



Database tool

CitrusKB: a comprehensive knowledge base for transcriptome and interactome of *Citrus* spp. infected by *Xanthomonas citri* subsp. *citri* at different infection stages

Adriano Ferrasa^{1,3}, Mayara M. Murata^{2,7}, Teresa D. C. G. Cofre²,
Juliana S. Cavallini², Gustavo Peron², Maria H. M. Julião²,
José Belasque⁴, Henrique Ferreira⁵, Maria Inês T. Ferro², Rui P. Leite Jr⁶,
Helen A. Penha^{2,8}, Flávia M. S. Carvalho², Alessandro M. Varani²,
Roberto H. Herai^{3,*} and Jesus A. Ferro^{2,*}

¹Departamento de Informática, Universidade Estadual de Ponta Grossa (UEPG), Av. Carlos Cavalcanti, 4748, 84030-900, Ponta Grossa, PR, Brazil, ²Departamento de Tecnologia, Universidade Estadual Paulista (UNESP), Via de Acesso Prof. Paulo Donato Castellane s/n, 14884-900, Jaboticabal, SP, Brazil, ³Graduate Program in Health Sciences, School of Medicine, Pontifícia Universidade Católica do Paraná (PUCPR), R. Imaculada Conceição, 1155, 80215-901, Curitiba, PR, Brazil, ⁴Departamento de Fitopatologia e Nematologia, Escola Superior de Agricultura 'Luiz de Queiroz', Universidade de São Paulo (USP), Av. Pádua Dias, 11, 13418-900, Piracicaba, SP, Brazil, ⁵Departamento de Biologia Geral e Aplicada, Instituto de Biociências, Universidade Estadual Paulista (UNESP), Av. 24-A, 1515, 13506-900, Rio Claro, SP, Brazil, ⁶Instituto Agrônômico do Paraná (IAPAR), Londrina, Rod. Celso Garcia Cid, km 375, 86047-902, Londrina, PR, Brazil, ⁷Diretoria de Pós-Graduação Stricto Sensu, Universidade Norte do Paraná (UNOPAR), Av. Paris, 675, 86041-100, Londrina, PR, Brazil and ⁸Inpreha Biotecnologia e Desenvolvimento Avançado LTDA, Fazenda Lagoinha, Estrada Velha de Taquaritinga, Km 04, 14870-970, Jaboticabal, SP, Brazil

*Corresponding author: Jesus Aparecido Ferro, Tel: +55 16 3209-2675; Email: jesus.ferro@unesp.br
Correspondence may also be addressed to Roberto Hirochi Herai. Tel: +55 41 3271-6341; Email: roberto.herai@pucpr.br

Citation details: Ferrasa,A., Murata,M.M., Cofre,T.D.C.G. *et al.* CitrusKB: a comprehensive knowledge base for transcriptome and interactome of *Citrus* spp. infected by *Xanthomonas citri* subsp. *citri* at different infection stages. *Database* (2020) Vol. XXXX: article ID baaa081; doi:10.1093/database/baaa081

Received 20 March 2020; Revised 5 August 2020; Accepted for publication 31 August 2020

Abstract

Citrus canker type A is a serious disease caused by *Xanthomonas citri* subsp. *citri* (*X. citri*), which is responsible for severe losses to growers and to the citrus industry worldwide. To date, no canker-resistant citrus genotypes are available, and there is limited information regarding the molecular and genetic mechanisms involved in the early stages of the citrus canker development. Here, we present the CitrusKB knowledge base. This is the first *in vivo* interactome database for different citrus cultivars, and it was produced to provide a valuable resource of information on citrus and their interaction with the citrus canker bacterium *X. citri*. CitrusKB provides tools for a user-friendly web interface to let users search and analyse a large amount of information regarding eight citrus

cultivars with distinct levels of susceptibility to the disease, with controls and infected plants at different stages of infection by the citrus canker bacterium *X. citri*. Currently, CitrusKB comprises a reference citrus genome and its transcriptome, expressed transcripts, pseudogenes and predicted genomic variations (SNPs and SSRs). The updating process will continue over time by the incorporation of novel annotations and analysis tools. We expect that CitrusKB may substantially contribute to the field of citrus genomics. CitrusKB is accessible at <http://bioinfo.deinfo.uepg.br/citrus>. Users can download all the generated raw sequences and generated datasets by this study from the CitrusKB website.

Introduction

Citrus canker A, caused by the Gram-negative bacterium *Xanthomonas citri* subsp. *citri* (*X. citri*), is one of the main diseases affecting citrus trees and is a threat for orange production in several countries around the world (1–3). The symptoms of citrus canker on susceptible trees include raised brownish circular lesions on leaves, stems and fruits (4). Several disease management procedures have been applied in attempts to control citrus canker, including pruning of infected trees, post-harvested treatment of fruits, decontamination of equipment and personnel and, within endemic regions of citrus canker, it is applied a spray of copper-containing chemicals to protect young leaves and fruit against bacterium infection (5, 6). Furthermore, additional measures to control the disease involve planting windbreaks in the citrus orchards, control of the citrus leaf miner *Phyllocnistis citrella* and production of healthy citrus nursery trees (6, 7).

Due to these challenging attempts for disease control, several efforts have been made to understand the mechanisms of plant–pathogen interaction and the disease development molecular basis, with the objective to establish more effective measures for the control of citrus canker.

