Pangenomic type III effector database of the plant pathogenic *Ralstonia spp.*

Cyrus Raja Rubenstein Sabbagh [Equal first author, 1], **Sébastien Carrère** [Equal first author, 1], **Fabien Lonjon** [2], **Fabienne Vailleau** [1], **Alberto P Macho** [3], **Stephane Genin** [1], **Nemo Peeters** [Corresp. 1]

[1] LIPM, Université de Toulouse, INRA, CNRS, Castanet-tolosan, France

[2] Department of Cell & Systems Biology, University of Toronto, Toronto, Canada

[3] Shanghai center for plant stress biology, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Corresponding Author: Nemo Peeters
Email address: nemo.peeters@inra.fr

Background. The bacterial plant pathogenic *Ralstonia* species belong to the beta-proteobacteria order and are soil-borne pathogens causing the vascular bacterial wilt disease, affecting a wide range of plant hosts. These bacteria form a heterogeneous group considered as a “species complex”,” gathering three newly defined species. Like many other Gram negative plant pathogens, *Ralstonia* pathogenicity relies on a type III secretion system, enabling bacteria to secrete/inject a large repertoire of type III effectors into their plant host cells. T3Es are thought to participate in generating a favorable environment for the pathogen (countering plant immunity and modifying the host metabolism and physiology). **Methods.** Expert genome annotation, followed by specific type III-dependent secretion, allowed us to improve our Hidden-Markov-Model and Blast profiles for the prediction of type III effectors. **Results.** We curated the T3E repertoires of 12 plant pathogenic *Ralstonia*strains, representing a total of 12 strains spread over the different groups of the species complex. This generated a pangenome repertoire of 102 T3E genes and 16 hypothetical T3E genes. Using this database, we scanned for the presence of T3Es in the 155 available genomes representing 140 distinct plant pathogenic *Ralstonia* strains isolated from different host plants in different areas of the globe. All this information is presented in a searchable database. A presence/absence analysis, modulated by a strain sequence/gene annotation quality score, enabled us to redefine core and accessory T3E repertoires.

Pangenomic type III effector database of the plant pathogenic *Ralstonia spp.*

Cyrus Raja Rubenstein Sabbagh[1], Sébastien Carrère[1], Fabien Lonjon[1, 2], Fabienne Vailleau[1], Alberto P. Macho[3], Stéphane Genin[1] and Nemo Peeters[1]

[1] LIPM, Université de Toulouse, INRA, CNRS, Castanet-Tolosan, France
[2] Department of Cell & Systems Biology, University of Toronto, Toronto, Ontario, Canada.
[3] Shanghai Center for Plant Stress Biology, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, Shanghai, 201602, China.

Corresponding Author:
Nemo Peeters[1]
LIPM, INRA CS52627, Chemin de borde Rouge, Auzeville, 31326 Castanet-Tolosan, France
Email address: nemo.peeters@inra.fr

Abstract

Background. The bacterial plant pathogenic *Ralstonia* species belong to the beta-proteobacteria order and are soil-borne pathogens causing the vascular bacterial wilt disease, affecting a wide range of plant hosts. These bacteria form a heterogeneous group considered as a “species complex”,” gathering three newly defined species. Like many other Gram negative plant pathogens, *Ralstonia* pathogenicity relies on a type III secretion system, enabling bacteria to secrete/inject a large repertoire of type III effectors into their plant host cells. T3Es are thought to participate in generating a favorable environment for the pathogen (countering plant immunity and modifying the host metabolism and physiology).

Methods. Expert genome annotation, followed by specific type III-dependent secretion, allowed us to improve our Hidden-Markov-Model and Blast profiles for the prediction of type III effectors.

Results. We curated the T3E repertoires of 12 plant pathogenic *Ralstonia* strains, representing a total of 12 strains spread over the different groups of the species complex. This generated a pangenome repertoire of 102 T3E genes and 16 hypothetical T3E genes. Using this database, we scanned for the presence of T3Es in the 155 available genomes representing 140 distinct plant pathogenic *Ralstonia* strains isolated from different host plants in different areas of the globe. All this information is presented in a searchable database. A presence/absence analysis, modulated by a strain sequence/gene annotation quality score, enabled us to redefine core and accessory T3E repertoires.

