Genome sequencing and analysis of *Ralstonia solanacearum* phylotype I strains FJAT-91, FJAT-452 and FJAT-462 isolated from tomato, eggplant, and chili pepper in China

Yidan Sun¹,²†, Keke Wang¹†, Carlos Caceres-Moreno¹,²†, Wei Jia¹, Aojun Chen¹, Heng Zhang¹, Renyi Liu¹* and Alberto P. Macho¹*

Abstract

*Ralstonia solanacearum* is an extremely destructive pathogen able to cause disease in a wide range of host plants. Here we report the draft genome sequences of the strains FJAT-91, FJAT-452 and FJAT-462, isolated from tomato, eggplant, and chili pepper, respectively, in China. In addition to the genome annotation, we performed a search for type-III secreted effectors in these strains, providing a detailed annotation of their presence and distinctive features compared to the effector repertoire of the reference phylotype I strain (GMI1000). In this analysis, we found that each strain has a unique effector repertoire, encoding both strain-specific effector variants and variations shared among all three strains. Our study, based on strains isolated from different hosts within the same geographical location, provides insight into effector repertoires sufficient to cause disease in different hosts, and may contribute to the identification of host specificity determinants for *R. solanacearum*.

Keywords: *Ralstonia*, Virulence, Effector, Tomato, Eggplant, Chili pepper, China

Introduction

*Ralstonia solanacearum* is often considered one of the most destructive bacterial pathogens, causing bacterial wilt disease in more than 250 plant species worldwide [1]. The pathogenicity of *R. solanacearum* heavily relies on the injection of proteins inside plant cells through a type-III secretion system (T3SS). The versatility of *R. solanacearum* strains correlates with the presence of a larger number of T3SS substrates, called type-III effectors (T3Es), encoded in their genomes, in comparison to other bacterial pathogens [2]. T3Es are important virulence factors required by most gram-negative pathogens to manipulate plant cells and cause disease [3, 4]. Bacteria from a single *R. solanacearum* strain can inject more than 70 T3Es (termed Rips for *Ralstonia* injected proteins) inside plant cells [2, 5]. Studies conducted in *Pseudomonas syringae* and *Xanthomonas axonopodis* strains indicate that T3E repertoires are highly variable among strains of these species, and led to the hypothesis that T3E composition may shape the host range of bacterial pathogens [6, 7]. Although the genome sequences and T3E repertoires have been defined for several *R. solanacearum* strains, repertoire comparisons have failed in identifying host specificity determinants so far [2], which may suggest that genome sequences from additional strains infecting different hosts are required for this analysis. Additionally, the diversity in the geographical origins of sequenced strains hinders this comparative analysis, since additional environmental factors, such as temperature, light, and humidity may have a significant impact on the requirement of effectors for a successful infection. In this project, we sequenced and annotated the...
genomes of the *R. solanacearum* strains FJAT-91, FJAT-452 and FJAT-462, isolated from tomato (*Solanum lycopersicum*), eggplant (*Solanum melongena*), and chili pepper (*Capsicum annum*), respectively, in the Fujian province (China) [8]. In addition, we performed a search for T3Es in these strains, providing a detailed annotation of their presence and distinctive features compared to the effector repertoire of the reference phylotype I strain (GMI1000). To our knowledge, this is the first report of genome sequences combined with T3E repertoire analysis performed in strains isolated from different hosts with the same geographical origin.

**Organism information**

**Classification and features**

*Ralstonia solanacearum* belongs to the order *Burkholderiales* of the class *Betaproteobacteria*. It is an aerobic, Gram-negative bacterium, naturally present in soil, water, infected plants or plant debris. It has a worldwide distribution, with higher incidence in tropical and subtropical regions, but also present in other temperate areas [9]. *R. solanacearum* is the agent causing bacterial wilt disease in multiple host plants, characterized by a sudden wilt of the whole plant. The strains sequenced in this study, FJAT-91, FJAT-452 and FJAT-462, were isolated from naturally infected tomato (*Solanum lycopersicum*), eggplant (*Solanum melongena*), and chili pepper (*Capsicum annum*) plants, respectively, in the Fujian province (China). Plants showing typical wilting symptoms were collected, surface-sterilized, and the tissue was homogenized with sterile water before plating serial dilutions to determine the causal agent [8, 9]. Sequence analysis determined that they belong to the *R. solanacearum* species complex [8]. The pathogenicity of FJAT-91 has been confirmed and used as positive control for pathogenicity assays in tomato plants in previous studies [10]. All three isolated strains displayed the typical physiological features of strains from the *R. solanacearum* species complex, showing aerobic growth in laboratory conditions, and were able to form 3–4 mm colonies within 2 days at 28 °C when grown on a rich laboratory medium containing tetrazolium chloride and high glucose content. For all three strains, colony shape was irregular, mucoid, and displayed a pink area in the middle of the colony and a large white edge (Fig. 1). Gene sequence analysis of PCR-amplified *fliC*, *hrpB* and *pehA* genes indicated that these strains belong to the phylotype I (represented by the reference strain GMI1000; Fig. 2), mostly formed by Asian strains [2]. The classification and general features of the three strains are summarized in the Tables 1, 2 and 3, and a phylogenetic tree is shown in the Fig. 2.

