of citrus and relatives and the newly sequenced Newhall genome. The 10 citrus accessions and relatives were classified into HLB- susceptible (6 citrus accessions) and-tolerant groups (4 citrus accessions) (Table S15). Here we have defined HLB-tolerant trees as those showing vigorous growth (such as thick canopy and not dying) in the presence of CLas and HLB symptoms, whereas HLB-susceptible trees refer to those without vigorous growth (such as with thin canopy and dying) in the presence of CLas and HLB symptoms based on the description in the original publications. Pan-genome analyses of the 10 genomes identified 50,442 gene families. The total number of gene sets continued to increase with the addition of each genome and was approaching but did not reach a plateau at n = 10 (Figure 2A), suggesting more high-quality genomes of citrus and its relatives are required. Specifically, 13,301, 4,559, 12,135, and 20,447 gene families were defined as core, soft-core, dispensable, and private genes, respectively (Figure 2B). The fraction of core genes (core plus soft core) in the citrus pan-genome (35%) was in line with previous studies (35–87%).[27–35] The dispensable and private gene families accounted for 64.6% of the total pan gene families, enlarging gene resources of citrus reference genome. The core gene families accounted for more than 51.7% genes of each genome (Figure 2C), suggesting conserved genomic features among citrus and relatives. In addition, wild citrus species, such as Ichang papeda, citron, and kumquat, and primitive citrus *Atlantia*, showed higher proportion of private genes than cultivated citrus accessions (Figure 2C). More than 47.2% of pan genome gene families were
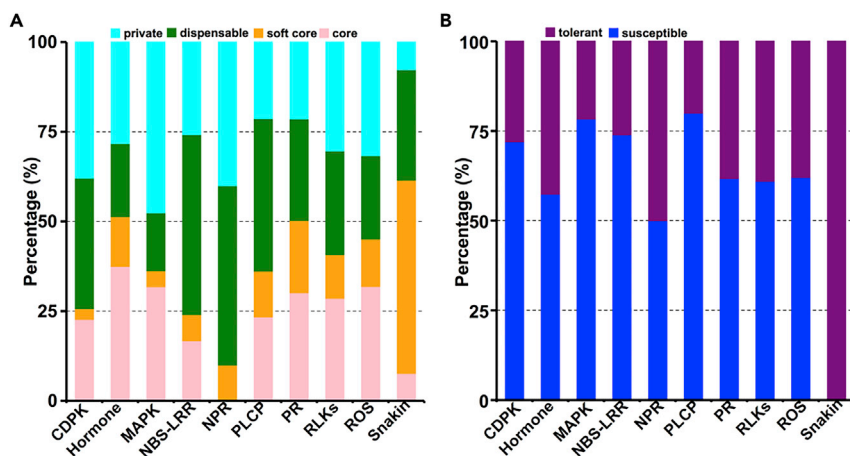
Figure 3. Plant immunity-related genes in citrus pan genome

(A) The composition of plant immunity-related genes in citrus pan genome, including core, soft core, dispensable and private genes.

(B) The HLB-tolerant (including resistant accessions) and-susceptible citrus accessions specific plant immunity-related genes. The plant immunity-related genes here refer to CDPK(Calcium-Dependent Protein Kinases), MAPK(Mitogen-Activated Protein Kinase Cascades), NBS-LRR (Nucleotide Binding Domain and Leucine-rich Repeat), NPR(Nonexpressor of Pathogenesis-related Genes), PLCP(Papain-Like Cysteine Proteases), PR (Pathogenesis-related Protein), RLKs (Receptor-Like Kinases), ROS(Reactive Oxygen Species), snakin, and plant hormone.

functionally unknown, which was higher than the diploid genome of *C. sinensis* 'Newhall' (42.3%) (Figure S6, Tables S13, and S16).

Whole-genome alignment between the *C. sinensis* Valencia genome and each of the 9 other genome assemblies was performed. A total of 77,609 InDels (37,355 deletions, 40,254 insertions) were identified in addition to 3 million single nucleotide polymorphisms (SNPs) (Figure S7, Table S17). We searched genes with presence/absence variation between the HLB tolerant and susceptible groups (Table S15). Because HLB is a pathogen-triggered immune disease,[10] we paid close attention to immunity related genes. The genes families involved in plant immunity, such as CDPK, MAPK, NBS-LRR, NPR, hormone, PLCP, PR, RLKs, ROS, and antioxidant genes, were presented mostly in dispensable and private gene sets (Figure 3A). Furthermore, HLB-susceptible citrus accessions were enriched for the plant immunity genes that were absent in HLB-tolerant accessions (Figure 3B, Table S18), suggesting a possible explanation for the stronger systemic and chronic immune responses observed in susceptible citrus accessions compared to those that are tolerant.

We further analyzed group (HLB-tolerant and-susceptible)-specific genes related to immunity (Table S15). Most of immunity related genes were present in all the accessions. Nevertheless, we have identified multiple interesting genes including orange1.1t03332.1 (NBS-LRR), orange1.1t04682.1 (NBS-LRR), orange1.1t05285.1 (PLCP, cysteine protease-like protein), Cs6g22310.1 (lectin), orange1.1t05183.1 (Leucine-rich repeat receptor-like protein kinase), Cs1g05340.1 (LRR-XII), Cs9g13810.1 (RLCK-XII/XIII), and Cs6g09910.5 (MAPKKK, Raf31), which were present in most HLB susceptible accessions (67–83%), but were absent in all HLB resistant accessions. On the contrary, only Cs2g10550.1 (Leucine-rich repeat receptor-like protein kinase), and Cs1g05370.1 (Serine-threonine protein kinase, plant-type) were present in 75% of four HLB tolerant accessions but absent in the 6 HLB susceptible accessions. In addition, 74 antioxidant biosynthesis and antioxidant enzyme genes were absent in susceptible citrus accessions but were present in one or more tolerant citrus accessions (Data S1).

Indel analyses of genes involved in plant immunity in 26 HLB-resistant,-tolerant, or-susceptible citrus accessions

The pan-genome analysis of 10 assembled citrus genomes included only accessions susceptible or tolerant to HLB, but none with HLB resistance (refers to citrus plants with no HLB symptoms in the presence of CLas or inhibiting CLas growth). To search for genomic signatures associated with resistance to HLB we expanded our interspecific analysis of genomic variation to a panel of 26 citrus accessions and relatives
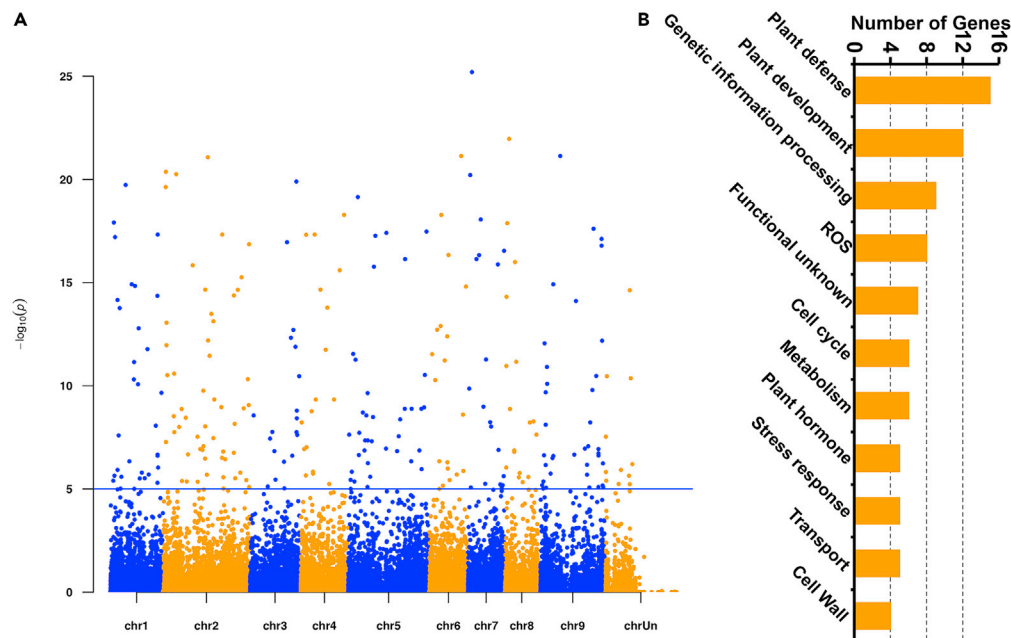
Figure 4. The genomic variations showing significant associations with citrus HLB based on GWAS
(A) Manhattan plots depicting HLB (A) showing significant associations with citrus genomic variations. Points over the blue line represent SNPs showing significant associations with HLB (corrected P-value < 1e-5).
(B) The functional pathway distribution of genes that contain the HLB associated SNPs.