The molecular basis and the genetic mechanisms involved in the early stages of the citrus canker development may be revealed by studies on citrus species and cultivars with different levels of susceptibility to the disease. For instance, kumquats (*Fortunella* spp.) and ‘Mexican’ lime (*Citrus aurantifolia* (Christm.) Swingle) are considered resistant and susceptible to citrus canker, respectively. Furthermore, other citrus cultivars exhibit intermediate levels of susceptibility and resistance to the disease, such as the sweet oranges (*Citrus sinensis* L. Osbeck) ‘Bahia’, ‘Hamlin’, ‘Valencia’ and ‘Pera’, and the mandarins ‘Ponkan’ (*Citrus reticulata* Blanco) and ‘Satsuma’ (*Citrus unshiu* Marcovitch) (7, 8) (Figure 1). Therefore, the study of the biological processes of the plant–pathogen interaction at molecular level in such diverse citrus genotypes may help to better understand citrus canker disease and may allow identifying potential targets for disease management. The use of the

high-throughput RNA-Seq technology can be used to detect the induced or repressed genes in the host during the bacterial infection process. Then, the transcriptome can unveil the genetic and molecular mechanisms that confer the different levels of citrus susceptibility and resistance against the bacterial pathogen. However, current transcriptome-based studies did not include a large range of citrus cultivars exhibiting different levels of resistance and susceptibility to citrus canker. For instance, a previous study was carried out using microarray-based analysis in sweet oranges ‘Pera’ and ‘Cristal’ varieties (*C. sinensis*) and ‘Mexican’ lime (*C. aurantifolia*) and ‘Siciliano’ lemon (*C. limon* (L.) Burm. f.) (9). In another investigation, based on RNA-Seq analysis, only two cultivars, the canker-resistant ‘Meiwa’ kumquat (*Fortunella crassifolia* Swingle) and the canker-susceptible ‘Newhall navel sweet orange’ (*C. sinensis*) were compared (10). In addition, RNA-Seq technique was also used to develop a high confident reference of citrus transcriptome based on 12 *Citrus* species from all main phylogenetic groups (11).

In this study, we aimed to determine early genetic expression changes in citrus plants under the development of the citrus canker disease using RNA-Seq approach to sequence the expressed mRNA of eight different citrus genotypes inoculated with *X. citri* at early stages of the infection. All sequenced data were analysed to detect the differentially expressed genes between compared genotypes. Thus, we have developed a transcriptome knowledge base of distinct citrus cultivars, the CitrusKB, to facilitate further research on the biology of citrus. Despite other citrus databases, such as *Citrus sinensis* annotation project (12) and Citrus Genome Database (<https://www.citrusgenomedb.org>), CitrusKB is the only web tool that comes to integrate a wide range of citrus species and cultivars exhibiting different levels of resistance to citrus canker. The CitrusKB is a user-friendly web-based interface that provides tools to allow researchers to visualize, search, analyse, recover and browse information on citrus and *X. citri* interactome.

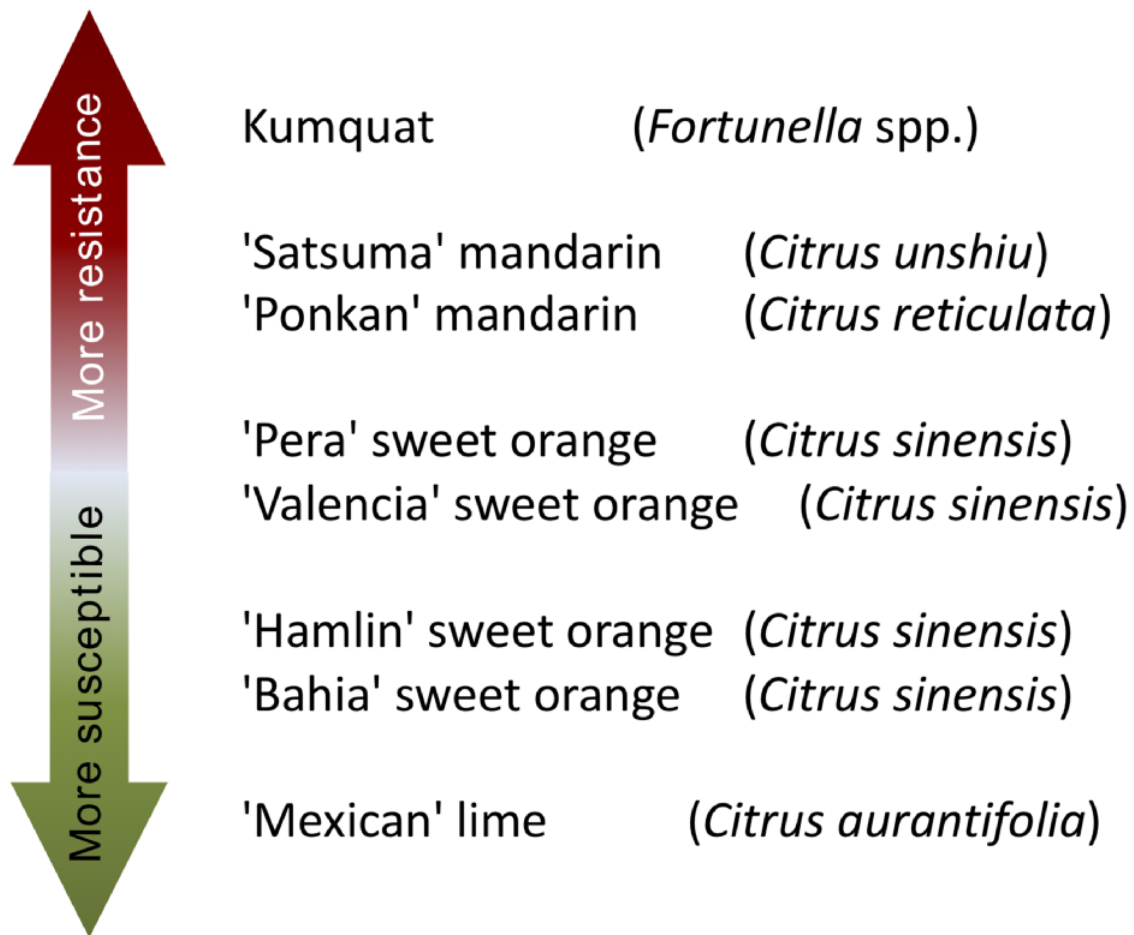


Figure 1. Citrus genotypes resistance and susceptibility scale to citrus canker A. 'More resistant' indicates that a citrus genotype is more resistant to citrus canker disease, as well as 'More susceptible' indicates that a citrus genotype is less resistant to citrus canker disease.