**Genome sequencing information**

**Genome project history**

This sequencing project was started in 2015, assembly and annotation was performed in 2016. Assembled draft genome sequences for the strains FJAT-91, FJAT-452 and FJAT-462 have been deposited to GenBank (Table 4).
Raw genomic reads have been deposited to the Sequence Read Archive with accession numbers SRP091690, SRR4431158, SRR4431159, SRR4428740.

Growth conditions and genomic DNA preparation

*R. solanacearum* strains were grown in rich medium (10 g/l bactopeptone, 1 g/l yeast extract and 1 g/l casamino acids). Genomic DNA was extracted from bacterial cultures grown to stationary phase for 18 h at 28 °C and shaking at 220 rpm (OD 600 = 1) using the Blood & Cell Culture DNA Mini kit (Qiagen), following manufacturer’s instructions for gram-negative bacteria. DNA concentration and quality were measured using a Qubit 2.0 Fluorometer (Invitrogen).

Genome sequencing and assembly

For each genome, we prepared a paired-end library with an average insert size of 470 bp and sequenced the library for 250 bp from both ends using Illumina HiSeq 2500. The number of raw read bases was greater than 300 million (>50x genome coverage) for each sequenced strain. The raw sequencing data were first preprocessed to remove adapter sequences, low-quality regions, and short sequences (less than 20 nucleotides) with Cutadapt [11] and SolexaQA [12]. The remaining clean reads were de novo assembled into contigs and scaffolds by using SOAPdenovo2 and GapCloser v1.12 [13]. Contigs and scaffolds were further assembled into chromosome, plasmid and scaffolds with CONTIGuator, using the GMI1000 genome as the reference. The resulting FJAT-91, FJAT-452 and FJAT-462 genomes are 4,620,128 bp, 5,334,434 bp and 5,083,617 bp, respectively (Table 5), close to the genome length of the *R. solanacearum* reference strain GMI1000 (5,810,922 bp) [14].

Table 1 Classification and general features of *Ralstonia solanacearum* FJAT-91 strain according to the MIGS recommendations [29]

| MIGS ID | Property      | Term                | Evidence code
|---------|---------------|---------------------|------------------|
|         | Classification| Domain Bacteria      | TAS [30]         |
|         | Phylum        | Proteobacteria       | TAS [31]         |
|         | Class         | Betaproteobacteria   | TAS [32, 33]     |
|         | Order         | Burkholderiales      | TAS [32, 33]     |
|         | Family        | Burkholderiaceae     | TAS [32, 33]     |
|         | Genus         | Ralstonia            | TAS [34, 35]     |
|         | Species       | *Ralstonia solanacearum* | TAS [34, 35] |
|         | Strain        | FJAT-91              |                  |
|         | Gram stain    | Negative             | IDA              |
|         | Cell shape    | Rod                  | IDA              |
|         | Motility      | Motile               | IDA              |
|         | Sporulation   | Non sporulating      | NAS              |
|         | Temperature   | Mesophile            | IDA              |
|         | Optimum       | 27 °C                | IDA              |
|         | pH range; Optimum | 5.5–8.0; 6.5  | NAS              |
|         | Carbon source | Dextrose, lactose, maltose, cellulose | IDA |

**Fig. 2** Phylogenetic tree showing the position of the *Ralstonia solanacearum* sequenced in this study, relative to other sequenced strains from the same species. The phylogenetic tree was constructed using concatenated alignments of the marker genes *flic*, *hrpB* and *pehA*. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model [27]. Evolutionary analyses were conducted in MEGA7 [28]. GenBank accession numbers are displayed within brackets. *Ralstonia pickettii* 12 J was used as an outgroup.