with members exhibiting all three classes of response to HLB infection. The additional 16 accessions were selected based on their phylogenetic relationships and high-quality genome resequencing data (20–126 x coverages) (Table S15). In total, we have identified 263,177, 26,891, and 1369 indels in the resistant, tolerant, and susceptible groups, respectively. Specifically, we identified indels in the coding region of 26 NBS-LRR, 3 receptor-like kinase, and 1 SOD genes in the resistant group, 4 NBS-LRR and 1 SOD genes in the tolerant group, 1 NBS-LRR and 1 receptor-like kinase genes in the susceptible group (Table S19). Consistent with the model that HLB is a pathogen-triggered immune disease,[10] HLB resistant/tolerant citrus accessions contain more indel mutations in plant immunity genes than HLB susceptible accessions that might contribute to the reduced immune responses in tolerant/resistant citrus genotypes compared to the susceptible genotypes.[6,7]

GWAS analysis of citrus genes that are potentially responsible for the HLB resistance/tolerance or susceptibility

GWAS has been widely used to understand the genetic basis of plant disease resistance and susceptibility. To further explore the genetic basis of HLB resistance, we performed GWAS analysis using a large panel of both HLB susceptible and resistant/tolerant varieties (Figure 4, Table S20, Data S2) using whole genome sequences of 447 citrus accessions and relatives.[36] We have sequenced 91 additional citrus genotypes (Table S21) to complement public sequence datasets of 356 accessions. A total of 7.59 million SNPs for 447 citrus accessions and relatives were generated, which were subsequently filtered by quality and sequence depth, and SNPs with minor allele frequency (MAF) more than 0.01 and individual level missingness less than 10%. A total of 252,357 high quality SNPs across the whole genome were used for HLB GWAS analysis (Figure S8). The principal component analysis based on SNP data suggested 447 citrus accessions and relatives showed population stratification, which was accounted in GWAS analysis (Figure S9). The Quantile-Quantile Plot suggested the robustness of our GWAS analysis (Figure S10). We found HLB-associated citrus genomic SNPs were from a large number of genes across whole genome (Figure 4A), suggesting that the citrus genetic effect on HLB may be explained by the omnigenic model (a large number of genes). Such a model indicates that complex traits are influenced by core genes with direct effects as well as by a modest number of genes or pathways with small effects.[37] 252 SNPs, including 86 from coding region and 166 from non-coding region were significantly associated with HLB resistance/tolerance and
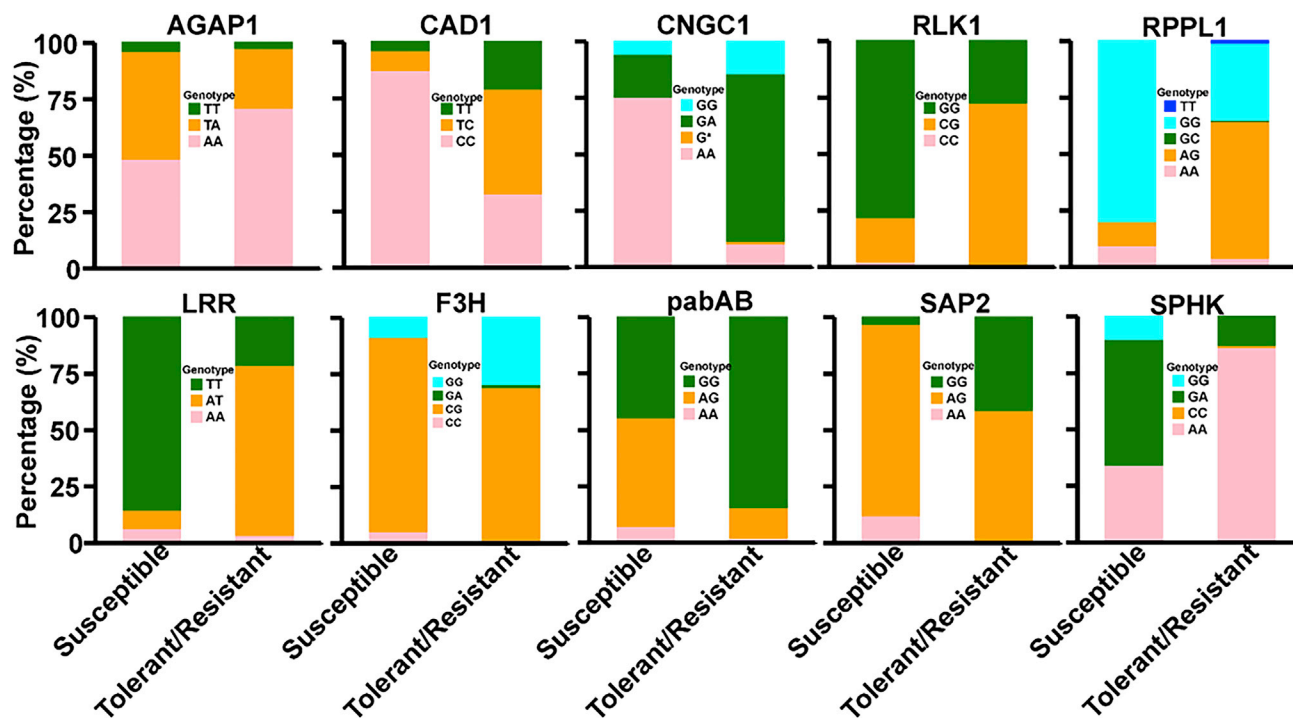
Figure 5. The type of citrus HLB disease associated SNPs among different citrus varieties
These SNPs were from genes involved in plant defense (AGAP1, LLR4, MIK2-LIKE, LRR2, RPPL1 and SUT1), and stress response (SEC11 and SPHK). AGAP1, ACYLATED GALACTOLIPID- ASSOCIATED PHOSPHOLIPASE 1; LLR4, MALECTIN-LIKE DOMAIN (MLD)- AND LEUCINE-RICH REPEAT (LRR)- CONTAINING PROTEIN 4; MIK2-LIKE, MDIS1-INTERACTING RECEPTORLIKE KINASE2 like; LRR2, Leucine-rich repeat protein 2; RPPL1, disease resistance RPP13-like protein 1; SUT1, SUPPRESSORS OF TOPP4-1; SEC11, signal peptidase I; SPHK, sphingosine kinase.

susceptibility (Figure 4A, Data S2, adjusted pvalue <1e-5). 37% of genes (32/86) containing SNPs with significant association with HLB resistance/tolerance and susceptibility were involved in plant immunity, ROS, and stress response (Figure 4B, Data S2). However, no group-specific SNPs were identified for HLB resistant/tolerant and susceptible groups even though SNPs from 10 genes showed different patterns between HLB susceptible and tolerant/resistant accessions (Figure 5, Data S2). These 10 genes included *PCS1* (*Phytochelatin synthase1*), mutation of which impairs callose deposition, bacterial pathogen defense and auxin content[38,39]; CNGC1 (Cyclic nucleotide-gated ion channel 1) that acts as Ca^{2+}-permeable channels involved in Ca^{2+} oscillations and receptor-mediated signaling during plant immunity[40]; two RLKs, *pabAB* (*para*-aminobenzoate synthetase) which has been shown to scavenge the reactive oxygen species *in vitro*[41]; and *RPPL1* (NLB RPP13-like protein 1).