Construction and content

Currently, CitrusKB hosts a database to study the effects of initial stages of citrus canker disease including *in vivo* RNA-Seq data of 8 citrus genotypes, which have been prepared during 3 infection time points, at 24, 48 and 72 hours post-inoculation (hpi) of *X. citri*. In addition, the web tool includes the sequences of other citrus reference genomes that were available in other databases.

Inoculation of citrus genotypes with *X. citri*

The eight citrus genotypes exhibiting different levels of resistance to citrus canker included in the study were: 'Kumquat' (*Fortunella* spp.); Mandarins 'Ponkan' (*Citrus reticulata*) and 'Satsuma' (*Citrus unshiu*); sweet oranges (*Citrus sinensis*) 'Bahia', 'Valencia', 'Pera' and 'Hamlin' and 'Mexican' lime (*Citrus aurantifolia*). The trees were maintained in a greenhouse under controlled temperature (28°C) and photoperiod conditions (light/dark periods of 12 h). The trees were pruned 3 weeks prior to inoculation in order to produce uniform young leaves, with 75% of leaf

expansion (ideal condition for bacterial inoculation test in citrus (13)). The citrus trees were grafted on 'Rangpur' lime (*Citrus limonia* Osbeck). The *X. citri* strain 306 (14) was streaked from glycerol stock on Nutrient Agar (0.3% beef extract, 0.5% peptone, 1.5% agar) for 72 h at 28°C. A single colony was transferred to a new Nutrient Agar plate and after 72 h at 28°C, a bacterial sample was diluted into sterile bi-distilled water and the inoculum was prepared by dilution to 10⁸ cfu/ml (0.3 OD reading at 600 nm). The bacterial suspension was infiltrated into the abaxial surface of 5 leaves of each of the 3 plants in the entire leaf blade by using a 1 mL syringe without needle. Similarly, leaves of other three check plants were infiltrated with sterile distilled water (control).

Citrus RNA extraction and sequencing

Treated and control plants were kept in a growth chamber under controlled temperature (28°C) and photoperiod conditions (light/dark periods of 12 h). At 24, 48 and 72 h after inoculation, 3 leaves inoculated with *X. citri* and showing canker symptoms (infected - I) and 3 leaves inoculated with

sterile water and without canker symptoms (control - NI) were collected from each plant and immediately frozen in liquid nitrogen and further stored at -80°C .

For each time point, frozen leaf tissue was ground with mortar and pestle using liquid nitrogen. Total RNA was extracted from macerated leaf samples using Trizol (Invitrogen, Carlsbad, California, USA), according to the manufacturer's protocol. RNA quality and yield were determined by using Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA) and Qubit 2.0 Fluorometer (Invitrogen). Only RNA having integrity number (RIN) ≥ 8.0 were used. The cDNA libraries were multiplexed and then subjected to high-throughput RNA-Seq using Illumina HiScanSQ (Illumina Inc., San Diego, California, USA). The sequenced samples (8 genotypes, 6 libraries each: I and NI for each time) produced 48 raw RNA-Seq libraries, corresponding to more than 50 millions of 50 bp single-end reads per library.

Transcriptome analysis

The raw data of each library were filtered using the software NGS-QC Toolkit (15) to remove bad quality, adaptors and contaminated sequences. The software was configured with the following parameters: Primer/Adaptor library: Rapid Library (Standard); Homopolymer trimming: On; Length of the homopolymer to be removed: 8; Cut-off read length for HQ: 70%; Cut-off quality score: 20. The libraries were then aligned to the *Citrus sinensis* reference genome (16) using the Tophat2 (v2.1.14) (17) plus Cufflinks (v2.2.1) (18) softwares with default parameters. The identification of novel transcripts and isoforms was performed based on Cufflinks (v2.2.1) software (18), with default parameters, using the reference transcriptome annotation of *Citrus sinensis* reference genome (16). All identified genes, including annotated and novel isoforms, are available at 'Download' section of CitrusKB web tool. A total of 49 564 individual transcripts with an average transcript length of 1.5 kb achieved a Benchmarking Universal Single Copy Orthologs (BUSCO) completeness of $\sim 95\%$ with [S: 77.9%, D: 17.0%, F: 4.7%, M: 0.4%, n : 430, according to the viridiplantae OrthoDB release 10 (19), and thus validating the CitrusKB as a comprehensive citrus transcriptome resource. These data were used in the gene expression analysis. In contrast, the *C. sinensis* and *C. clementina* transcriptomes available at Phytozome database (20) achieved a similar BUSCO completeness values, i.e. C: 94.2% [S: 52.1%, D: 42.1%], F: 4.9%, M: 0.9%, n : 430, and C: 98.6% [S: 72.6%, D: 26.0%], F: 1.2%, M: 0.2%, n : 430, respectively.

Table 1. *Citrus sinensis* transcriptome: assembly and annotation status. Annotation type indicates the analysed characteristic of the transcriptome, Amount indicates a size or a number of molecules.