*Genome annotation*

Genome annotation was performed using Prokka (v1.11) [15] with the option for non-coding RNA (ncRNA) search. The COG database [16] and Pfam v30.0 [17] were used for functional annotation of genes. T3Es in the three newly sequenced strains were identified and...
annotated in two steps: first, 52, 62 and 60 of the T3Es from the *R. solanacearum* species complex [2] were identified in FJAT-91, FJAT-452 and FJAT-462, respectively, based on Prokka annotations; second, known T3Es protein sequences [2] were used as query to search the assembled genome sequences of three strains using BLAST [18] with a stringent significance cutoff of e-value $< 1e^{-30}$, identity $> 60$, and coverage on the query T3E protein sequence being over 50% or at least 100 aa in length. As a result, 72, 78 and 75 T3Es were identified in FJAT-91, FJAT-452 and FJAT-462, respectively. These two sets of T3E genes were merged together to generate the final lists of T3E genes in the three genomes. To identify the sequence variations within T3E genes between three strains and the reference strain, the clean reads from the three newly

| Table 2 | Classification and general features of *Ralstonia solanacearum* FJAT-452 strain according to the MIGS recommendations [29] |
|---------|-------------------------------------------------------------------------------------------------------------------------------|
| MIGS ID | Property Term                                                                                                                  | Evidence code  |
|---------|-------------------------------------------------------------------------------------------------------------------------------|----------------|
| Classification | Domain Bacteria                                                                                                               | TAS [30] |
|          | Phylum Proteobacteria                                                                                                          | TAS [31] |
|          | Class Betaproteobacteria                                                                                                        | TAS [32, 33] |
|          | Order Burkholderiales                                                                                                          | TAS [32, 33] |
|          | Family Burkholderiaceae                                                                                                         | TAS [32, 33] |
|          | Genus Ralstonia                                                                                                                | TAS [34, 35] |
|          | Species Ralstonia solanacearum                                                                                                  | TAS [34, 35] |
|          | Strain: FJAT-452                                                                                                               |                |
| Gram stain | Negative                                                                                                                      | IDA |
| Cell shape | Rod                                                                                                                           | IDA |
| Motility   | Motile                                                                                                                        | IDA |
| Sporulation | Non sporulating                                                                                                              | NAS |
| Temperature range | Mesophile                                                                                                                    | IDA |
| Optimum temperature | 27 °C                                                                                                                        | IDA |
| pH range; Optimum | 5.5–8.0; 6.5                                                                                                                  | NAS |
| Carbon source | Dextrose, lactose, maltose, celloobiene                                                                                      | IDA |
| MIGS-6 Habitat | Eggplants. Soil.                                                                                                              | TAS [8, 10] |
| MIGS-6.3 Salinity | $<2.0\%$                                                                                                                     | TAS [36] |
| MIGS-22 Oxygen requirement | Aerobic                                                                                                                      | IDA |
| MIGS-15 Biotic relationship | free-living                                                                                                                  | IDA |
| MIGS-14 Pathogenicity | Pathogenic                                                                                                                   | TAS [8, 10] |
| MIGS-4 Geographic location | Ningde, Fujian, China.                                                                                                       | TAS [8] |
| MIGS-5 Sample collection | 2011                                                                                                                        | TAS [8] |
| MIGS-4.1 Latitude | 26°39'56.22"N                                                                                                                   | NAS |
| MIGS-4.2 Longitude | 119°32'52.56"E                                                                                                                   | NAS |
| MIGS-4.4 Altitude | 2.89 m                                                                                                                        | NAS |

*Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [37]*

| Table 3 | Classification and general features of *Ralstonia solanacearum* FJAT-462 strain according to the MIGS recommendations [29] |
|---------|-------------------------------------------------------------------------------------------------------------------------------|
| MIGS ID | Property Term                                                                                                                  | Evidence code  |
|---------|-------------------------------------------------------------------------------------------------------------------------------|----------------|
| Classification | Domain Bacteria                                                                                                               | TAS [30] |
|          | Phylum Proteobacteria                                                                                                          | TAS [31] |
|          | Class Betaproteobacteria                                                                                                        | TAS [32, 33] |
|          | Order Burkholderiales                                                                                                          | TAS [32, 33] |
|          | Family Burkholderiaceae                                                                                                         | TAS [32, 33] |
|          | Genus Ralstonia                                                                                                                | TAS [34, 35] |
|          | Species Ralstonia solanacearum                                                                                                  | TAS [34, 35] |
|          | Strain: FJAT-462                                                                                                               |                |
| Gram stain | Negative                                                                                                                      | IDA |
| Cell shape | Rod                                                                                                                           | IDA |
| Motility   | Motile                                                                                                                        | IDA |
| Sporulation | Non sporulating                                                                                                              | NAS |
| Temperature range | Mesophile                                                                                                                    | IDA |
| Optimum temperature | 27 °C                                                                                                                        | IDA |
| pH range; Optimum | 5.5–8.0; 6.5                                                                                                                  | NAS |
| Carbon source | Dextrose, lactose, maltose, celloobiene                                                                                      | IDA |
| MIGS-6 Habitat | Chili pepper plants. Soil.                                                                                                    | TAS [8, 10] |
| MIGS-6.3 Salinity | $<2.0\%$                                                                                                                     | TAS [36] |
| MIGS-22 Oxygen requirement | Aerobic                                                                                                                      | IDA |
| MIGS-15 Biotic relationship | free-living                                                                                                                  | IDA |
| MIGS-14 Pathogenicity | Pathogenic                                                                                                                   | TAS [8, 10] |
| MIGS-4 Geographic location | Ningde, Fujian, China.                                                                                                       | TAS [8] |
| MIGS-5 Sample collection | 2011                                                                                                                        | TAS [8] |
| MIGS-4.1 Latitude | 26°39'56.22"N                                                                                                                   | NAS |
| MIGS-4.2 Longitude | 119°32'52.56"E                                                                                                                  | NAS |
| MIGS-4.4 Altitude | 2.89 m                                                                                                                        | NAS |

*Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [37]*
sequenced strains were mapped to the reference genome GMI1000 using BWA (v0.7.12) [19]. SNPs and INDELs were identified using Samtools (v0.1.19) [20] and vcftools (v0.1.12) [21] and were further annotated using SnpEff (v4.0) [22].

**Genome properties**
The genome of *R. solanacearum* strain FJAT-91 has 329 scaffolds and the average GC content of the genome is 60.6% (Table 5). A total of 6,522 genes (6457 CDSs and 65 ncRNAs) were predicted. Of the protein-coding genes, 2544 (39.4%) had functions assigned while 3913 were considered hypothetical (Table 5). 42.03% of the CDSs could be assigned to one COG functional category and 36.56% contained one or more conserved PFAM-A domains (Table 6). The genome of *R. solanacearum* strain FJAT-452 has 309 scaffolds and the average GC content of the genome is 62.33% (Table 5). A total of 6729 genes (6658 CDSs and 71 ncRNAs) were predicted. Of the protein-coding genes, 3075 (46.19%) had functions assigned while 2544 (39.4%) had functions assigned while 3913 were considered hypothetical (Table 5).

### Table 4 Project information

| MIGS ID | Property                  | FJAT-91  | FJAT-452 | FJAT-462 |
|---------|---------------------------|----------|----------|----------|
| MIGS 31 | Finishing quality         | Draft    | Draft    | Draft    |
| MIGS 28 | Libraries used            | Vazyme   | Vazyme   | Vazyme   |
| MIGS 29 | Sequencing platforms      | Illumina | Illumina | Illumina |
| MIGS 31.2| Fold coverage            | >50X     | >50X     | >50X     |
| MIGS 30 | Assemblers                | SOAPdenovo2; GapCloser v1.12; CONTIGuator | SOAPdenovo2; GapCloser v1.12; CONTIGuator | SOAPdenovo2; GapCloser v1.12; CONTIGuator |
| MIGS 32 | Gene calling method       | Prokka v1.11 (ncRNAs search enabled) | Prokka v1.11 (ncRNAs search enabled) | Prokka v1.11 (ncRNAs search enabled) |
| Locus Tag | NA                      | NA       | NA       |
| Genbank ID      | MLYU000000000 | MLYS000000000 | MLYT000000000 |
| GenBank Date of Release | December 25, 2016 | December 25, 2016 | December 25, 2016 |
| GOLD ID          | NA                      | NA       | NA       |
| BIOPROJECT       | PRJNA347535             | PRJNA347535 | PRJNA347535 |
| MIGS 13 | Source Material Identifier | SAMN05892025 | SAMN05892026 | SAMN05892027 |
| Project relevance | Plant pathogen         | Plant pathogen | Plant pathogen |