Allele-specific expression (ASE)

Next, we used the phased diploid genome assembly of *C. sinensis* to investigate whether ASE contributes to the differences in HLB response between *C. sinensis* 'Valencia', an HLB susceptible cultivar, and Sugar Belle mandarin LB9-9, an HLB tolerant cultivar. Both cultivars mainly contain genes originated from either mandarin or pummelo. Mandarin LB9-9 is tolerant to HLB, whereas Valencia sweet orange is susceptible to HLB.[5] We conducted ASE analyses for two set of RNA-seq data of *C. sinensis* 'Valencia' and Sugar Belle mandarin LB9-9 including both HLB symptomatic and asymptomatic samples.[42] A total of 1623 and 1668 genes with ASE were identified for Valencia and LB9-9, respectively (Data S3). For Valencia, expression of 892 genes of the mandarin allele was higher than the allele with pummelo ancestry, with 731 displaying the reverse pattern with higher expression of the pummelo allele. For LB9-9, 1012 genes of mandarin origin had higher expression whereas 656 genes of pummelo origin had higher expression (Data S3). We further compared the ASE genes and differentially expressed genes (DEGs) between HLB symptomatic vs. asymptomatic Valencia or LB9-9. Intriguingly, 615 DEGs in symptomatic vs asymptomatic Valencia and 484 DEGs of LB9-9 were also ASE genes. Only 177 genes were shared between the DEG and ASE analysis including 166 genes showing similar response to HLB in both Valencia and LB9-9 and only 11 genes showed opposite

expression pattern to HLB (Data S3). The 11 genes included orange1.1t03406 (RLK), Cs9g12460 (raffinose synthase), orange1.1t03953 (haloacid dehalogenase-like hydrolase domain-containing protein), Cs5g18710 (licodione synthase), Cs5g21200 (indole-3-acetate beta-glucosyltransferase), Cs4g03330 (mandrin 4-coumarate—CoA ligase 1), Cs2g19490 (leucine-rich repeat, cysteine-containing type RLK), Cs2g19440 (Glucose-6-phosphate dehydrogenase), Cs2g04330 (LRR-VIII-2 RLK), Cs2g01740 (peptidase aspartic), and Cs1g12660 (caffeoyl-CoA O-methyltransferase). Nine genes, including orange1.1t03406, orange1.1t03953, Cs5g18710, Cs5g21200, Cs4g03330, Cs2g19490, Cs2g19440, Cs2g01740, and Cs1g12660 were upregulated in sweet orange and downregulated in LB9-9 under CLas infection condition. Two genes, including Cs9g12460 and Cs2g04330 were downregulated in sweet orange and upregulated in LB9-9 under CLas infection condition (Data S3).

DISCUSSION

We have completed the phased diploid genome assembly of *C. sinensis* 'Newhall'. There are currently nine assembled genomes for citrus and relatives.[1,21–25] Even though citrus and its relatives are diploid, none of the nine assembled genomes were chromosome-level phased genomes. A phased genome assembly with allelic information is critical for dissecting the special genetic characteristics, accurately evaluating somatic mutation calling and gene expression, as well as conducting allelic level analysis.[43] The first chromosome-level phased genome of *C. sinensis* was completed for Valencia sweet orange.[43] 'Newhall' navel is also sweet orange but with distinct characteristics from Valencia sweet orange, including the harvest time, navel, seeds, and juice contents. Completion of the chromosome-level phased genome for both Newhall and Valencia will provide useful resource to investigate the underlying genetic determinants for such traits. In addition, Newhall chromosome-level phased genome was successfully used in this study to identify multiple candidate genes contributing to the difference in HLB tolerance between mandarin LB9-9 and Valencia.

We have sequenced 91 citrus accessions in this study. In total, 447 citrus accessions and relatives have been sequenced so far. Importantly, the power of GWAS is boosted when sample size is large.[44] Imai and colleagues have demonstrated GWAS is suitable for citrus analysis based on analysis of 110 citrus accessions composed of landraces, modern cultivars, and in-house breeding lines.[45] The large amount of whole genome sequencing data of citrus and relatives enabled GWAS analysis of putative genetic determinants of HLB resistance/tolerance and susceptibility. Mattia et al. conducted GWAS analysis of genes responsible for flavonoid biosynthesis of diverse mandarin accessions.[46] Another study by Minamikawa et al. used GWAS analysis to identify putative genetic traits controlling fruit quality.[47] In both studies, SNP arrays were used. With the advances of high-throughput sequencing, whole genome sequencing has become a viable genotyping technology for use in GWAS analyses, offering the potential to analyze a broader range of genome-wide variations.[36] Whole genome sequence provides a significant advantage over array-based methods, with the potential to detect and genotype all variants present in a sample, not only those present on an array or imputation reference panel.[36] However, few GWAS analyses have been performed for citrus using whole genome sequencing data to date, probably owing to the challenges in handling a large amount of genomic data. This GWAS analysis using whole genome sequencing data in this study has advanced our understanding of HLB resistance, tolerance or susceptibility in citrus accessions.

Citrus genomic resources generated in the past and in this study enabled identification of candidate genes underlying the HLB resistance, tolerance, or susceptibility via Pan genome analysis of genes presence and absence in 10 citrus accessions, indel analyses of 26 citrus accessions, GWAS analysis of 447 citrus accessions, and ASE analysis of Valencia and mandarin LB9-9. Among the identified genes, NBS-LRR genes are the most abundant. This probably evolves from that CLas resides inside the phloem sieve elements[2] and NBS-LRRs are involved in direct recognition of CLas proteins either on the surface or released.[48–50] More unique NLRs genes were found in HLB-susceptible citrus accessions than in HLB-tolerant accessions, suggesting CLas might trigger more severe immune responses in HLB susceptible citrus genotypes. More indel mutations were identified in NLR genes in the HLB resistant group than in the susceptible group, which suggest that such mutations might lead to reduced immune responses to CLas in the HLB resistant group. In addition to NBS-LRR, genes encoding CDPK, PRRs, and RLCKs were also identified. CDPKs play critical roles in plant immunity, including regulation of oxidative burst, gene expression, and hormone signal transduction.[51] PRRs including receptor-like kinases (RLKs) and receptor-like proteins (RLPs) usually localize on the membrane to detect MAMPs in the apoplast. It is possible that some PAMPs of CLas, such as

LPS and flagella,[52,53] are released into the apoplast to be sensed by citrus cells. RLCKs including BOTRYTIS-INDUCED KINASE 1 (BIK1) and related PBS1-like kinases act synergistically with multiple PRRs to allow subsequent phosphorylation of the transmembrane NADPH-oxidase RESPIRATORY BURST OXIDASE D (RBOHD), the main producer of ROS during pathogen infection.[54–56] Consistent with our observation, it was reported that CLas causes less expression changes in immunity genes for the HLB-tolerant US-897 than the HLB-susceptible 'Cleopatra' mandarin,[7] which might result from the differences in the NLR, RLK, RLCK, and CDPK genes. Intriguingly, 74 antioxidant biosynthesis or antioxidant enzyme genes were absent in susceptible citrus accessions but were present in one or more tolerant citrus accessions. This is consistent with that higher antioxidant levels and antioxidant enzyme activities were reported to account for the higher tolerance of Persian triploid lime than Mexican lime.[6] Because the presence/absence pattern of antioxidant biosynthesis or antioxidant enzyme genes was not universal, they are likely responsible for some citrus genotypes, but not all genotypes. For example, HLB-tolerant Sugar Belle mandarin LB9-9 has similar antioxidant enzyme activities as HLB-susceptible Valencia sweet orange but has higher phloem regeneration capacity than the later. Antimicrobial peptides were also reported to be responsible for HLB resistance in Australian Finger lime.[11]

Overall, this study has provided a phased chromosome-level genome assembly for *C. sinensis* 'Newhall', and sequenced 91 new citrus accession, which were used for identification of putative genes responsible for HLB resistance, tolerance, or susceptibility. Such genes should be further verified using other approaches such as mutagenesis or gene silencing. Identification of genetic determinants responsible for HLB resistance, tolerance, or susceptibility in citrus accessions provide useful targets for developing of HLB-resistant/tolerant citrus cultivars using the CRISPR genome editing tool.[57,58]

Limitations of the study

By analyzing HLB susceptible, tolerant, or resistant citrus genotypes, we have identified putative genetic determinants of HLB pathogenicity. However, none of them have been verified using experimental approaches such as RNAi or CRISPR technologies.[59,60] We have conducted pan-genome analyses using 10 genomes. However, the pan-genome of Citrus did not reach a plateau, which requires more high-quality genomes of citrus and its relatives.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Plant materials
  - DNA and RNA extraction
  - Library construction and sequencing
  - *De novo* genome assembly for Newhall navel sweet orange
  - Genome quality assessment
  - Repeat element identification
  - Gene structure annotation
  - Gene functional annotation
  - Non-coding RNA annotation
  - Identification of genomic variations
  - Pan genome construction
  - GWAS
  - Allelic specific expression of transcriptomes of valencia sweet orange and Mandarin LB9-9 in response to CLas infection
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis

AUTHOR CONTRIBUTIONS

N.W. conceived and supervised the project. Y.G., J.X., and N.W. designed the experiment. Y.G., Y.Z., N.R., and Z.L. collected samples and extracted DNA. Y.G., J.X., Z.L., and N.W. analyzed the data. Y.G., J.X. and N.W. wrote the manuscript. N.W., Y. G., J.X., Z.L., Z.X., Z.O., X.L., Z.L., D.S., B.Z., and N.W. revised the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

REFERENCES

1. Wu, G.A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., Perrier, X., Ruiz, M., Scalabrin, S., Terol, J., et al. (2014). Sequencing of diverse Mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat. Biotechnol. *32*, 656–662. https://doi.org/10.1038/nbt.2906.