Introduction

Plant pathogenic *Ralstonia* species (Peeters et al. 2013b) were ranked among the 10 most important plant bacterial pathogens (Mansfield et al. 2012). These soil-resident bacteria are indeed important, as they affect many different plant species, ranging from solanaceous crops to other important crops like banana and peanut, in different parts of the world. Recently, new plant species have been found to be infected and present symptoms of bacterial wilt, like blueberry shrubs in Florida, USA (Bocsanczy et al. 2019), ornamental roses in the Netherlands (Bergsma-Vlami et al. 2018), or pumpkin in China (She et al. 2017). This bacterium is also particularly studied as it has one of the largest known repertoire of T3Es among all plant or animal pathogenic bacteria. The secretion system (T3SS) of Gram negative phytopathogenic bacteria is essential for virulence, and Type III effectors (T3Es hereafter) have been found to contribute in many different and sometimes redundant manners to the fitness of the bacterium in interaction with its host (Buttner 2016).

Plant pathogenic *Xanthomonas* spp., and animal pathogens like *Escherichia* spp., *Shigella* spp. *or Yersinia* spp. have around 30 T3Es per strain (Dong et al. 2015; Schwartz et al. 2015). Classically known strains of *Pseudomonas* spp. have also around 30-40 T3Es (Wei et al. 2015),

with some more rare cases of up to 50 T3Es in a given strain (Dillon et al. 2019). It was reported that *legionella spp.* can secret in their host cells up to 300 effectors type IV effectors) (Gomez-Valero et al. 2019). Plant pathogenic *Ralstonia* spp. have between 46 to 71 T3Es (Peeters et al. 2013a).

In this work, we curated the genome of 2 new phylotype I strains bringin the total curated strains to 12 plant pathogenic *Ralstonia* strains, representing the known diversity of phylotypes (Wicker et al. 2012), more recently subdivided into three species (Safni et al. 2014). This generated new and updated profiles for the prediction of 102 Rips ("Ralstonia injected Proteins") and 16 hypothetical Rips, to be compared with the previous 94 Rips and 16 hypothetical Rips (Peeters et al. 2013a). Two hypothetical Rips from the reference strain CMR15, Psi07, and GMI1000 were experimentally confirmed as being *bona fide* Rips (and were named RipBM and RipBO).

The new and improved prediction profiles were used to analyze the effector repertoires of the 155 genomic sequences available in genbank. This dataset represents 140 different strains spread over the three newly defined species: 54 *Ralstonia solanacearum* (16 Phylotype IIA and 38 Phylotype IIB strains), 59 *Ralstonia pseudosolanacearum* (57 Phylotype I and 2 Phylotype III strains) and 27 *Ralstonia syzygii* (27 Phylotype IV strains). The prediction of all 118 Rips (including hypothetical Rips) over the whole dataset of 155 genomes/140 different strains is available as a browsable database, accommodating for direct comparisons between strain repertoires, from presence/absence tables to multiple alignments of DNA and protein sequences. This dataset was then further analyzed to evaluate how conserved the Rips are among these 140 strains. This was performed taking into account the host of isolation as a strong (but limited) host cue, or the phylogenetic group, to identify host or kinship repertoire conservation.

Materials & Methods

155 published whole genomes

The genbank genome data repository was scanned for the presence of complete genome sequences of *Ralstonia* species complex strains. The total number of genomes gathered was 155,

with some strains sequenced multiple times by different research groups, yielding sequence data for 140 distinctive strains. Owing to the fact that for a same strain different isolates could be slightly different, and also to the fact that sequence quality is important for gene repertoire completeness, we decided to keep all strain duplicates (in the database duplicates and triplicates are indicated as "-2" and "-3", respectively). Strains in duplicates are the following: FJAT-1458, FJAT-91, PSS4, CFBP2957, K60, CFBP6783, IBSBF1503, IPO1609, Po82, and UW163. Molk2 strain was present in the database with three independent sequence files, and UW551 with four independent sequence files. Table S1 contains all the available data on the 155 genome files. Whenever available, data for the following fields were also recorded : strain synonym; pubmed ID of reference articles; Species name (Safni et al. 2014); Phylotype; Geographical origin (isolation site); Plant isolated from; Genome assembly size; Assembly score; Number of contigs; Number of scaffolds; Bioproject. Figure S1 provides a *mutS* phylogeny (Wicker et al. 2012) indicating the strain relatedness.