### Table 5 Genome and annotation statistics of the three newly sequenced *Ralstonia solanacearum* strains

| Attribute                      | FJAT-91 | % of total | FJAT-452 | % of total | FJAT-462 | % of total |
|--------------------------------|---------|------------|----------|------------|----------|------------|
| Genome size (bp)               | 4,620,128| 100.00     | 5,334,434| 100.00     | 5,083,617| 100.00     |
| DNA coding (bp)                | 3,003,037| 65.00      | 3,696,229| 69.29      | 3,397,556| 66.83      |
| DNA G+C (bp)                   | 2,799,660| 60.60      | 3,324,908| 62.33      | 3,123,835| 61.45      |
| DNA scaffolds                  | 329      | 100.00     | 309      | 100.00     | 358      | 100.00     |
| Total genes                    | 6522     | 100.00     | 6729     | 100.00     | 6758     | 100.00     |
| Protein coding genes           | 6457     | 99.00      | 6658     | 98.94      | 6696     | 99.08      |
| RNA genes                      | 65       | 1.00       | 71       | 1.06       | 62       | 0.92       |
| Pseudo genes                   | NA       | NA         | NA       | NA         | NA       | NA         |
| Genes in internal clusters     | NA       | NA         | NA       | NA         | NA       | NA         |
| Genes with function prediction | 2544     | 39.40      | 3075     | 46.19      | 2855     | 42.64      |
| Genes assigned to COGs         | 2714     | 42.03      | 3263     | 49.01      | 3046     | 45.49      |
| Genes with Pfam domains        | 2361     | 36.56      | 2948     | 44.28      | 2674     | 39.93      |
| Genes with signal peptides     | 270      | 4.18       | 334      | 5.02       | 303      | 4.53       |
| Genes with transmembrane helices | 291     | 4.51       | 349      | 5.24       | 311      | 4.64       |
| CRISPR repeats                 | 0        | -          | 0        | -          | 0        | -          |
3583 were considered hypothetical (Table 5). 49.01% of the CDSs could be assigned to one COG functional category and 44.28% contained one or more conserved PFAM-A domains (Table 6). The genome of R. solanacearum strain FJAT-462 has 358 scaffolds and the average GC content of the genome is 61.45% (Table 5). A total of 6758 genes (6696 CDSs and 62 ncRNAs) were predicted. Of the protein-coding genes, 2855 (42.64%) had functions assigned while 3,841 were considered hypothetical (Table 5). 45.49% of the CDSs could be assigned to one COG functional category and 39.93% contained one or more conserved PFAM-A domains (Table 6).

**Insights from the genome sequence**

**Comparative analysis of virulence-related genes**

T3E proteins are essential virulence factors in most gram-negative bacterial pathogens, such as R. solanacearum [2, 5], although they can also be perceived by resistant hosts as invasion signals, leading the development of plant defense responses [23]. The expression of genes encoding T3Es and structural components of the T3SS is activated after the perception of plant signals, and coordinated by a well-studied signaling pathway [24]. We analyzed the presence of genes involved in plant sensing and virulence regulation in the newly sequenced strains, and found that all the major regulators are present in the three strains (Table 7). These genes displayed a high percentage of similarity when compared to their homologs in the GMI1000 reference strain, ranging from 98.97 to 100% at the DNA level and from 99.19 to 100% at the amino acid level (Table 7).

The composition of T3E repertoires often defines the host range of specific strains. In this regard, we have identified over 70 T3Es in each strain based on comparisons with effector sequences in public databases (Table 8).

### Table 6 Number of genes that are associated with different COG functional categories

| Code | FJAT-91 | FJAT-452 | FJAT-462 | Description |
|------|---------|---------|---------|-------------|
| J    | 173     | 189     | 185     | Translation, ribosomal structure and biogenesis |
| A    | 2       | 2       | 2       | RNA processing and modification |
| K    | 221     | 273     | 264     | Transcription |
| L    | 104     | 119     | 115     | Replication, recombination and repair |
| B    | 2       | 2       | 3       | Chromatin structure and dynamics |
| D    | 47      | 51      | 53      | Cell cycle control, cell division, chromosome partitioning |
| V    | 52      | 60      | 61      | Defense mechanisms |
| T    | 167     | 191     | 180     | Signal transduction mechanisms |
| M    | 172     | 217     | 197     | Cell wall/membrane/envelope biogenesis |
| N    | 105     | 116     | 109     | Cell motility |
| Z    | 1       | 1       | 1       | Cytoskeleton |
| W    | 3       | 3       | 3       | Extracellular structures |
| U    | 107     | 109     | 119     | Intracellular trafficking, secretion, and vesicular transport |
| O    | 131     | 144     | 150     | Posttranslational modification, protein turnover, chaperones |
| X    | 65      | 120     | 71      | Mobilome: prophages, transposons |
| C    | 191     | 231     | 213     | Energy production and conversion |
| G    | 151     | 185     | 166     | Carbohydrate transport and metabolism |
| E    | 232     | 288     | 263     | Amino acid transport and metabolism |
| F    | 62      | 77      | 71      | Nucleotide transport and metabolism |
| H    | 107     | 140     | 127     | Coenzyme transport and metabolism |
| I    | 153     | 193     | 180     | Lipid transport and metabolism |
| P    | 131     | 147     | 143     | Inorganic ion transport and metabolism |
| Q    | 58      | 69      | 60      | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 175     | 213     | 200     | General function prediction only |
| S    | 182     | 232     | 211     | Function unknown |