2. Bové, J.M. (2006). Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. J. Plant Path. 7–37.

3. Pandey, S.S., Hendrich, C., Andrade, M.O., and Wang, N. (2022). Liberibacter: from movement, host responses, to symptom development of citrus huanglongbing. Phytopathology *112*, 55–68. https://doi.org/10.1094/PHYTO-08-21-0354-FI.

4. Folimonova, S.Y., Robertson, C.J., Garnsey, S.M., Gowda, S., and Dawson, W.O. (2009). Examination of the responses of different genotypes of citrus to huanglongbing (citrus greening) under different conditions. Phytopathology *99*, 1346–1354. https://doi.org/10.1094/PHYTO-99-12-1346.

5. Deng, H., Achor, D., Exteberria, E., Yu, Q., Du, D., Stanton, D., Liang, G., and Gmitter, F.G. (2019). Phloem regeneration is a mechanism for huanglongbing-tolerance of "bearss" lemon and "LB8-9" sugar Belle.

Front. Plant Sci. *10*, 277. https://doi.org/10.3389/fpls.2019.00277.

6. Sivager, G., Calvez, L., Bruyere, S., Boisne-Noc, R., Brat, P., Gros, O., Ollitrault, P., and Morillon, R. (2021). Specific physiological and anatomical traits associated with polyploidy and better detoxification processes contribute to improved huanglongbing tolerance of the Persian lime compared with the Mexican lime. Front. Plant Sci. *12*, 685679. https://doi.org/10.3389/fpls.2021.685679.

7. Albrecht, U., and Bowman, K.D. (2012). Transcriptional response of susceptible and tolerant citrus to infection with Candidatus Liberibacter asiaticus. Plant Sci. *185–186*, 118–130. https://doi.org/10.1016/j.plantsci.2011.09.008.

8. Alves, M.N., Lopes, S.A., Raiol-Junior, L.L., Wulff, N.A., Girardi, E.A., Ollitrault, P., and Peña, L. (2020). Resistance to 'Candidatus liberibacter asiaticus,' the huanglongbing associated bacterium sexually and/or graft-compatible citrus relatives. Front. Plant Sci. *11*, 617664. https://doi.org/10.3389/fpls.2020.617664.

9. Cifuentes-Arenas, J.C., Beattie, G.A.C., Peña, L., and Lopes, S.A. (2019). Murraya paniculata and Swinglea glutinosa as short-term transient hosts of 'Candidatus

liberibacter asiaticus' and implications for the spread of huanglongbing. Phytopathology *109*, 2064–2073. https://doi.org/10.1094/phyto-06-19-0216-r.

10. Ma, W., Pang, Z., Huang, X., Xu, J., Pandey, S.S., Li, J., Achor, D.S., Vasconcelos, F.N.C., Hendrich, C., Huang, Y., et al. (2022). Citrus Huanglongbing is a pathogen-triggered immune disease that can be mitigated with antioxidants and gibberellin. Nat. Commun. *13*, 529. https://doi.org/10.1038/s41467-022-28189-9.

11. Huang, C.Y., Araujo, K., Sánchez, J.N., Kund, G., Trumble, J., Roper, C., Godfrey, K.E., and Jin, H. (2021). A stable antimicrobial peptide with dual functions of treating and preventing citrus Huanglongbing. Proc. Natl. Acad. Sci. USA *118*. e2019628118. https://doi.org/10.1073/pnas.2019628118.

12. Lu, Y., and Tsuda, K. (2021). Intimate association of PRR- and NLR-mediated signaling in plant immunity. Mol. Plant Microbe Interact. *34*, 3–14. https://doi.org/10.1094/mpmi-08-20-0239-ia.

13. Ge, D., Yeo, I.C., and Shan, L. (2022). Knowing me, knowing you: self and non-self recognition in plant immunity. Essays Biochem. *66*, 447–458. https://doi.org/10.1042/EBC20210095.

14. Tena, G. (2021). PTI and ETI are one. Nat. Plants 7, 1527. https://doi.org/10.1038/s41477-021-01057-y.

15. Ngou, B.P.M., Ahn, H.K., Ding, P., and Jones, J.D.G. (2021). Mutual potentiation of plant immunity by cell-surface and intracellular receptors. Nature 592, 110–115. https://doi.org/10.1038/s41586-021-03315-7.

16. Yuan, M., Jiang, Z., Bi, G., Nomura, K., Liu, M., Wang, Y., Cai, B., Zhou, J.M., He, S.Y., and Xin, X.F. (2021). Pattern-recognition receptors are required for NLR-mediated plant immunity. Nature 592, 105–109. https://doi.org/10.1038/s41586-021-03316-6.

17. Zhai, K., Liang, D., Li, H., Jiao, F., Yan, B., Liu, J., Lei, Z., Huang, L., Gong, X., Wang, X., et al. (2022). NLRs guard metabolism to coordinate pattern- and effector-triggered immunity. Nature 601, 245–251. https://doi.org/10.1038/s41586-021-04219-2.

18. Pruitt, R.N., Locci, F., Wanke, F., Zhang, L., Saile, S.C., Joe, A., Karelina, D., Hua, C., Fröhlich, K., Wan, W.-L., et al. (2021). The EDS1–PAD4–ADR1 node mediates Arabidopsis pattern-triggered immunity. Nature 598, 495–499. https://doi.org/10.1038/s41586-021-03829-0.

19. Tian, H., Wu, Z., Chen, S., Ao, K., Huang, W., Yaghmaiean, H., Sun, T., Xu, F., Zhang, Y., Wang, S., et al. (2021). Activation of TIR signalling boosts pattern-triggered immunity. Nature 598, 500–503. https://doi.org/10.1038/s41586-021-03987-1.

20. Chang, M., Chen, H., Liu, F., and Fu, Z.Q. (2022). PTI and ETI: convergent pathways with diverse elicitors. Trends Plant Sci. 27, 113–115. https://doi.org/10.1016/j.tplants.2021.11.013.

21. Xu, Q., Chen, L.L., Ruan, X., Chen, D., Zhu, A., Chen, C., Bertrand, D., Jiao, W.B., Hao, B.H., Lyon, M.P., et al. (2013). The draft genome of sweet orange (Citrus sinensis). Nat. Genet. 45, 59–66. https://doi.org/10.1038/ng.2472.

22. Wang, X., Xu, Y., Zhang, S., Cao, L., Huang, Y., Cheng, J., Wu, G., Tian, S., Chen, C., Liu, Y., et al. (2017). Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. Nat. Genet. 49, 765–772. https://doi.org/10.1038/ng.3839.

23. Zhu, C., Zheng, X., Huang, Y., Ye, J., Chen, P., Zhang, C., Zhao, F., Xie, Z., Zhang, S., Wang, N., et al. (2019). Genome sequencing and CRISPR/Cas9 gene editing of an early flowering Mini-Citrus (Fortunella hindsii). Plant Biotechnol. J. 17, 2199–2210. https://doi.org/10.1111/pbi.13132.

24. Wang, L., He, F., Huang, Y., He, J., Yang, S., Zeng, J., Deng, C., Jiang, X., Fang, Y., Wen, S., et al. (2018). Genome of wild Mandarin and domestication history of Mandarin. Mol. Plant 11, 1024–1037. https://doi.org/10.1016/j.molp.2018.06.001.

25. Peng, Z., Bredeson, J.V., Wu, G.A., Shu, S., Rawat, N., Du, D., Parajuli, S., Yu, Q., You, Q., Rokhsar, D.S., et al. (2020). A chromosome-scale reference genome of trifoliate orange (Poncirus trifoliata) provides insights into disease resistance, cold tolerance and genome evolution in Citrus. Plant J. 104, 1215–1232. https://doi.org/10.1111/tpj.14993.

26. Wu, B., Yu, Q., Deng, Z., Duan, Y., Luo, F., and Gmitter, F., Jr. (2023). A chromosome-level phased genome enabling allele-level studies in sweet orange: a case study on citrus Huanglongbing tolerance. Hortic. Res. 10, uhac247. https://doi.org/10.1093/hr/uhac247.

27. Sun, X., Jiao, C., Schwaninger, H., Chao, C.T., Ma, Y., Duan, N., Khan, A., Ban, S., Xu, K., Cheng, L., et al. (2020). Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. Nat. Genet. 52, 1423–1432. https://doi.org/10.1038/s41588-020-00723-9.

28. Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D.M., Burzynski-Chang, E.A., Fish, T.L., Stromberg, K.A., Sacks, G.L., et al. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nat. Genet. 51, 1044–1051. https://doi.org/10.1038/s41588-019-0410-2.