Genome quality

Some genome sequences deposited by their authors were of insufficient quality to be included in the Refseq database; this is the case for the genomic entries FJAT-452, FJAT-462, T110, T12, T25 and UW700. These sequences were left on the complete database but were excluded for the further analysis in this work.

We then devised an assembly score in order to sort all the strains and to distinguish draft from complete and "polished" genomes. This score is Log_{10}(assembly size/number of contigs, with contigs being the N-free scaffolds or assembled pseudomolecules that were spliced by us on N stretches), and is a good general score to assess the overall completeness of the genome sequence. One exception to this is the strain CFBP2957, with a high score (6.287), but for which only the chromosome was available (and not the megaplasmid, see the bipartite nature of plant pathogenic *Ralstonia* genomes (Salanoubat et al. 2002)), and thus artificially increased the quality score.

This assembly score was insufficient, as we also felt the need to indirectly rate the gene annotation and prediction if it were to be further used in T3E repertoire comparisons. This is why we decided to generate two stringency cutoffs using the total T3E gene prediction performed by our prediction pipeline: for each strain the content in multicopy paralogous genes ("MULTI"),

the single defined genes ("OK"), the frame-shifted genes ("FS"), and the pseudogenes ("PG"), were computed for the 102 T3Es. We applied two levels of stringency. For "stringency 1", we kept only the strains for which the total number of pseudogenes plus frame-shifted genes is lower than 10: (PG+FS) < 10 ; this yielded a total of 123 genomes corresponding to 114 different strains. For "stringency 2", we only kept the strains that also had more than 50 T3Es in total ; this yielded to a set of 88 genomes corresponding to 84 different strains. Table S1 contains two columns identifying the 123 "stringency 1" and 88 "stringency 2" strains.
This "stringency" ranking is an artificial cutoff, but we believe this is a valid method to further compare the complete gene repertoires. The two strains T110 and UW700 have high genome assembly scores (6.45 and 5.06 respectively), but performed badly in this stringency test, with only 30 and 21 well-predicted T3Es (excluding them from "stringency 2" group) and 35 and 25 frameshifted and pseudogenes (excluding them from "stringency 1" group).

Gene presence/absence

For each strain, the Table S1 contains the presence/absence scoring for all the 102 Rips and 16 hypothetical Rips. We used the prediction data for each strain (see further) as highlighted on the database website (https://iant.toulouse.inra.fr/T3E). Frameshifted genes are rare in well-sequenced genomes. Indeed, out of the 67 strains reported with two scaffolds (corresponding to the expected chromosome and megaplasmid (Salanoubat et al. 2002)), 52 have no frame-shifted genes, and nine only one frameshifted gene (see Table S1 for the data). We thus hypothesized that a frameshift is more due to sequencing errors than representing true genomic data. As a consequence, when making a binary scale for scoring the presence/absence of T3Es, we considered all "MULTI" (recently duplicated genes), "OK" (single gene) and "FS" (frameshifted) as "1" (or "present"); when absence "NO" and pseudogene ("PG") were considered as "0" (or "Absent"), this same reasoning was used previously (Peeters et al. 2013a).

T3E prediction improvements

We have improved our first T3E prediction pipeline (Peeters et al. 2013a), by adding databases of confirmed T3Es from *Xanthomonas* spp. (www.xanthomonas.org/t3e.html) and *Pseudomonas* spp. strains (www.pseudomonas-syringae.org/T3SS-Hops.xls). In order to capture more distantly related Rips, we lowered the tblastN/blastX thresholds (Query coverage per subject 60 %, and

Percentage of identical matches 60 %), this was well exemplified by the RipBN case, an AvrRpt2 ortholog clearly present in CMR15strain (Eschen-Lippold et al. 2016) and also detectable by blast, but not without slightly lower thresholds. We also rewrote some parts of the pipeline in order to speed up the prediction engine. The updated pipeline is outlined in Fig. 1.