The total is based on the total number of protein coding genes in the genome.
Comparisons with the reference GMI1000 strain suggest that the FJAT-91 strain lacks the T3E genes ripAG, ripS4, ripP3, hyp16, ripAI and ripY; the FJAT-452 strain lacks the T3E genes ripP3, hyp16 and ripM, and the FJAT-462 strain lacks the T3E genes ripAI, ripS4, ripP3, hyp16, ripM and ripAM. On the other hand, several T3E genes that are not present in GMI1000 were found in the three newly sequenced strains, including ripBE (in FJAT-462), ripS7 (in FJAT-452), hyp7 (in FJAT-452 and FJAT-462) and ripAL and ripF2 (in all 3 strains). The presence of most new T3E genes was confirmed by sequence analysis of PCR-amplified fragments from the three strains, being 100% identical among them and very similar or identical (78.84–100%) to their closest orthologs from other sequenced strains (Fig. 3). However, the hyp7 gene from FJAT-462 has a 1206 bp insertion annotated as a transposase 180 bp downstream the start codon (Fig. 3). By comparing the sequences of the T3E genes that are shared by the three newly sequenced strains and the reference strain GMI1000, we identified 652, 798 and 692 variant sites in the three strains (Table 7). These variations were classified into 7 types: missense variant, synonymous variant, frame shift variant, intron deletion, intron insertion, stop codon gain, and stop codon loss. Among them, 351 variations are shared by the three newly sequenced strains (Fig. 4). For example, the effector ripAI has both missense and synonymous variants, ripAZ1 have a frame-shift variant, and ripX has an intrame deletion in all three strains (Fig. 5).

Besides T3Es, R. solanacearum employs several additional virulence factors to achieve infection, such as EPS. The signaling cascade leading to the production of EPS involves several different regulatory components [25]. We analyzed the presence of genes involved in the regulation of EPS production, and found that all the major regulators are present in the three strains (Table 7). These genes displayed a high percentage of similarity when compared to their homologs in the GMI1000 reference strain, with most genes ranging from 98.35%–100% at the DNA level (98.26–100% at the amino acid level), with the exception of phcB, which shows a lower similarity in the FJAT-91 and FJAT-452 strains (86.41% at the DNA level in both strains) (Table 7). Other genes encoding putative virulence factors, such as egl (encoding an endoglucanase) and pehB (encoding an exo-poly-α-d-galacturonosidase) were also present in the three strains, with >99% similarity at the DNA and amino acid level compared to GMI1000 (Table 7).

Table 7 Sequence similarity of genes involved in R. solanacearum virulence between the three newly sequenced strains and the reference strain GMI1000

| Gene          | FJAT-91 | FJAT-452 | FJAT-462 |
|---------------|---------|----------|----------|
| Plant sensing and virulence regulation |         |          |          |
| hrgG          | 99.46   | 99.59    | 99.46    |
| phrA          | 99.54   | 99.77    | 100.00   |
| phrJ          | 99.23   | 100.00   | 99.94    |
| phrR          | 100.00  | 100.00   | 99.19    |
| Non-T3SS virulence factors |         |          |          |
| egl           | 99.05   | 99.37    | 99.20    |
| epsA          | 99.56   | 100.00   | 99.83    |
| epsB          | 99.63   | 99.86    | 99.58    |
| epsC          | 99.82   | 99.74    | 99.82    |
| epsD          | 99.68   | 100.00   | 99.68    |
| epsE          | 99.40   | 100.00   | 99.62    |
| epsF          | 99.75   | 99.75    | 98.35    |
| epsP          | 99.77   | 100.00   | 99.77    |
| epsR          | 99.86   | 100.00   | 99.86    |
| pehB          | 99.05   | 99.43    | 99.15    |
| phcA          | 99.62   | 99.71    | 99.71    |
| phcB          | 86.41   | 85.84    | 99.93    |
| vsrA          | 98.34   | 97.84    | 99.65    |
| vsrB          | 99.89   | 100.00   | 99.40    |
| vsrC          | 99.70   | 100.00   | 99.70    |
| vsrD          | 100.00  | 100.00   | 100.00   |
| xpsR          | 99.78   | 99.67    | 99.67    |

Table 8 Annotation and comparison of Type III effector genes in the three newly sequenced strains

| Number of T3E genes by Prokka annotation | FJAT-91 | FJAT-452 | FJAT-462 |
|------------------------------------------|---------|----------|----------|
| Total T3E genes after homology search    | 72      | 78       | 75       |
| Number of T3E genes not present in GMI1000 | 2       | 4        | 4        |
| Annotation of T3E genes not present in GMI1000 | ripAL   | ripAL    | ripAL    |
|                                               | ripF2   | ripF2    | ripF2    |
|                                               | ripS7   | ripS7    | ripS7    |
|                                               | hyp7    | hyp7     | hyp7     |
| Number of T3E genes in GMI1000 but not found in newly sequenced strain | 7       | 3        | 6        |

Conclusions

Earlier studies on the T3E repertoires of different plant pathogens suggested that T3E composition might shape the host range [6, 7]. In this study, we sequenced and analysed the genome of three R. solanacearum strains isolated from different host plants
with similar geographical origin (Fujian province, China). Our analysis indicates that each one of these strains have a unique effector repertoire (Table 7). In contrast to what we observed for T3E genes, all the analysed genes involved in the perception of plant signals and the regulation of virulence factors were present in all strains, and displayed a high degree of similarity between the newly sequenced strains and the GMI1000 reference strains (Table 7), suggesting that the mechanism of perception of plant signals does not differ significantly among bacteria infecting different plant species.