Annotation type	Amount
Total number of transcripts	49 564
Total size of transcripts	73 991 731 bp
GC%	41.06
Longest transcript	16 810 bp
Transcripts > 1 kb	29 599
Mean transcript size	1493 bp
Median transcript size	1241 bp
Blast2GO annotated transcripts	33 086
GO Terms	56 655
-Biological process	18 271
-Cellular component	12 090
-Molecular function	26 294
EggNOG annotated transcripts	35 330
-Unique GO terms	6265
-Unique KEEG Kos	3541
-Unique BiGG reactions	284

Transcriptome annotation

The assembled transcripts and multi-species citrus transcriptome were annotated with the tools InterProScan (v5.29.68.0) (21), eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) (v4.5.1) (22) and Blast2GO (v5.0) (23). The WEGO tool (24) was used to plot the GO annotations. Both the WEGO 'GO Annotation file format' (GAF) file and eggNOG annotation results and further analysis are available for download in the CitrusKB 'Download' section. An overview of the Citrus transcriptome Distribution of euKaryotic Orthologous Group (KOG or NOGs) and Gene Ontology annotation is also available for visualization (Table 1, Figure 2).

Gene expression analysis

For differential gene expression analysis (DEG), the high-quality reads used for transcriptome annotation step were mapped against the citrus reference genome using the STAR software (v2.4.1) with default parameters (25). Next, the absolute number of aligned reads was extracted for each individual transcript using the HTSeq software (v0.6.1) (26). Further, the normalized expression variation with statistical significance analysis for each individual transcript was calculated between samples. The False Discovery Rate (FDR) correction (27) over the statistical significance found between the samples was applied to control false-positive significance transcript expression variation. These analysis

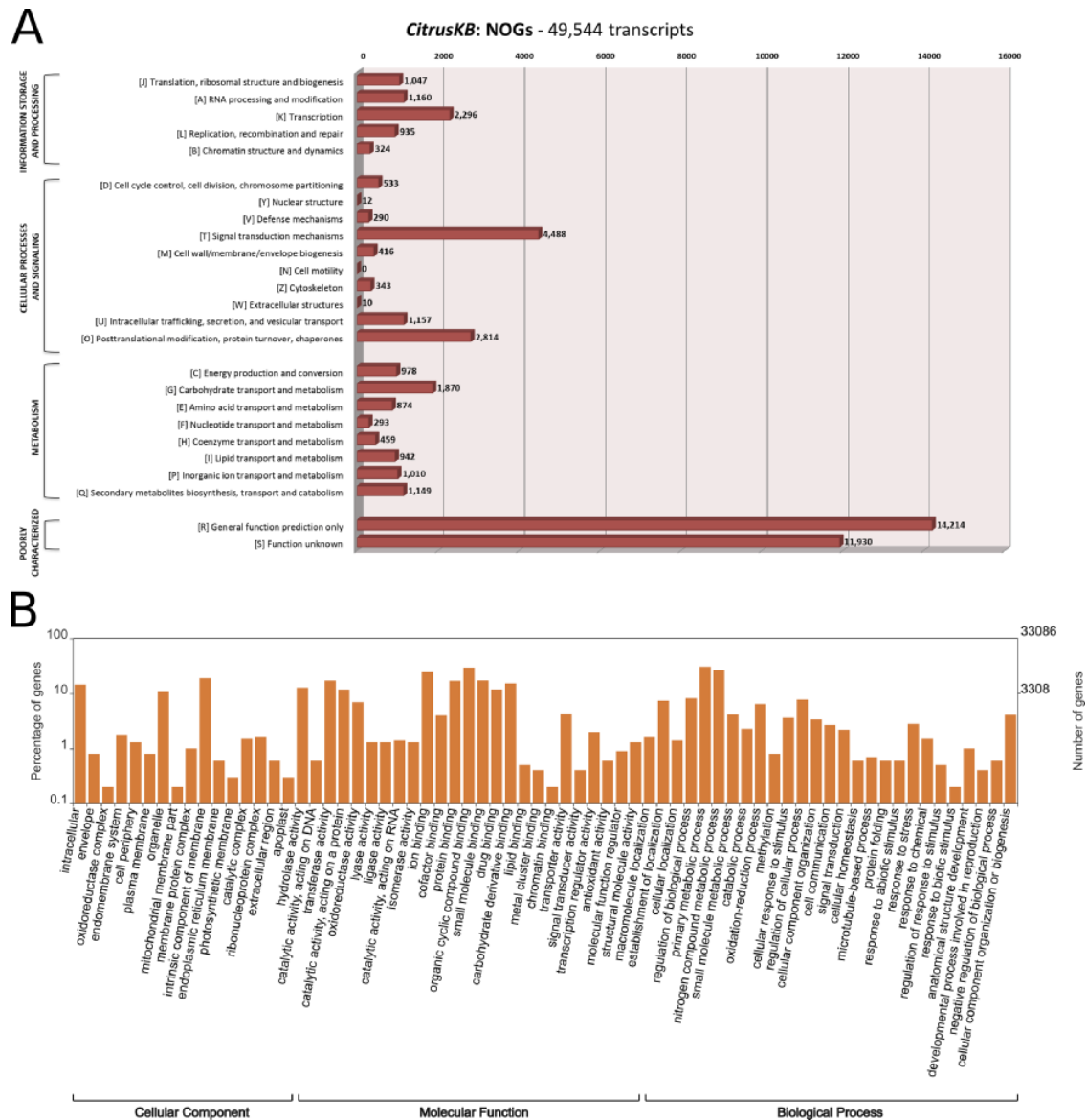


Figure 2. Annotation of *Citrus sinensis* CitrusKB transcriptome. EggNOG (A) and representative GO terms (B). Functionally annotated orthologous groups are represented in the Y axis of eggNOG annotation. For each annotated group, we have the identification of the functional category of the group (J, A, K, L, B, D, Y, V, T, M, N, Z, W, U, O, C, G, E, F, H, I, P, Q, R, S) followed by its description. The Y axis of GO annotations is presented in log (10) scale. Both, eggNOG and GO terms are fully searchable through the Gene expression tables.

were performed using DESeq, a Bioconductor R package (28). The transcripts were considered as differentially expressed when statistical significance (adjusted *P*-value) was less than 0.05.