T3Edb v3 specificities

The new database version is very similar to the previous version (Peeters et al. 2013a), except that a first set of curated strains ("curated repertoire") available to inspect. This set is composed of the following strains: 244 (phylotype I) (Ramesh et al. 2014); GMI1000 (I) (Salanoubat et al. 2002); YC45 (I) (She et al. 2015); CFBP2957 (IIA) (Remenant et al. 2010); CMR15 (III) (Remenant et al. 2010); IPO1609 (IIB) (Gonzalez et al. 2011); Molk2 (IIB) (Remenant et al. 2010); Po82 (IIB) (Xu et al. 2011); UW551 (IIB) (Gonzalez et al. 2011); PSI07 (IV) (Remenant et al. 2010); BDBR229 (IV) (Remenant et al. 2011), and R24 (IV) (Remenant et al. 2011).
In order to "build profiles" of Rip prediction in different strains to compare the strains and/or to generate multifasta files (of nucleotide or protein sequences of specific Rips), one can now sort the whole set of 155 complete genomes on the different headers available, namely these are: "status (curated or not); "code" (abbreviated name); "synonym", "species name" (Safni et al. 2014); "phylotype"; "plant isolated from"; "assembly size"; "number of contigs"; "number of scaffolds"; "assembly score" (see definitions above).

Type III secretion dependence

The type III-dependent secretion of two previously defined hypothetical T3Es in strains GMI1000, CMR15, and PSI07 (Peeters et al. 2013a) was demonstrated in this work. The coding sequences of PSI07_1860 (formerly Hyp15) and CMR15v4_mp10184 (formerly Hyp15), were ordered as DNA synthesis from Sangon (Shanghai, China). RSc3174, from the reference strain GMI1000 (formerly Hyp16) was amplified in two steps. The first PCR was performed using the following primers: Forward : 5' GGAGATAGAACCATGAAAGTCGGCAACCAATC-3' and Reverse 5' CAAGAAAGCTGGGTCTCCACGTGATAAGTTGTAGCG-3'. The second PCR was performed using 1 µl of the first PCR as matrix and attB universal primers (oNP291 5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG-3' and oNP292 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTC-3'. Then, Rsc3174, PSI07_1860

and CMR15v4_mp10184 were cloned into pDONR207 vector using a BP reaction and in pNP329 using a LR reaction following the instructions of the manufacturer. The final expression vectors were transformed into *R. pseudosolanacearum* GMI1000 strain and in the *hrcV* mutant (type III secretion defective mutant, used as a negative control) as previously described (Perrier et al. 2018). In-vitro Secretion assays and Western blot analysis were performed as previously described (Lonjon et al. 2018).

Results

Curation of 2 new Phylotype I strains and identification of 8 New Rips

Because strain GMI1000 was the only curated *R. pseudosolanacearum* strain in the former RalstoT3Edb (Peeters et al. 2013a), we conducted a manual curation of the Type III effectome in two other *R. pseudosolanacearum* strains, both differing in host range from GMI1000. Strain Rs-10-244 was isolated from chilli pepper (*Capsicum annuum*) in Andaman Islands (India) (Ramesh et al. 2014) and strain YC45 was isolated from a monocotyledoneous host, aromatic ginger (*Rhizoma kaempferiae*) in Southern China (She et al. 2015). Manual curation identified 73 Rip genes (+1 candidate) in strain YC45 and 77 Rip genes (+3 candidates) in Rs-10-244. Novel Rip effectors and candidates were identified in these strains (Table 1).

RipBJ was identified by secretome analysis of the GMI1000 strain (*R. pseudosolanacearum*) (Lonjon et al. 2016). RipBK and RipBL were identified in the process of curation of the strain YC45 (*R. pseudosolanacearum)*, owing to their homology to HopAM1 (Chang et al. 2005; Goel et al. 2008) and HopAO1 (Chang et al. 2005; Macho et al. 2014). RipBM and RipBO, formerly known as Hyp15 and Hyp16, respectively (Peeters et al. 2013a), were experimentally confirmed to be secreted by the T3SS in GMI1000 (Fig. 2, Fig. S2). RipBN was identified by sequence