In addition to their presence or absence in specific strains, T3E genes may undergo several types of mutations that change or disrupt their coding sequence. As a consequence, the encoded proteins may lose the original function, become unstable, or gain a new function. This allelic diversification may be imposed by the host defense system, and allows

![Fig. 3 Schematic diagram of ripBE (a), hyp7 (b) and ripAL (c) sequence alignment. The nucleotide sequence of these genes in the strains sequenced in this study is 100% identical to each other, except for hyp7, which has an insertion annotated as a transposase in FJAT-462 (numbers indicate the insertion site). The percentage of identity compared to the orthologs in other sequenced strains is indicated in the figure.]

| Type               | FJAT-91 | FJAT-452 | FJAT-462 |
|--------------------|---------|----------|----------|
| All variants       | 652     | 798      | 692      |
| Missense variant   | 268     | 350      | 299      |
| Synonymous variant | 378     | 439      | 382      |
| Frameshift variant | 2       | 4        | 6        |
| Inframe deletion   | 1       | 1        | 2        |
| Inframe insertion  | 3       | 3        | 2        |
| Stop codon gain    | 0       | 0        | 1        |
| Stop codon loss    | 0       | 1        | 0        |

Table 9 Numbers and types of sequence variations (SNPs and INDELs) identified in the Type III effector genes (SNPs and INDELs) between the three newly sequenced strains and the reference strain GMI1000.
pathogens to avoid perception by the immune system of resistant host plants, in a phenomenon called pathoadaptation [26]. We identified alterations in effector sequences that were conserved in the three sequenced strains (Fig. 4). These sequence modifications may be due to the geographical distribution of these strains in comparison with the GMI1000 reference strain, originally isolated from French Guyana (South America), and may have functional relevance in the subversion of host functions in specific environmental conditions. Similarly, it is noteworthy that the FJAT-91 strain lacks 7 T3Es compared to GMI1000, while both are able to cause disease in tomato plants. Comparative analyses using the same tomato cultivars in controlled conditions will determine whether (i) these effectors are really dispensable to infect tomato, (ii) these effectors are dispensable in specific tomato cultivars, (iii) these effectors trigger immunity in specific tomato cultivars, or (iv) the environmental conditions in the FJAT-91 isolation site are more favourable to *R. solanacearum* infection, rendering unnecessary their virulence activities. The strain-specific absence of T3E genes or strain-specific loss-of-function variants (Fig. 4) may be caused by adaptation of these strains to specific hosts. Similarly, the transposase insertion in hyp7 (specific from FJAT-462; Fig. 3) is likely to alter or abolish the function of the encoded T3E in this strain, and may suggest that this T3E is not needed (or its alteration is actually required) to cause disease in chili pepper plants. Additional functional characterization will be required to determine whether these effectors induce immune responses in eggplant or chili pepper, and may allow the identification of novel sources of resistance against *R. solanacearum*. Our analysis shows that these unique effector repertoires are sufficient to cause disease in different hosts within a similar geographical location, allowing us to reduce the impact of environmental conditions in the analysis of the requirement of T3Es to cause infection. This information, together with the increasing number of sequenced *R.
solanacearum strains, constitutes one more step towards the identification of host specificity determinants for *R. solanacearum*.

**Abbreviations**

BLAST: Basic local alignment search tool; CDSS: Coding sequences; COG: Clusters of orthologous groups; EPS: Exopolysaccharide; E-value: Expect value; INDEL: Insertion-deletion; ncRNAs: Noncoding RNAs; Pfam: Protein families; SNP: Single nucleotide polymorphism; T3E: Type-III effector; TSSS: Type-III secretion system

**Acknowledgements**

We thank Suomeng Dong and Qinghe Chen for their help in providing the isolates sequenced in this work.

**Funding**

This work was supported by funds from the Chinese Academy of Sciences. CCM is sponsored by a CAS-TWAS President’s Fellowship for International PhD Students. RL is supported by the One Hundred Talent-Program of the Chinese Academy of Sciences. APM is supported by the Chinese 1000 Talents Program.