29. Contreras-Moreira, B., Cantalapiedra, C.P., García-Pereira, M.J., Gordon, S.P., Vogel, J.P., Igartua, E., Casas, A.M., and Vinuesa, P. (2017). Analysis of plant pan-genomes and transcriptomes with GET_HOMOLOGUES-EST, a clustering solution for sequences of the same species. Front. Plant Sci. 8, 184. https://doi.org/10.3389/fpls.2017.00184.

30. Hurgobin, B., Golicz, A.A., Bayer, P.E., Chan, C.K.K., Tirnaz, S., Dolatabadian, A., Schiessl, S.V., Samans, B., Montenegro, J.D., Parkin, I.A.P., et al. (2018). Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid Brassica napus. Plant Biotechnol. J. 16, 1265–1274. https://doi.org/10.1111/pbi.12867.

31. Golicz, A.A., Bayer, P.E., Barker, G.C., Edger, P.P., Kim, H., Martinez, P.A., Chan, C.K.K., Severn-Ellis, A., McCombie, W.R., Parkin, I.A.P., et al. (2016). The pangenome of an agronomically important crop plant Brassica oleracea. Nat. Commun. 7, 13390. https://doi.org/10.1038/ncomms13390.

32. Montenegro, J.D., Golicz, A.A., Bayer, P.E., Hurgobin, B., Lee, H., Chan, C.K.K., Visendi, P., Lai, K., Doležel, J., Batley, J., and Edwards, D. (2017). The pangenome of hexaploid bread wheat. Plant J. 90, 1007–1013. https://doi.org/10.1111/tpj.13515.

33. Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M., Zheng, T., Fuentes, R.R., Zhang, F., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature 557, 43–49. https://doi.org/10.1038/s41586-018-0063-9.

34. Li, Y.H., Zhou, G., Ma, J., Jiang, W., Jin, L.G., Zhang, Z., Guo, Y., Zhang, J., Sui, Y., Zheng, L., et al. (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits.

Nat. Biotechnol. 32, 1045–1052. https://doi.org/10.1038/nbt.2979.

35. Gordon, S.P., Contreras-Moreira, B., Woods, D.P., Des Marais, D.L., Burgess, D., Shu, S., Stritt, C., Roulin, A.C., Schackwitz, W., Tyler, L., et al. (2017). Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure. Nat. Commun. 8, 2184. https://doi.org/10.1038/s41467-017-02292-8.

36. McMahon, A., Lewis, E., Buniello, A., Cerezo, M., Hall, P., Sollis, E., Parkinson, H., Hindorff, L.A., Harris, L.W., and MacArthur, J.A.L. (2021). Sequencing-based genome-wide association studies reporting standards. Cell Genom. 1, 100005. https://doi.org/10.1016/j.xgen.2021.100005.

37. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An expanded view of complex traits: from polygenic to omnigenic. Cell 169, 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038.

38. Morita-Yamamuro, C., Tsutsui, T., Sato, M., Yoshioka, H., Tamaoki, M., Ogawa, D., Matsuura, H., Yoshihara, T., Ikeda, A., Uyeda, I., and Yamaguchi, J. (2005). The Arabidopsis gene CAD1 controls programmed cell death in the plant immune system and encodes a protein containing a MACPF domain. Plant Cell Physiol. 46, 902–912. https://doi.org/10.1093/pcp/pci095.

39. De Benedictis, M., Brunetti, C., Brauer, E.K., Andreucci, A., Popescu, S.C., Commisso, M., Guzzo, F., Sofo, A., Ruffini Castiglione, M., Vatamaniuk, O.K., and Sanità di Toppi, L. (2018). The. Front. Plant Sci. 9, 19. https://doi.org/10.3389/fpls.2018.00019.

40. Dietrich, P., Moeder, W., and Yoshioka, K. (2020). Plant cyclic nucleotide-gated channels: new insights on their functions and regulation. Plant Physiol. 184, 27–38. https://doi.org/10.1104/pp.20.00425.

41. Lu, Z., Kong, X., Lu, Z., Xiao, M., Chen, M., Zhu, L., Shen, Y., Hu, X., and Song, S. (2014). Para-aminobenzoic acid (PABA) synthase enhances thermotolerance of mushroom Agaricus bisporus. PLoS One 9, e91298. https://doi.org/10.1371/journal.pone.0091298.

42. Ribeiro, C., Xu, J., Hendrich, C., Pandey, S.S., Yu, Q., Gmitter, F., and Wang, N. (2022). Seasonal transcriptome profiling of susceptible and tolerant citrus cultivars to citrus Huanglongbing. Phytopathology. https://doi.org/10.1094/phyto-05-22-0179-r.

43. Wu, B., Yu, Q., Deng, Z., Duan, Y., Luo, F., and Gmitter, F. (2022). A chromosome-level phased Citrus sinensis genome facilitates understanding Huanglongbing tolerance mechanisms at the allelic level in an irradiation-induced mutant. Preprint at bioRxiv. https://doi.org/10.1101/2022.02.05.479263.

44. Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis

with GWAS: a review. Plant Methods 9, 29. https://doi.org/10.1186/1746-4811-9-29.

45. Imai, A., Nonaka, K., Kuniga, T., Yoshioka, T., and Hayashi, T. (2018). Genome-wide association mapping of fruit-quality traits using genotyping-by-sequencing approach in citrus landraces, modern cultivars, and breeding lines in Japan. Tree Genet. Genomes 14, 24. https://doi.org/10.1007/s11295-018-1238-0.

46. Mattia, M.R., Du, D., Yu, Q., Kahn, T., Roose, M., Hiraoka, Y., Wang, Y., Munoz, P., and Gmitter, F.G. (2022). Genome-wide association study of healthful flavonoids among diverse Mandarin accessions. Plants 11, 317. https://doi.org/10.3390/plants11030317.

47. Minamikawa, M.F., Nonaka, K., Kaminuma, E., Kajiya-Kanegae, H., Onogi, A., Goto, S., Yoshioka, T., Imai, A., Hamada, H., Hayashi, T., et al. (2017). Genome-wide association study and genomic prediction in citrus: potential of genomics-assisted breeding for fruit quality traits. Sci. Rep. 7, 4721. https://doi.org/10.1038/s41598-017-05100-x.

48. Prasad, S., Xu, J., Zhang, Y., and Wang, N. (2016). SEC-translocon dependent extracytoplasmic proteins of Candidatus liberibacter asiaticus. Front. Microbiol. 7, 1989. https://doi.org/10.3389/fmicb.2016.01989.

49. Clark, K., Franco, J.Y., Schwizer, S., Pang, Z., Hawara, E., Liebrand, T.W.H., Pagliaccia, D., Zeng, L., Gurung, F.B., Wang, P., et al. (2018). An effector from the Huanglongbing-associated pathogen targets citrus proteases. Nat. Commun. 9, 1718. https://doi.org/10.1038/s41467-018-04140-9.

50. Pang, Z., Zhang, L., Coaker, G., Ma, W., He, S.Y., and Wang, N. (2020). Citrus CsACD2 is a target of Candidatus liberibacter asiaticus in huanglongbing disease. Plant Physiol. 184, 792–805. https://doi.org/10.1104/pp.20.00348.

51. Liese, A., and Romeis, T. (2013). Biochemical regulation of in vivo function of plant calcium-dependent protein kinases (CDPK). Biochim. Biophys. Acta 1833, 1582–1589. https://doi.org/10.1016/j.bbamcr.2012.10.024.

52. Zou, H., Gowda, S., Zhou, L., Hajeri, S., Chen, G., and Duan, Y. (2012). The destructive citrus pathogen, 'Candidatus Liberibacter asiaticus' encodes a functional flagellin characteristic of a pathogen-associated molecular pattern. PLoS One 7, e46447. https://doi.org/10.1371/journal.pone.0046447.

53. Andrade, M.O., Pang, Z., Achor, D.S., Wang, H., Yao, T., Singer, B.H., and Wang, N. (2020). The flagella of 'Candidatus Liberibacter asiaticus' and its movement in planta. Mol. Plant Pathol. 21, 109–123. https://doi.org/10.1111/mpp.12884.

54. Kadota, Y., Sklenar, J., Derbyshire, P., Stransfeld, L., Asai, S., Ntoukakis, V., Jones, J.D., Shirasu, K., Menke, F., Jones, A., and Zipfel, C. (2014). Direct regulation of the

NADPH oxidase RBOHD by the PRR-associated kinase BIK1 during plant immunity. Mol. Cell 54, 43–55. https://doi.org/10.1016/j.molcel.2014.02.021.

55. Liang, X., Ding, P., Lian, K., Wang, J., Ma, M., Li, L., Li, L., Li, M., Zhang, X., Chen, S., et al. (2016). Arabidopsis heterotrimeric G proteins regulate immunity by directly coupling to the FLS2 receptor. Elife 5, e13568. https://doi.org/10.7554/eLife.13568.