homology in the strain CMR15 (*R. pseudosolanacearum*) (Eschen-Lippold et al. 2016). RipBP (homolog to HopW1 (Zumaquero et al. 2010)) was identified in the strain OE1-1 (*R. pseudosolanacearum*), and RipBQ (homolog to HopK1 (Chang et al. 2005)) in the strain KACC10722 (*R. syzygii*). RipBP and RipBQ are considered here as Rips by applying the rule of homology with a known T3E (Peeters et al. 2013a). These two latter Rips have been highlighted on the curated list of strains although none of these curated strains harbor these effectors. This is the same for RipBE, specific to RS1000 strain (Mukaihara & Tamura 2009; Peeters et al. 2013a). Table 1 mentions the reference sequences for new Rip genes. Two new hypothetical Rips were also identified; named Hyp17 and Hyp18 and have two associated reference sequences (See Table 1). Considering that in the previous database (Peeters et al. 2013a), some Rips were only represented by pseudogenes, we corrected this by attributing new reference sequences to RipBA (strain Rs-10-244, sequence RS244_c002320), RipBE (strain YC40-M, sequence YC40-M_00170) and RipP3 (strain Rs-10-244, sequence RS244_c031810).

Improved Rip-scanning pipeline

Thanks to the increased dataset of 102 total Rips (and 16 hypothetical T3Es), on a total of 155 genomes (totaling 140 different strains), we were able to generate new effector profiles for the better prediction of these Rips and candidate Rips in newly available *Ralstonia* genomes. The "scan your genome" tool is available on the database website. For large dataset analysis, we prefer to be contacted directly (RalstoT3E-toulouse@inra.fr), to prevent server overload. A Blast tool, as well as all the files and results of predictions for the 155 genomes are also available on our website. A convenient tool is the availability of multifasta files for the nucleotide or protein sequences for a given Rip, containing the genome/strain sequences that one queried for comparison in the first place.

Core effectors

We wanted to have a new look at the number of conserved Rips among this new diverse set of strains. As a principle, as more strains are compared, the smaller the core set of Rips will become. In order to have a pertinent set of strains to compare, we decided to limit the core comparisons to the "stringency 2" set of strains (the 84 distinct strains having less than 10 pseudogenes or frameshifted genes and, at the same time, more than 50 predicted Rip genes).

Fig. 3 shows a phylogenetic tree built using the *mutS* gene sequence of these 84 strains to be able to judge the relatedness of this set of strains. We are aware of the risk of excluding some strains based on this stringent selection. This could in particular be the case for the known strains that have seen genome reduction and hence have fewer T3Es. This is the case for the Moko disease, or blood disease bacterium BDBR229 (Remenant et al. 2011), and the *R. syzygii* strain insect-transmitted clove-tree infecting R24 (Remenant et al. 2011), which each have respectively 54 and 48 Rip genes (as defined under the "stringency 2" criteria). Moreover, both strains are already left out under "stringency 1" criteria, for having more that 10 (respectively 20 and 12), frameshifted and pseudogenes.

We then decided to use the host of isolation as an interesting criterion to compare strains. Of course there are numerous examples of strains isolated on one host and later shown in laboratory settings to be able to infect other host plants. GMI1000, isolated from tomato (Salanoubat et al. 2002), was shown to be very well capable of wilting *Medicago truncatula* (Vailleau et al. 2007) or *Arabidopsis thaliana* (Deslandes et al. 1998). As laboratory settings are hard to compare between labs, and as thorough host-compatibility has been done only for a handful of strains, we preferred to stick with the host of isolation information, without excluding that the host range might be much wider for some strains, and restricted for others. We decided to compare the conservation of Rip repertoires among the 84 "stringency 2" strains, classifying them into hosts of isolation; *Solanaceae* strains ("SOL"), tomato strains ("TOM", 15 strains), Eggplant ("EGG", 9 strains), potato ("POT", 30 strains) and banana ("BAN", 15 strains). The larger, encompassing category being the Solanaceae group, with 58 strains (for the list of strains see Table S1). Table 2 indicates the list, per host-of-isolation category of the core set of Rip genes. For a set of n total strains, we decided to still consider core, the Rip genes present in the interval (n; n-5%) number of strains. For instance, for the 58 "SOL" strains, the core Rip genes are the ones present in 58 to 55 strains. Obviously, the larger the number of strains, the least Rips are conserved. For instance, for the 9 "EGG" strains, there are 44 strictly conserved Rips (in all 9 strains), whereas there are only 27 conserved Rips in the 15 "BAN" strains (in 14 to 15 strains). Figure 4 show two Venn diagram comparing these sets of conserved Rip between host-of-isolation groups.