Simple sequence repeats

The search for Simple Sequence Repeats (SSR) was performed using the TandemSmart software (version v1.0, unpublished). This approach performs custom SSR search using a reference sequence combined with high-throughput sequencing alignment data. This data is then used to validate detected SSRs and providing normalized occurrence

values in a RPKM-based calculation for each SSR marker. Higher RPKM-based values denote higher expressed markers. The detected markers through all citrus genotypes, stored in a PostgreSQL database, describe the following information: sample name for each SSR occurred, sequence ID, sequence motif, start and end coordinates for SSR in reference sequence, motif length, length of repeat and RPKM-based value.

Single nucleotide polymorphism

To detect single nucleotide polymorphisms (SNPs), RNA-Seq data were aligned to *Citrus sinensis* reference genome

CitrusKB

Knowledge Base for Transcriptome of Citrus Interactome



The figure displays the CitrusKB website interface, showing a navigation menu at the top and several tool screenshots below. The navigation menu includes: Home, Sequencing, Local blast, Transcriptome analysis, Download, Tools, and Contact. The tool screenshots include:

- Local Blast:** A form for selecting a program to use, inputting sequences, and performing a BLAST search.
- Gene Expression:** A table showing gene expression data for various samples and genes, with search parameters and a genome browser view.
- Create Venn Diagram:** A Venn diagram showing the overlap of genes across different samples, with a configuration panel for displaying counts and labels.
- Genetic Markers (SSR):** A table listing Simple Sequence Repeat (SSR) markers with columns for Sample, Sequence, Repeat, Position, Ref. Allel, and SNP.
- Genome Browser (Jbrowse):** A visualization of the genome browser showing gene models and alignments.
- Genetic Markers (SNP):** A table listing Single Nucleotide Polymorphism (SNP) markers with columns for Sample, Genomic Sequence, Mapped Transcript, Position, Ref. Allel, and SNP.
- Sequence Search:** A search interface for finding sequences in the transcriptome database.
- HeatMap Tool:** A heatmap visualization tool for gene expression data, with options for data file, plot type, and styling.

Figure 3. Citrus basic knowledge database (CitrusKB) website resources. Main functionalities are represented by the following screenshots: Local Blast (Local Blast navigation menu); Gene Expression, Genome Browser, Sequence Search, Genetic Markers (SSR) and Genetic Variants (SNP) (Transcriptome analysis navigation menu); transcriptome analysis tools Jvenn and Heatmapper (Tools navigation menu).

using the STAR aligner (25) with default parameters. Next, the genetic variants having expressed all alternate alleles were extracted using the SAMtools package (29). The data generated were stored within the database and are represented by the following attributes: sequence ID, position, reference allele and SNP.

Utility description

CitrusKB implementation

CitrusKB was developed on a web-based environment and is executed on a Linux operating system (Debian v7.7), which includes several common software packages: Apache HTTP server, PostgreSQL database, PHP, Java and Perl. The website (Figure 3) was developed using PHP, Java and HTML languages, and Twitter Bootstrap was used to achieve an enhanced user interaction. A Basic Local Alignment Search Tool (BLAST) (30) server is available in CitrusKB environment. As genome browser for transcriptome visualization, JBrowse (31) was implemented, showing all details of the transcripts and their alignments at nucleotide level.

Local blast

To provide an intuitive web-based graphical user interface and a rapidly search in a large volume of data available in the CitrusKB, we used the standalone BLAST server (v2.2.26) (32). To create the BLAST alignment database, we used the makeblastdb program of the NCBI BLAST (v2.2.30) software package (33). Using web forms, users can perform nucleotide searches using the BLASTN tool and protein searches, using BLASTP tool.

Sequence search

To perform searches for sequences from transcriptome, users can use as parameter a single or multiple sequence IDs, annotation keywords or gene names. The transcript sequence is retrieved by an ‘in-house’ script that performs the search over the stored transcriptome.

Gene expression

In this section, users can view the gene expression profile of the sequenced libraries. For each citrus cultivar, differential analysis was performed to determine expression

variation between pairs of conditions, including the following information: sequence ID, annotation, InterProScan (21) domains prediction, Fragments per Kilobase of Exon per Million Fragments Mapped per sample expression, Log_2 (Fold-change), P -value and adjusted P -value (FDR-corrected). They are shown in a table that allows users to dynamically sort them. In addition, users can perform custom searches based on attributes of the analysis and limit the search by applying specific minimum and maximum filtering criteria for: gene expression values, fold-change alteration, P -value and adjusted P -value. Filtered values can be exported to tables and users can download the results for further analysis.

Genome browser

CitrusKB uses JBrowse to dynamically browse and let users visualize, as multiple configurable tracks, the data corresponding to genome sequence, reference transcriptome, alignments information (sequencing coverage, splice junctions) and genetic variants (SNP, Insertions, Deletions). Available tracks are classified into three distinct categories, named as 'Alignments', 'Reference Sequence' and 'Variants'. At 'Reference Sequence' category, users can visualize a reference genome track and a reference transcriptome track that provides a visualization of annotated information and the gene exon-intron structure. Users can also examine the expression level, as well as sequencing depth and coverage of each gene at the 'Alignments' category, including details related to alignment such as splice junctions and mismatched positions. All the 48 tracks were created from BAM alignment files to display short RNA-Seq reads aligned to the citrus reference genome. Users can also visualize single nucleotide variants for all tracks into 'Variants' category. The information is showed by JBrowse from VCF files.

Genetic molecular markers

The SSR and SNP can be visualized by two distinct tables. In the first table, detected SSR in all samples are available for search. This table consists of sample ID, sequence ID, sequence, annotation, motif, start and end coordinates for SSR, motif and repeat length, and RPKM value. In the second table, SNPs are listed with the following attributes: sample ID, sequence name (scaffold from reference *C. sinensis* genome), position, reference allele and SNP. Both tables allow to dynamically sort the information, and the user also can configure the search by choosing a specific citrus sample.