56. Couto, D., Niebergall, R., Liang, X., Bücherl, C.A., Sklenar, J., Macho, A.P., Ntoukakis, V., Derbyshire, P., Altenbach, D., Maclean, D., et al. (2016). The arabidopsis protein phosphatase PP2C38 negatively regulates the central immune kinase BIK1. PLoS Pathog. 12, e1005811. https://doi.org/10.1371/journal.ppat.1005811.

57. Jia, H., Omar, A.A., Orbović, V., and Wang, N. (2022). Biallelic editing of the. Phytopathology 112, 308–314. https://doi.org/10.1094/PHYTO-04-21-0144-R.

58. Huang, X., Wang, Y., and Wang, N. (2021). Highly efficient generation of canker-resistant sweet orange enabled by an improved CRISPR/Cas9 system. Front. Plant Sci. 12, 769907.

59. Ding, S.W. (2022). Transgene silencing, RNA interference, and the antiviral defense mechanism directed by small interfering RNAs. Phytopathology. https://doi.org/10.1094/phyto-10-22-0358-ia.

60. Huang, X., Wang, Y., Xu, J., and Wang, N. (2020). Development of multiplex genome editing toolkits for citrus with high efficacy in biallelic and homozygous mutations. Plant Mol. Biol. 104, 297–307. https://doi.org/10.1007/s11103-020-01043-6.

61. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods 18, 170–175. https://doi.org/10.1038/s41592-020-01056-5.

62. Zhang, X., Zhang, S., Zhao, Q., Ming, R., and Tang, H. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. Nat. Plants 5, 833–845. https://doi.org/10.1038/s41477-019-0487-8.

63. Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3, 99–101. https://doi.org/10.1016/j.cels.2015.07.012.

64. Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol. Biol. Evol. 38, 4647–4654. https://doi.org/10.1093/molbev/msab199.

65. Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes.

Bioinformatics 23, 1061–1067. https://doi.org/10.1093/bioinformatics/btm071.

66. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595. https://doi.org/10.1093/bioinformatics/btp698.

67. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580. https://doi.org/10.1093/nar/27.2.573.

68. Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 35, W265–W268. https://doi.org/10.1093/nar/gkm286.

69. Lian, S., Chen, X., Wang, P., Zhang, X., and Dai, X. (2016). A complete and accurate ab initio repeat finding algorithm. Interdiscip. Sci. 8, 75–83. https://doi.org/10.1007/s12539-015-0119-6.

70. Tempel, S. (2012). Using and understanding RepeatMasker. Methods Mol. Biol. 859, 29–51. https://doi.org/10.1007/978-1-61779-603-6_2.

71. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinf. 10, 421. https://doi.org/10.1186/1471-2105-10-421.

72. Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and genomewise. Genome Res. 14, 988–995. https://doi.org/10.1101/gr.1865504.

73. Alioto, T., Blanco, E., Parra, G., and Guigó, R. (2018). Using geneid to identify genes. Curr. Protoc. Bioinformatics 64, e56. https://doi.org/10.1002/cpbi.56.

74. Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268, 78–94. https://doi.org/10.1006/jmbi.1997.0951.

75. Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20, 2878–2879. https://doi.org/10.1093/bioinformatics/bth315.

76. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.

77. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084.

78. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based

genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.

79. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/nbt.3122.

80. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. *9*, R7. https://doi.org/10.1186/gb-2008-9-1-r7.

81. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60. https://doi.org/10.1038/nmeth.3176.

82. Lowe, T.M., and Chan, P.P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. *44*, W54–W57. https://doi.org/10.1093/nar/gkw413.

83. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

84. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

85. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158.

86. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

87. Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., Li, Y., Ye, J., Yu, C., Li, Z., et al. (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. GigaScience *7*, 1–6. https://doi.org/10.1093/gigascience/gix120.

88. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

89. Zhou, X., and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat. Methods *11*, 407–409. https://doi.org/10.1038/nmeth.2848.

90. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169. https://doi.org/10.1093/bioinformatics/btu638.

91. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

92. Rana, M.M., Aycan, M., Takamatsu, T., Kaneko, K., Mitsui, T., and Itoh, K. (2019). Optimized nuclear pellet method for extracting next-generation sequencing quality genomic DNA from fresh leaf tissue. Methods Protoc. *2*, 54. https://doi.org/10.3390/mps2020054.

93. Kozarewa, I., Ning, Z., Quail, M.A., Sanders, M.J., Berriman, M., and Turner, D.J. (2009). Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat. Methods *6*, 291–295. https://doi.org/10.1038/nmeth.1311.

94. Belton, J.M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. Methods *58*, 268–276. https://doi.org/10.1016/j.ymeth.2012.05.001.

95. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. USA *117*, 9451–9457. https://doi.org/10.1073/pnas.1921046117.

96. Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. *110*, 462–467. https://doi.org/10.1159/000084979.

97. Nachtweide, S., and Stanke, M. (2019). Multi-genome annotation with AUGUSTUS. Methods Mol. Biol. *1962*, 139–160. https://doi.org/10.1007/978-1-4939-9173-0_8.

98. Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., and Bairoch, A. (2007). UniProtKB/Swiss-Prot. Methods Mol. Biol. *406*, 89–112. https://doi.org/10.1007/978-1-59745-535-0_4.

99. Gertz, E.M., Yu, Y.K., Agarwala, R., Schäffer, A.A., and Altschul, S.F. (2006). Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol. *4*, 41. https://doi.org/10.1186/1741-7007-4-41.

100. Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: protein domains identifier. Nucleic Acids Res. *33*, W116–W120. https://doi.org/10.1093/nar/gki442.

101. Servant, F., Bru, C., Carrère, S., Courcelle, E., Gouzy, J., Peyruc, D., and Kahn, D. (2002). ProDom: automated clustering of homologous domains. Brief. Bioinform. *3*, 246–251. https://doi.org/10.1093/bib/3.3.246.

102. Attwood, T.K., Beck, M.E., Bleasby, A.J., and Parry-Smith, D.J. (1994). PRINTS–a database of protein motif fingerprints. Nucleic Acids Res. *22*, 3590–3596.

103. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: the protein families database in 2021. Nucleic Acids Res. *49*, D412–D419. https://doi.org/10.1093/nar/gkaa913.

104. Lou, F., Song, N., Han, Z., and Gao, T. (2020). Single-molecule real-time (SMRT) sequencing facilitates Tachypleus tridentatus genome annotation. Int. J. Biol. Macromol. *147*, 89–97. https://doi.org/10.1016/j.ijbiomac.2020.01.029.

105. Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nat. Protoc. *8*, 1551–1566. https://doi.org/10.1038/nprot.2013.092.

106. Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., and Thomas, P.D. (2019). Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat. Protoc. *14*, 703–721. https://doi.org/10.1038/s41596-019-0128-8.

107. Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P.S., Pagni, M., and Sigrist, C.J.A. (2006). The PROSITE database. Nucleic Acids Res. *34*, D227–D230. https://doi.org/10.1093/nar/gkj063.

108. Gene Ontology Consortium (2015). Gene Ontology consortium: going forward. Nucleic Acids Res. *43*, D1049–D1056. https://doi.org/10.1093/nar/gku1179.

109. Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., et al. (2009). InterPro: the integrative protein signature database. Nucleic Acids Res. *37*, D211–D215. https://doi.org/10.1093/nar/gkn785.

110. Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. *35*, D61–D65. https://doi.org/10.1093/nar/gkl842.

111. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. *45*, D353–D361. https://doi.org/10.1093/nar/gkw1092.

112. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Research 47, D309–D314.

113. Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155. https://doi.org/10.1126/science.290.5494.1151.

114. Gao, Y., Zhao, H., Jin, Y., Xu, X., and Han, G.Z. (2017). Extent and evolution of gene duplication in DNA viruses. Virus Res. 240, 161–165. https://doi.org/10.1016/j.virusres.2017.08.005.

115. Kalvari, I., Nawrocki, E.P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., Griffiths-Jones, S., Toffano-Nioche, C., Gautheret, D., Weinberg, Z., et al. (2021). Rfam 14: expanded coverage of metagenomic, viral and microRNA families. Nucleic Acids Res. 49, D192–D200. https://doi.org/10.1093/nar/gkaa1047.

116. Wang, L., Huang, Y., Liu, Z., He, J., Jiang, X., He, F., Lu, Z., Yang, S., Chen, P., Yu, H., et al. (2021). Somatic variations led to the selection of acidic and acidless orange cultivars. Nat. Plants 7, 954–965. https://doi.org/10.1038/s41477-021-00941-x.