The set of 140 strains is evenly spread over the three newly defined *Ralstonia* species (see Fig. S1 for a *mutS* phylogeny (Wicker et al. 2012) of the 140 strain/155 genome sequences of

this study). One possible caveat is the few phylotype III strains (two strains: CMR15 and CFBP3059), now classified with phylotype I strains among the *R. pseudosolanacearum*. One interesting way to look at the conservation of Rips is to make specific species groups. The 84 "stringency 2" strains belong to *R. pseudosolanacearum* (phylotypes I and III), for 38 strains; to *R. solanacearum* (phylotypes IIA and IIB), for 25 strains; and to *R. syzygii* (phylotype IV), for 21 strains, see Fig. 3 for the *mutS* phylogeny of these 84 strains (Wicker et al. 2012). Table 3 represents the Rip distribution among these three species, together with the conservation in the total set of 84 strains. Figure 5 displays the Venn diagram corresponding to this triple comparison.

Discussion

In this work, we significantly updated the *Ralstonia* type III secretion effector database (T3Edb) (Peeters et al. 2013a). This latter work, providing a new nomenclature for these essential virulence proteins was widely accepted and cited by the community. Here, we reported on the curation of new plant pathogenic *Ralstonia* strains, adding new T3Es to this database. These are the new series from RipBJ to RipBQ, among which both RipBM and RipBO were shown in this work to be indeed secreted by the GMI1000 (*R. pseudosolanacearum)* T3SS. One of these newly defined Rips, RipBN, was identified for being an ortholog to the *Pseudomonas syringae* AvrRpt2 T3E (Eschen-Lippold et al. 2016), and recently shown to function similarly in the triggering of resistance in Ptr1-tomato lines (Mazo-Molina et al. 2019).

The newly defined Rip profiles (102 Rips and 16 Hypothetical Rips), were then used to predict the T3E repertoire of the 155 genome sequences available, representing a total of 140 different strains, compared to the 12 genomes previously available. This large set of strains allows us to provide an updated database with a better representation of each of the three phylogenetic clades of this species complex. These are: Phylotypes I and III, or the new proposed species name *R. pseudosolanacearum (Safn*i et al. 2014), phylotypes IIA and IIB, or *R. solanacearum* and phylotype IV, or *R. syzygii*. For a better view of strain relatedness, *mutS* phylogenies are displayed in Fig. 3 (set of 84 strains), and in Fig. S1 (all 140 strains). In order to understand the contribution of these T3Es to the virulence of these bacteria on their host plant, it is particularly interesting to analyze which T3Es are conserved among the different strains. Our comparison

results highlight two ways to explore these repertoires: either by host plant or by phylogenetic relatedness.

Ideally, each of the well-sequenced and well-annotated strains (the "stringency 2" list of 84 strains) should be tested on a panel of host plants in order to define their actual host range. As these data is not available we focused on the host of isolation as a limited but natural host definition factor. This is a strong limitation in this comparison, as it is reported or known (and shared through personal communications) that some strains are also compatible on other, and sometimes distantly related), hosts. We decided to add the published information on the compatibility on other host plants in the Table S1. A few research groups added a significant amount of host-compatibility information for a set of strains (Ailloud et al. 2015; Cho et al. 2018; Lebeau et al. 2011). Other groups have performed host-compatibility experiments and shared this information with us, *e.g.* tobacco strain CQPS-1 (Liu et al. 2017) is also mildly pathogenic on tomato (Personal communication Prof. Ding W.), when the blueberry strains P816, P822 and P824 (Bocsanczy et al. 2019), are very aggressive on tomato (personal communication Dr Norman DJ). In our repertoire comparison, we allow a tolerance of presence for the Rip in the interval of strains between the total number of strains n, and n-5%, this allows to compensate the effect of unequal set of strains to compare. Table 2 and Fig. 4A show that, unsurprisingly strains isolated from Eggplant "EGG", tomato "TOM" and potato "POT", share a significative amount of their conserved T3Es (22 shared out of 44 "EGG", 44 "TOM" and 30 "POT"), this number is probably largely underestimated as we know that some of these *Solanaceae*-isolated strains are compatible on other *Solanaceae* (Lebeau et al. 2011). Another comparison showed (Fig. 4B) is between the "TOM", "POT" and banana "BAN" isolated strains. Here, we can see that there could be more T3Es shared between "BAN" and "TOM" (21 out of 27 "BAN" strains) than between "BAN" and "POT" (15 out of 27 "BAN" strains). To evaluate this potential difference, one has to keep in mind that, although banana and *Solanaceae* are distantly related, it has been shown that 9 out of the 27 "BAN" strains are also compatible on tomato and potato, when only one strain (BDBR229) was shown to be incompatible on these two *Solanaceae* hosts (Ailloud et al. 2015). When considering all *Solanaceae* (SOL) as host of isolation strains (58 strains from the "stringency 2" set of 84 strains), the core set of T3Es (as defined to be present in 55 to 58 strains) is represented by a list of 27 T3Es (see Table 2). It is