Tools page

To facilitate the interpretation and to visualize the transcriptome annotation and differential expression analysis results, CitrusKB provides two different third-part commonly used tools for transcriptome analysis: (a) Venn diagrams, based on jvenn (34), and (b) HeatMaps, implemented by the Heatmapper software (35). Further details on how to use these tools are shown in the topic 'Using CitrusKB: a study case'.

Download page

CitrusKB provides a user-friendly interface for the download of entire data sets. The HTTP links provide the download of all RNA-Seq raw sequencing reads from the 48 libraries, available in FASTQ format and compressed as Gzipped files. Users can also download the following data: reference genome, assembled reference transcriptome sequences and gene structure in GFF format, all the gene expression analysis, and other details of SSRs and SNPs genetic markers. The raw data are also available under BioProject number PRJNA470961.

Using CitrusKB and discussion

The following section summarizes one example of CitrusKB usability. It is based on a study of differentially expressed genes (DEGs), revealing important transcripts found in the spatial and temporal citrus and *X. citri* interactome. The used thresholds during transcript search in the CitrusKB gene expression tables to consider a DEG were set as: fold-change ≥ 2 for up-regulated gene, fold-change ≤ -2 for down-regulated gene, both with an adjusted P -value ≤ 0.05 (FDR-corrected). The resulting tables were exported in the 'TSV' file format ('export data' button) and directly imported in the CitrusKB Venn Diagram Tool ('Tools' tab). The GAF file available in the 'download section' of the CitrusKB was used for the exploration of the enriched GO terms, with the use of the WEGO tool. In addition, the TSV file can be also imported in the CitrusKB HeatMap Tool ('Tools' tab) and interactively inspected in the form of heat maps. The annotation products, Gene Ontology terms and IDs, predicted NOGs and enzyme code of each DEG transcript can be retrieved through the gene expression tables or exported in the TSV file that can be further explored for other biological interpretations and/or downstream analysis.

It is worth to mention that the following study consist in a general example of the application of the CitrusKB which was not intended to show a comprehensive transcriptome analysis, but to illustrate the database use, the built-in

tools to facilitate the transcriptome data integration, visualization and comparison. The objective to include this example was to present the power of the CitrusKB to reveal the plant–pathogen related gene targets for further and more advanced studies. Researchers working with different plant–pathogen interaction topics can raise specific questions using the CitrusKB gene expression tables search engine and genome browser.

A case of study: global overview of the DEG of the *Citrus* and *X. citri* interactome with identification of transcripts potentially related to plant–pathogen interactions

The analysis of the predicted DEGs of eight citrus genotypes in three-time stages (24, 48, and 72 h) revealed a plethora of DEG transcripts which can be strictly related with plant pathogen interaction, such as defense mechanisms and resistance and susceptibility to *X. citri*. In this analysis, the eight genotypes were grouped in five clusters: the first contains the highly resistant ‘Kumquat’ genotype; the second contains the resistant Mandarins ‘Ponkan’ and ‘Satsuma’; the third with the less-resistant sweet oranges ‘Pera’ and ‘Valencia’; the fourth with the more susceptible genotypes ‘Hamlin’ and ‘Bahia’, and the fifth with the highly susceptible ‘Mexican’ lime genotype. In summary, ‘Kumquat’, ‘Ponkan’ and ‘Satsuma’, ‘Hamlin’ and ‘Bahia’ stand as the genotypes with the larger number of up-regulated transcripts (average of 2350), whereas

‘Kumquat’ exhibits the largest number of down-regulated transcripts (1602) (Figure 4). The ‘Kumquat’ genotype also exhibits the largest number of exclusive up-regulated (961) and down-regulated transcripts (1602). The list of the transcripts used to generate the Venn diagrams is provided as supplementary material (Table S1) for further exploitation.

In general, the presented analysis also revealed 3827 (7%) up-regulated and 4889 (9%) down-regulated unique DEGs, and thus representing an overall pattern of *C. sinensis* response to *X. citri* infection. Among those, 642 up-regulated and 782 down-regulated transcripts correspond to hypothetical or uncharacterized proteins, and thus, may be considered as potential new targets for functional studies of plant–pathogen interaction mechanisms (Table S2). In addition, 2307 up-regulated and 2743 down-regulated transcripts are classified in NOGs (Table 2). The most prevalent DEGs NOGs belongs to ‘transcription’, ‘signal transduction mechanisms’, ‘post-translational modification, protein turnover, chaperones’, ‘carbohydrate transport and metabolism’ and ‘secondary metabolites biosynthesis, transport and catabolism’. There are more than 40% of up-regulated DEGs related to ‘replication, recombination and repair’, and ‘cytoskeleton’ NOGs categories in comparison to down-regulated DEGs. In contrast, much more enriched down-regulated DEG were identified, such as those related to ‘transcription’, ‘cell wall/membrane/envelope biogenesis’, ‘carbohydrate transport and metabolism’, ‘coenzyme transport and