117. Zhang, Y., Barthe, G., Grosser, J.W., and Wang, N. (2016). Transcriptome analysis of root response to citrus blight based on the newly assembled Swingle citrumelo draft genome. BMC Genom. 17, 485. https://doi.org/10.1186/s12864-016-2779-y.

118. Storey, J.D. (2003). The positive false discovery rate: abayesian interpretation and the q-value. Ann. Statist. 31, 2013–2035.

119. Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., and Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. 45, 122–129. https://doi.org/10.1093/nar/gkx382.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
C. sinensis Osbeck cv. Newhall	National Navel Orange Engineering Research Center, Ganzhou, Jiangxi Province, China	N/A
Leaf and root samples of citrus germplasms	Lower Variety Grove, California Citrus State Historic Park, Riverside, California	N/A
Deposited data		
Genome sequencing data of *C. sinensis*'Newhall'	This study	SRA Bioproject: PRLNA810206.
Genome sequencing data of new sequenced 91 citrus accessions	This study	SRA Bioproject: PRJNA698060
Software and algorithms		
Hifiasm v 0.15.4	Cheng et al., 2021[61]	https://github.com/chhylp123/hifiasm
ALLHIC v 0.9.8	Zhang et al., 2019[62]	https://github.com/tangerzhang/ALLHiC
juicebox v1.11.08	Durand et al., 2016[63]	https://github.com/aidenlab/Juicebox
BUSCO V10	Manni et al., 2021[64]	https://github.com/WenchaoLin/BUSCO-Mod
CEGMA v2	Parra et al., 2007[65]	https://github.com/KorfLab/CEGMA_v2
BWA v 0.7.17	Li and Durbin, 2010[66]	https://github.com/lh3/bwa
TRF	Benson, 1999[67]	https://github.com/Benson-Genomics-Lab/TRF
LTR_FINDER	Xu and Wang, 2007[68]	https://github.com/xzhub/LTR_Finder
RepeatScout	Lian et al., 2016[69]	https://github.com/mmcco/RepeatScout
RepeatMasker	Tempel, 2012[70]	https://github.com/rmhubley/RepeatMasker
BLAST v2.2.26	Camacho et al., 2009[71]	https://github.com/ncbi/blast_plus_docs
GeneWise v2.4.1	Birney et al., 2004[72]	https://bio.tools/genewise
Geneid v1.4	Alioto et al., 2018[73]	https://github.com/guigolab/geneid
Genescan v1.0	Burge and Karlin, 1997[74]	http://hollywood.mit.edu/GENSCAN.html
GlimmerHMM v3.04	Majoros et al., 2004[75]	https://github.com/kblin/glimmerhmm
Trinity v2. 1. 1	Grabherr et al., 2011[76]; Haas et al., 2013[77]	https://github.com/trinityrnaseq/trinityrnaseq
Hisat v2.0.4	Kim et al., 2019[78]	https://github.com/DaehwanKimLab/hisat2
StringTie v1.3.3	Pertea et al., 2015[79]	https://github.com/gpertea/stringtie
EvidenceModeler (EVM) v1. 1. 1	Haas et al., 2008[80]	https://github.com/EVidenceModeler
DIAMOND v0.8.22	Buchfink et al., 2015[81]	https://github.com/bbuchfink/diamond
tRNAscan-SE	Lowe and Chan, 2016[82]	https://github.com/UCSC-LoweLab/tRNAscan-SE
SAMtools v1.10	Li et al., 2009[83]	https://github.com/samtools/
GATK v4.1.9.0	McKenna et al., 2010[84]	https://github.com/broadgsa/gatk
CD-HIT v4.8	Li and Godzik, 2006[85]	https://github.com/weizhongli/cdhit
OrthoFinder v2.2.7	Emms and Kelly, 2019[86]	https://github.com/davidemms/OrthoFinder
SOAPnuke v1.5.3	Chen et al., 2018[87]	https://github.com/BGI-flexlab/SOAPnuke
Bowtie2 v 2.2.6	Langmead and Salzberg, 2012[88]	https://github.com/BenLangmead/bowtie2
GEMMA v 0.98.1	Zhou and Stephens, 2014[89]	https://github.com/genetics-statistics/GEMMA
HTSeq-count v0.11.2	Anders et al., 2015[90]	https://github.com/simon-anders/htseq
DESeq2	Love et al., 2014[91]	https://github.com/mikelove/DESeq2

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to, and will be fulfilled by, the lead contact, Nian Wang (nianwang@ufl.edu).

Materials availability

Besides data, this study did not generate any new reagents or materials.

Data and code availability

The raw sequencing reads of new sequenced genome for *C. sinensis*'Newhall' were deposited in the NCBI database under the accession number BioProject: PRJNA810206. The raw sequencing reads of new sequenced 91 citrus accessions were deposited in the NCBI database under the accession number Bio-Project: PRJNA698060. The raw sequencing of other citrus accessions was downloaded from public databases with accession number in Tables S15 and S20. The RNA-seq data used in ASE analysis were from published bioprojects, including BioProject: PRJNA739184 for Valencia and BioProject: PRJNA739186 for SB mandarin. All the published assembly citrus genomes were downloaded from two public citrus genome databases, including CPBD: Citrus Pan-genome to Breeding Database (http://citrus.hzau.edu.cn/index.php) and Citrus Genome Database (https://www.citrusgenomedb.org/).

This paper does not report original code. All the data analysis were performed using published programs, which have been described in the method details.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

This study does not include experimental model or subjects.

METHOD DETAILS

Plant materials

The Newhall navel sweet orange (*C. sinensis* Osbeck cv. Newhall*)* was sequenced in this study. The plant materials of Newhall navel sweet orange were obtained for genomic DNA and RNA extractions from the National Navel Orange Engineering Research Center, Ganzhou, Jiangxi Province, China. Leaf and root samples of citrus germplasms were collected from the Lower Variety Grove, California Citrus State Historic Park, Riverside, California for GWAS study.

DNA and RNA extraction

Genomic DNA for Illumina sequencing was extracted using the phenol-chloroform method.[92] Genomic DNA was isolated using Nanobind Plant Nuclei Big DNA Kit (Circulomics Inc., Baltimore, MD, USA) following the manufacturer's instructions for PacBio and Hi-C sequencing. For full length transcript sequencing (Iso-Seq), the RNA was extracted using an RNAprep Plant Kit (Qiagen, Valencia, CA, USA). Genomic DNA for BGI sequencing was extracted using MoBio Powersoil DNA extraction kit (MoBio Laboratories Inc. Carlsbad, CA, USA) following the manufacturer's instructions. The quality of genomic DNA and RNA was evaluated using agarose gel electrophoresis and Qubit 3.0 Fluorometer (Life Technologies, USA).

Library construction and sequencing

Following the manufacturer's protocol of short read DNA sequencing from Illumina,[93] the library was prepared. After quality control, quantification, and normalization of the DNA libraries, 150-bp paired-end reads were generated using the Illumina NovaSeq 6000 platform according to the manufacturer's instructions.

PacBio HiFi SMRTbell Library with 15 kb DNA fragment was constructed following the manufacturer's protocol. The constructed library was then sequenced by Pacbio Sequel II platform according to the manufacturer's instructions.

Hi-C libraries were prepared by a standard procedure.[94] Formaldehyde solution was used to fix plant cells. Then, 2.5 M glycine was added to quench the cross-linking reaction. The crosslink DNA was treated with restriction enzymes (e.g., HindIII), creating a gap on both sides of the crosslinking point. The exposed DNA ends were repaired and covalently linked with biotin-14-dCTP (Invitrogen Life Technologies, Carlsbad, CA), and then connected by T4 DNA ligase (Invitrogen Life Technologies, Carlsbad, CA) to form a closed random circular DNA structure. Proteinase K (Invitrogen Life Technologies, Carlsbad, CA) was used to digest proteins at the junction point to disconnect the crosslink of the protein and DNA. Genomic DNA was extracted and later fragmented into 350 bp. Capturing and enriching biotin-labeled fragments was conducted via the affinity of streptomycin to biotin, which was used to construct Illumina library. Hi-C sequencing libraries were amplified by PCR and sequenced on Illumina HiSeq-2500 platform (PE 125 bp). The 91 citrus genomes for GWAS study were sequenced using BGISEQ500 platform. Shotgun genomic library preparation and sequencing were performed per the manufacturer's protocol at BGI-Shenzhen, China. Briefly, 500 ng of input DNA was used for library generation and fragmented ultrasonically to yield 400 to 600 bp of fragments. DNA fragments were then end-repaired and A-tailed, and adaptors with specific barcodes were added. PCR amplification of DNA fragments was carried out to generate a single-strand circular DNA library. The DNA libraries were sequenced by BGISEQ500 using a paired-end 100-bp sequencing strategy. On average, more than 41.9 Gb of raw data were generated for each genomic sample for GWAS analysis.