only when host-compatibility is confronted in detail with T3E repertoire that we can start to potentially associate the presence (or the presence of specific alleles) to be required (or deleterious) for specific host-compatibility (Cho et al. 2019; Wang et al. 2016).

A second and maybe less ambiguous way to compare lists of conserved T3Es is to group the strains by their phylogenetic origin. Table 3 summarizes the T3Es conserved within each of the three phylogenetic groups of strains (Wicker et al. 2012). These groups are: the 38 strains from phylotypes I and III, or *R. pseudosolanacearum;* the 25 strains from phylotypes IIA and IIB, or *R. solanacearum;* and the 21 strains from phylotype IV or *R. syzygii* ; Fig. 3 display the *mutS* phylogeny of these 84 strains. Some strong phylogenetic associated presence/absence can be highlighted, like the systematic presence in *R. pseudosolanacearum* and *R. syzygii* and systematic absence, in *R. solanacearum* conserved T3Es of RipA2, RipG5 and RipZ. Some T3Es are systematically associated with only one of these phylogenetic groups, like RipC1, RipI, RipAC, RipD in *R. solanacearum*; RipL, RipQ, RipS4 in *R. pseudosolanacearum* and RipA5, RipM, RipS5, RipAQ in *R syzygii.* Sixteen (16) T3Es are conserved among the phylogenetic groups (Fig. 5 and Table 3). Eight of them are conserved in the different species: RipB (absent only in *R. pseudosolanacearum* CQPS-1 (Liu et al. 2017)); RipH2 (absent only in *R. pseudosolanacearum* RSCM isolated from Cucurbita maxima in China (She et al. 2017)); RipR (absent only in *R solanacearum* UW181, a plantain banana strain (Wicker et al. 2012), and *R. syzygii* BDB_RUN1347, no host of isolation reported); RipW (absent only in *R. pseudosolanacearum* strain SL3822 isolated form potato in Korea (Cho et al. 2018); RipAB (absent only in *R. pseudosolanacearum* strain YC40-M, no host of isolation reported, and *R. solanacearum* strain MolK2 (Remenant et al. 2010)); RipAI (absent only in *R. pseudosolanacearum* strain HA4-1 a Chinese peanut strain); RipAO (absent only in *R. pseudosolanacearum* strain SL3755 isolated form potato in Korea (Cho et al. 2018)). The only strictly conserved T3E among these 84 strains being RipAJ. Eight other are slightly underrepresented in one species out of the three (number of strains in which the T3E is present is indicated in brackets in Table 3). Among these two (2) are less conserved in *R. pseudosolanacearum* (Phylotype I and III): RipAN and RipAY; four (4) are less conserved in *R. solanacearum* (Phylotype II): RipA2, RipG5, RipG6, RipAM; two (2) are less conserved in *R. syzygii* (Phylotype IV): RipU and RipV1.

Conclusions

This work describes the methods and strains used to build a comprehensive database of the type III effectors (T3Es) from the *Ralstonia solanacearum* Species Complex (*R. solanacearum, R. pseudosolanacearum* and *R. syzygii*). Representing a resource to both study and identify new allelic version of specific T3Es, indeed the database contains all the specific T3E sequences (102 T3Es and 16 hypothetical T3Es over 155 strains), but also allows to identify new T3E ortholog by scanning DNA sequence (partial, shotgun or complete genomes) originating from original isolates.