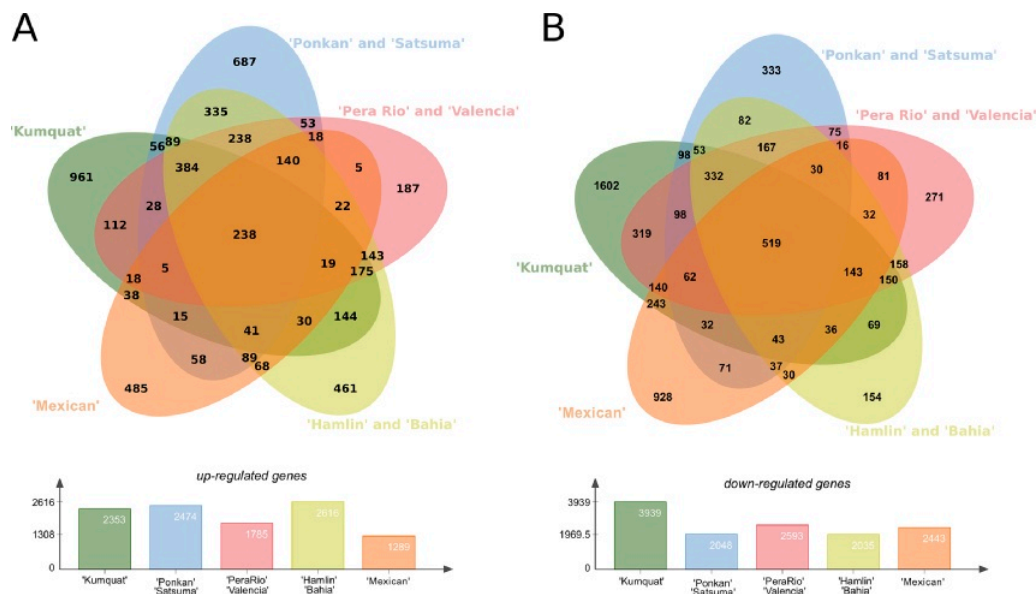


Figure 4. Differentially expressed genes (DEGs) present in the eight citrus genotypes considering the three analysed time stages (24, 48 and 72 h) post *X. citri* inoculation. (A) Upper panel, Venn diagram with the Up-regulated transcripts; Lower panel, total number of DEGs for each analysed genotype group. (B) Upper panel, venn diagram with the Down-regulated transcripts; Lower panel, total number of DEGs for each analysed genotype group.

Table 2. EggNOG classification of up- and down-regulated DEGs identified among the eight citrus genotypes in the all-time stages. Transcripts with only one NOG were considered (Table S1 for a comprehensive list of the transcripts and annotation).

	EggNOG category	Up-regulated transcripts	Down-regulated transcripts
Information Storage and Processing	[J] Translation, ribosomal structure and biogenesis	91	96
	[A] RNA processing and modification	28	53
	[K] Transcription	179	261
	[L] Replication, recombination and repair	72	22
	[B] Chromatin structure and dynamics	7	7
Cellular Processes and Signaling	[D] Cell cycle control, cell division, chromosome partitioning	42	26
	[V] Defense mechanisms	36	33
	[T] Signal transduction mechanisms	552	566
	[M] Cell wall/membrane/envelope biogenesis	24	69
	[Z] Cytoskeleton	54	27
	[W] Extracellular structures	2	2
	[U] Intracellular trafficking, secretion and vesicular transport	59	55
Metabolism	[O] Posttranslational modification, protein turnover, chaperones	271	280
	[C] Energy production and conversion	92	112
	[G] Carbohydrate transport and metabolism	184	283
	[E] Amino acid transport and metabolism	101	99
	[F] Nucleotide transport and metabolism	30	26
	[H] Coenzyme transport and metabolism	23	88
	[I] Lipid transport and metabolism	87	115
	[P] Inorganic ion transport and metabolism	119	170
	[Q] Secondary metabolites biosynthesis, transport and catabolism	141	245

metabolism’, ‘inorganic ion transport and metabolism’ and ‘secondary metabolites biosynthesis, transport and catabolism’.

These findings may suggest for a global set of genes potentially related to the plant response against the pathogen infection. For instance, among the identified NOGs (Figure 2A–2B), 440 out of 2296 (19%), 1118 out of 4488 (24%) and 551 out of 2814 (19%) represent DEGs related to ‘transcription’, ‘signal transduction mechanisms’, and ‘posttranslational modification, protein turnover, chaperones’, respectively, which may regulate the citrus metabolism, through the inactivation and activation of genes in a putative attempt to control the infection caused by the pathogen. Conversely, 69 out of 290 (23%) DEGs related to ‘defense mechanisms’ were also identified, revealing several transcripts with detoxification related functions (Figure 2A, Table S2) which may be involved in the scavenge of the reactive oxygen species (ROS) generated by aerobic metabolism, and also by other processes, such as growth development, programmed cell death and response to biotic and abiotic environmental stimuli. All these processes can be linked to *X. citri* infection. Overall, these results summarized the use of CitrusKB to identify several transcripts (Table S2) with potential correlation to plant–pathogen interactions.

Conclusions

The knowledge base for transcriptome of Citrus interactome, CitrusKB, was created as a web tool to researches on citrus biology, in particular to the genes involved on plant–pathogen interaction. The CitrusKB also provides tools on a user-friendly web interface that allows researchers to visualize, search, analyse and browse information regarding citrus and their interaction with the citrus canker pathogen *X. citri*. To the best of our knowledge, this is the first *in vivo* database for analysis of citrus infected by a pathogen, and we expect that it will bring a substantial contribution for researches in the area, particularly in the field of molecular plant–pathogen interactions. We will continue updating CitrusKB by incorporating novel literature data as they get available over time.

Acknowledgements

We thank Applied Bioinformatics Laboratory (LBA) from Brazilian Agricultural Research Corporation (Embrapa/CNPq, Campinas, SP, Brazil) by providing computational and bioinformatics software support for data analysis. We also thank Fundo de Defesa da Citricultura (FUNDECITRUS) for providing all the plants used in this study.

Funding

JAF is recipient of Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), [Grant # 307986/2012-8] and from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES – BIOCUMPUTACIONAL), [Grant # 3371/2013]. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. AF is PhD fellowship from CAPES. MMM was recipient of a Master (MS) fellowship from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), [Grant # 2011/05477-4] and TDCGC, JSC, and MHM] were recipient of Master (MS) fellowships from CAPES. HAP was recipient of PhD fellowship from CAPES. GP was a recipient of a scientific initiation scholarship from PIBIC/CNPq/UNESP program. RHH is recipient of Fundação Araucária (FA) [Grant # CP09/2016].