**Authors’ contributions**

Conceived the project: HZ, RL, APM. Prepared and sequenced libraries: WJ, AC. Assembled and annotated the genome: YS. Performed effector gene annotation: YS. Validated sequencing data and PCR-based analysis: CCM, KW. Analyzed and interpreted results: YS, KW, CCM, HZ, RL, APM. Performed effector secretion system

**Competing interests**

The authors declare that they have no competing interests.

**Publisher’s Note**

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

**Author details**

1. Shanghai Center for Plant Stress Biology, CAS Center for Excellence in Molecular Plant Sciences; Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 201602, China. 2. University of Chinese Academy of Sciences, Beijing 100039, China. 3. National Institute of Genetics, Nagoya 464-8602, Japan. 4. Chinese Academy of Agricultural Sciences, Beijing 100081, China. 5. Chinese Academy of Sciences. 6. National Oceanic and Atmospheric Administration, Pacific Marine Environmental Laboratory, Seattle, WA 98112, USA. 7. University of British Columbia, Vancouver, BC, Canada. 8. The Pennsylvania State University, University Park, PA 16802, USA. 9. National Institute for Agronomic Research, INRA, Montpellier, France. 10. Department of Biochemistry, National University of Singapore, 11 Kent Ridge Crescent, Singapore 119260, Singapore. 11. School of Biological Sciences, University of Hong Kong, Pokfulam, Hong Kong. 12. School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230001, China. 13. National Institute of Genetics, Nagoya 464-8602, Japan. 14. National Institute of Genetics, Nagoya 464-8602, Japan. 15. University of Hong Kong, Pokfulam, Hong Kong. 16. Beijing Genomics Institute, Shenzhen, Guangdong, China. 17. Finnish Centre for Genomics, University of Helsinki, Finland. 18. University of Arizona, Tucson, AZ, USA. 19. University of Hong Kong, Pokfulam, Hong Kong. 20. Chinese Academy of Plant Sciences, Beijing 100081, China. 21. Chinese Academy of Plant Sciences, Beijing 100081, China. 22. University of Hong Kong, Pokfulam, Hong Kong. 23. National Institute of Genetics, Nagoya 464-8602, Japan. 24. University of Hong Kong, Pokfulam, Hong Kong. 25. University of Hong Kong, Pokfulam, Hong Kong. 26. University of Hong Kong, Pokfulam, Hong Kong. 27. University of Hong Kong, Pokfulam, Hong Kong. 28. University of Hong Kong, Pokfulam, Hong Kong. 29. University of Hong Kong, Pokfulam, Hong Kong. 30. University of Hong Kong, Pokfulam, Hong Kong. 31. University of Hong Kong, Pokfulam, Hong Kong. 32. University of Hong Kong, Pokfulam, Hong Kong.

**Received: 9 January 2017 Accepted: 7 April 2017**

**Published online: 17 April 2017**

**References**

1. Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, Dow M, Baker J, Raskin I, Kaur A, ed. *Plant Pathogen Diversity and Evolution of the Type III Effector Gene Set in the Ralstonia solanacearum Species Complex*. St. Paul: APS Press; 2005. p. 9–28.
2. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Hof Richter E, Heger A, Heidbrink K, Helming K, Holm L, Mistry J, et al. Pfam: the protein families database. Nucleic Acid Res. 2014;42(Database issue):D223–230.
3. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
4. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
5. Li H. A statistical framework for SNP calling, discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987–93.
33. Garrity GM, Bell JA, Lilburn T. Class II. Betaproteobacteria class. nov. In: Brenner DJ, Krieg NR, Staley JT, editors. Bergey’s Manual® of Systematic Bacteriology: Volume Two The Proteobacteria Part C The Alpha-, Beta-, Delta-, and Epsilonproteobacteria. Boston: Springer US; 2005. p. 575–922.

34. Editor L. Validation List No. 57. Validation of the publication of new names and new combinations previously effectively published outside the IJSB. Int J Syst Bacteriol. 1996;46:625–6.

35. Yabuuchi E, Kosako Y, Yano I, Hotta H, Nishiuchi Y. Transfer of two Burkholderia and an Alcaligenes species to Ralstonia gen. Nov.: Proposal of Ralstonia pickettii (Ralston, Palleroni and Doudoroff 1973) comb. Nov., Ralstonia solanacearum (Smith 1896) comb. Nov. and Ralstonia eutropha (Davis 1969) comb. Nov. Microbiol Immunol. 1995;39(11):897–904.

36. Denny T, Hayward A, Schaad N, Jones J, Chun W. II Gram-negative bacteria. F. Ralstonia: Laboratory guide for identification of plant pathogenic bacteria. 2001. p. 151–74.

37. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–9.