De novo genome assembly for Newhall navel sweet orange

Hifiasm v 0.15.4[61] with default parameters was used to assemble HiFi reads into scaffolds, which were corrected using short Illumina DNA reads. Based on Hi-C sequencing data, the assembled scaffold sequences were mounted to the near-chromosome level using ALLHIC v 0.9.8.[62] According to chromosome interaction intensity using juicebox v1.11.08,[63] the near-chromosome level genome was manually corrected into a chromosome-level genome.

Genome quality assessment

To assess the assembly quality of the assembled genomes, the completeness of the assembly was evaluated using Benchmarking Universal Single-Copy Orthologs (BUSCO) v10,[64] and the Conserved Core Eukaryotic Gene Mapping Approach (CEGMA) v2.[65] The coverage of assembled genomes was also calculated by mapping Illumina short reads to the assembly using Burrows-Wheeler Aligner (BWA).[66]

Repeat element identification

Repeat elements of whole genome were identified using both homology alignment and de novo prediction. Tandem repeat was extracted using TRF[67] by ab initio prediction. The de novo repetitive element library was constructed using LTR_FINDER,[68] RepeatScout,[69] and RepeatModeler.[95] Homolog repeats were predicted based on Repbase database[96] employing RepeatMasker software[70] and its in-house scripts RepeatProteinMask with default parameters. Using uclust program, a non-redundant transposable element (TE) library (a combination of homolog repeats and de novo TE) was generated, which was applied to mask the genome using RepeatMasker software.

Gene structure annotation

Homology-based prediction, ab initio prediction, and RNA-Seq assisted prediction were employed to perform the gene model prediction. Genomic sequences were aligned to homologous proteins using tblastn v2.2.26[71] with a threshold of E-value $\leq$ 1e−5. Based on the matched proteins from reference genomes, GeneWise (v2.4.1)[72] software was used to predict gene structure. Augustus v3.2.3,[97] Geneid v1.4,[73] Genescan v1.0,[74] GlimmerHMM v3.04,[75] and SNAP_2013-11-29 were used for the automated de novo gene prediction. The transcriptome assembly was performed using Trinity v2. 1. 1[77,76] for the genome annotation. To identify exon region and splice positions, the RNA-Seq reads from leaf, root, and fruit tissues were aligned to genome using Hisat v2.0.4[78] with default parameters. The alignment results were then used as input for StringTie v1.3.3[79] with default parameters for genome-based transcript assembly. The non-redundant reference gene set was generated by merging genes predicted from three methods using EvidenceModeler (EVM) v1. 1. 1.[80]

Gene functional annotation

Gene functions were assigned according to the best match by aligning the protein sequences to the Swiss-Prot[98] using Blastp[71] with a threshold of E-value $\leq$ 1e−5.[99] The motifs and domains were annotated using

InterProScan v5.31[100] by searching publicly available databases, including Pro-Dom,[101] PRINTS,[102] Pfam,[103] SMRT,[104] PANTHER[105,106] and PROSITE.[107] The Gene Ontology (GO)[108] IDs for each gene were assigned according to the corresponding Inter-Pro[109] entry. We predicted the proteins function by transferring annotations from the closest BLAST hit with a threshold of E-value $<10^{-5}$ from the Swiss-Prot[98] and the NR databases[110] using DIAMOND v0.8.22.[81] We also mapped gene set to the KEGG[111] and eggNOG databases[112] to generate more comprehensive gene functional information. To infer the evolutionary trajectories of genes in Newhall navel orange, we calculated the number of substitutions per synonymous site (Ks) and the number of substitutions per nonsynonymous site (Ka) for each homologous gene pairs. A Ka/Ks ratio more than 1, less than 1, and equal to 1 indicates positive selection, purifying selection and neutral evolution, respectively.[113,114]

Non-coding RNA annotation

tRNAs were predicted using the program tRNAscan-SE.[82] rRNA sequences were predicted using Blast[71] with relative species' rRNA sequences. Other ncRNAs, including miRNAs and snRNAs, were identified by searching the Rfam database[115] with default parameters.

Identification of genomic variations

We used For SNPs and InDels identification, short sequencing reads were aligned to the high-quality sweet orange genome as the reference genome[116] using BWA software v 0.7.17 (Li and Durbin, 2010). Duplicated mapping reads and unmapped reads were removed with the SAMtools v1.10.[83] All genotype information for the polymorphic sites was generated using the GATK v4.1.9.0 population method.[84] The generated SNPs and InDels were filtered by quality and sequence depth.

Pan genome construction

To construct the citrus pan genome, we generated the gene families from 10 citrus and relative genomes (Table S1). For each genome, a gene containing CDS with 100% similarity to other genes was removed by using the cd-hit-est of CD-HIT v4.8 toolkit.[85] Protein sequences of the remaining genes were subjected to homologous searching by BLASTp[71] with default parameters. OrthoFinder v2.2.7 [86] was used to deal with the BLAST result with default parameters to make gene family clustering. The gene families shared among all accessions were defined as core gene families. The gene families that were missed in one or two accessions were defined as softcore gene families. The gene families that were missed in more than two accessions were defined as dispensable gene families, and gene families that only existed in one accession were defined as private gene families. We used eggNOG database to generate gene functional information for the citrus pan genome. The eggNOG pathway enrichment analysis of core, softcore, dispensable, and private gene families was performed using Fisher exact test in R program.

GWAS

To generate comprehensive SNP genotyping data for GWAS analysis, we sequenced 91 new citrus genomes and collected 356 representative genomic data of citrus and its relatives from public database (Tables S20 and S21). Raw reads for each genomic sample were filtered by SOAPnuke (v1.5.3) with the parameters set as "filterMeta-Q 2-S-L 15-N 3 –P 0.5-q 20-L 60-R 0.5–5 0".[87] The trimmed reads were mapped to the sweet orange genomes,[21,116] C. clementina,[1] pummelo, Citron, kumquat, Atlantia, Papeda,[22] Citrus mangshanensis,[24] and Swingle citrumelo[117] genomes using Bowtie2 software[88] to identify the citrus reads. The high quality paired-end short citrus genomic reads were mapped to pummelo (Citrus maxima)[22] reference genome, which is of the highest-quality among sequenced genomes, using BWA software v 0.7.17.[66] Duplicated mapping reads were removed with the SAMtools package v1.10.[83] All genotype information for the polymorphic sites was generated using the GATK v4.1.9.0 population method.[84] The generated SNPs were filtered by quality and sequence depth, and SNPs with minor allele frequency (MAF) more than 0.01 and individual level missingness less than 10% were retained for GWAS analysis.

To perform the association between HLB and citrus genomic variations, we collected the HLB resistant or tolerant information for each citrus genotypes from previous studies (Table S20). The population structure of citrus was estimated using PCA method.[47] GWAS for HLB was conducted using the univariate linear mixed model in the GEMMA package v0.98.1, which accounts for population stratification using the first two principal components (PCs) of population structure and kinship matrices.[89] P-values for multiple testing were corrected using the FDR method in R program.[118] All items with corrected P-values below

1e-5 were considered significant. Significantly associated SNPs were mapped to the citrus reference genome to acquire gene annotations. The genes containing significantly associated SNPs were functionally annotated using the KEGG and agriGO 2.0 databases.[119]

Allelic specific expression of transcriptomes of valencia sweet orange and Mandarin LB9-9 in response to CLas infection

We generated the allelic specific expression of genes (ASE) of Valencia sweet orange and Mandarin LB9-9.[27] Briefly, cleaned RNA-seq reads were first mapped against both alleles of genome of Newhall navel sweet orange using HISAT2 v2.0.4[78] and SAMtools v1.10[83] to select good-quality alignments. Reads uniquely aligned to the same chromosome in both alleles were preserved and assigned to the allele if the alignment had a higher score and fewer mismatches than to the other allele. Reads that failed to be assigned, which were mainly from homozygous genomic region or regions that were unassembled or haplotype unresolved, were not considered in downstream analyses. The allele-specific reads were mapped to the Newhall navel sweet orange consensus genome using HISAT2 v2.0.4 and SAMtools v1.10. Raw counts for each genes were quantified using HTSeq-count v0.11.2.[90] Genes with total counts >10 in all biological replicates were analyzed for ASE using DESeq2 in R program.[91] A gene was considered to have ASE if the expression difference of the two alleles was significantly greater than 2-fold (adjusted $p < 0.05$).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

The statistical analysis details have been described in methods and results. All the statistical analysis was perform using R program.