Availability of data and material

The datasets generated and/or analysed during the current study are available at the institutional website: <http://bioinfo.deinfo.uepg.br/citrus>.

Conflict of interest None declared.

Authors' contributions

AF, JBJ, HF, MITF, RPLJ, AMV, RHH and JAF designed the study (data generation and web tool); FMSC, MMM, TDCGC, JSC, GP and JB prepared and generated the sequencing data; AF, MITF, HAP, AMV, RHH and JAF processed, analysed and annotated the data; AF, HAP, AMV, RHH, JB and JAF wrote and revised the manuscript.

Supplementary material

[Supplementary data](#) are available at *Database* Online.

References

- Li, W., Song, Q., Bransky, R.H. *et al.* (2007) Genetic diversity of citrus bacterial canker pathogens preserved in herbarium specimens. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 18427–18432.
- Talon, M. and Gmitter, F.G. (2008) Citrus genomics. *Int. J. Plant Genomics*, **2008**, 528361.
- Civerolo, E.L. (1984) Bacterial canker disease of citrus. *J. Rio Gd. Val. Hort. Soc.*, **37**, 127–146.
- Rossetti, V.V. (2001) Manual Ilustrado de Doenças dos Citros. Fealq/Fundecitrus, Piracicaba, 2001.
- Behlau, F., Amorim, L., Belasque, J. *et al.* (2010) Annual and polyetic progression of citrus canker on trees protected with copper sprays. *Plant Pathol.*, **59**, 1031–1036.
- Barbosa, J.C., Gimenes-Fernandes, N., Massari, C.A., Ayres, A. *et al.* (2001) Incidência e distribuição de cancro cítrico em pomares comerciais Estado de São Paulo e sul do Triângulo Mineiro. *Summa Phytopathol.*, **27**, 30–35.
- de Carvalho, S.A., de Carvalho Nunes, W.M., Belasque, J. *et al.* (2015) Comparison of resistance to Asiatic citrus canker among different genotypes of *Citrus* in a long-term canker-resistance field screening experiment in Brazil. *Plant Dis.*, **99**, 207–218.
- Khalaf, A., Moore, G.A., Jones, J.B. *et al.* (2007) New insights into the resistance of Nagami kumquat to canker disease. *Physiol. Mol. Plant Pathol.*, **71**, 240–250.
- Cernadas, R.A., Camillo, L.R. and Benedetti, C.E. (2008) Transcriptional analysis of the sweet orange interaction with the citrus canker pathogens *Xanthomonas axonopodis* pv. *citri* and *Xanthomonas axonopodis* pv. *aurantifolii*. *Mol. Plant Pathol.*, **9**, 609–631.
- Fu, X.Z., Gong, X.Q., Zhang, Y.X. *et al.* (2012) Different transcriptional response to *Xanthomonas citri* subsp. *citri* between kumquat and sweet orange with contrasting canker tolerance. *PLoS One*, **7**(7), 1–16.
- Terol, J., Tadeo, F., Ventimilla, D. *et al.* (2016) An RNA-Seq-based reference transcriptome for citrus. *Plant Biotechnol. J.*, **14**, 938–950.
- Wang, J., Chen, D., Lei, Y. *et al.* (2014) Citrus sinensis Annotation Project (CAP): a comprehensive database for sweet orange genome. *PLoS One*, **9**(1), 1–7.
- Murata, M.M., Omar, A.A., Mou, Z. *et al.* (2018) Novel plastid-nuclear genome combinations enhance resistance to citrus canker in cybrid grapefruit. *Front. Plant Sci.*, **9**, 1858.
- da Silva, A.C.R., Ferro, J.A., Reinach, F.C. *et al.* (2002) Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature*, **417**, 459–463.
- Patel, R.K. and Jain, M. (2012) NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE*, **7**(2), 1–7.
- Wu, G.A., Prochnik, S., Jenkins, 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.
- Kim, D., Pertea, G., Trapnell, C. *et al.* (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36.
- Trapnell, C., Williams, B.A., Pertea, G. *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.
- Seppy, M., Manni, M. and Zdobnov, E.M. (2019) BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol.*, **1962**, 227–245.
- Goodstein, D.M., Shu, S., Howson, R. *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.*, **40**, D1178.
- Quevillon, E., Silventoinen, V., Pillai, S. *et al.* (2005) InterProScan: protein domains identifier. *Nucleic Acids Res.*, **33**, W116–W120.
- Jensen, L.J., Julien, P., Kuhn, M. *et al.* (2007) eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.*, **36**, D250–D254.
- Conesa, A. and Götz, S. (2008) Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics*, **2008**, 1–12.

24. Ye, J., Fang, L., Zheng, H. *et al.* (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.*, **34**, W293.
25. Dobin, A., Davis, C.A., Schlesinger, F. *et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
26. Anders, S., Pyl, P.T., and Huber, W. (2015) HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**(2), 166–169.
27. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.*, **57**, 289–300.
28. Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
29. Li, H., Handsaker, B., Wysoker, A. *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
30. Altschul, S.F., Gish, W., Miller, W. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
31. Buels, R., Yao, E., Diesh, C.M. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
32. Johnson, M., Zaretskaya, I., Raytselis, Y. *et al.* (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.*, **36**, W5–9.
33. Camacho, C., Coulouris, G., Avagyan, V. *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinform.*, **10**, 421.
34. Bardou, P., Mariette, J., Escudié, F. *et al.* (2014) jvenn: an interactive Venn diagram viewer. *BMC Bioinform.*, **15**, 293.
35. Babicki, S., Arndt, D., Marcu, A. *et al.* (2016) Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res.*, **44**, W147–W153.