Acknowledgements

references

Ailloud F, Lowe T, Cellier G, Roche D, Allen C, and Prior P. 2015. Comparative genomic analysis of Ralstonia solanacearum reveals candidate genes for host specificity. *BMC Genomics* 16:270. 10.1186/s12864-015-1474-8

Bergsma-Vlami M, van de Bilt JLJ, Tjou-Tam-Sin NNA, Westenberg M, Meekes ETM, Teunissen HAS, and Van Vaerenbergh J. 2018. Phylogenetic Assignment of Ralstonia

pseudosolanacearum (Ralstonia solanacearum Phylotype I) Isolated from Rosa spp. *Plant Dis* 102:2258-2267. 10.1094/PDIS-09-17-1345-RE

Bocsanczy AM, Espindola AS, and Norman DJ. 2019. Whole-Genome Sequences of Ralstonia solanacearum Strains P816, P822, and P824, Emerging Pathogens of Blueberry in Florida. *Microbiol Resour Announc* 8. 10.1128/MRA.01316-18

Buttner D. 2016. Behind the lines-actions of bacterial type III effector proteins in plant cells. *FEMS Microbiol Rev* 40:894-937. 10.1093/femsre/fuw026

Chang JH, Urbach JM, Law TF, Arnold LW, Hu A, Gombar S, Grant SR, Ausubel FM, and Dangl JL. 2005. A high-throughput, near-saturating screen for type III effector genes from Pseudomonas syringae. *Proc Natl Acad Sci U S A* 102:2549-2554. 10.1073/pnas.0409660102

Cho H, Song ES, Heu S, Baek J, Lee YK, Lee S, Lee SW, Park DS, Lee TH, Kim JG, and Hwang I. 2019. Prediction of Host-Specific Genes by Pan-Genome Analyses of the Korean Ralstonia solanacearum Species Complex. *Front Microbiol* 10:506. 10.3389/fmicb.2019.00506

Cho H, Song ES, Lee YK, Lee S, Lee SW, Jo A, Lee BM, Kim JG, and Hwang I. 2018. Analysis of Genetic and Pathogenic Diversity of Ralstonia solanacearum Causing Potato Bacterial Wilt in Korea. *Plant Pathol J* 34:23-34. 10.5423/PPJ.FT.09.2017.0203

Deslandes L, Pileur F, Liaubet L, Camut S, Can C, Williams K, Holub E, Beynon J, Arlat M, and Marco Y. 1998. Genetic characterization of RRS1, a recessive locus in Arabidopsis thaliana that confers resistance to the bacterial soilborne pathogen Ralstonia solanacearum. *Mol Plant Microbe Interact* 11:659-667. 10.1094/MPMI.1998.11.7.659

Dillon MM, Thakur S, Almeida RND, Wang PW, Weir BS, and Guttman DS. 2019. Recombination of ecologically and evolutionarily significant loci maintains genetic cohesion in the Pseudomonas syringae species complex. *Genome Biol* 20:3. 10.1186/s13059-018-1606-y

Dong X, Lu X, and Zhang Z. 2015. BEAN 2.0: an integrated web resource for the identification and functional analysis of type III secreted effectors. *Database (Oxford)* 2015:bav064. 10.1093/database/bav064

Eschen-Lippold L, Jiang X, Elmore JM, Mackey D, Shan L, Coaker G, Scheel D, and Lee J. 2016. Bacterial AvrRpt2-Like Cysteine Proteases Block Activation of the Arabidopsis Mitogen-Activated Protein Kinases, MPK4 and MPK11. *Plant Physiol* 171:2223-2238. 10.1104/pp.16.00336

Goel AK, Lundberg D, Torres MA, Matthews R, Akimoto-Tomiyama C, Farmer L, Dangl JL, and Grant SR. 2008. The Pseudomonas syringae type III effector HopAM1 enhances virulence on water-stressed plants. *Mol Plant Microbe Interact* 21:361-370. 10.1094/MPMI-21-3-0361

Gomez-Valero L, Rusniok C, Carson D, Mondino S, Perez-Cobas AE, Rolando M, Pasricha S, Reuter S, Demirtas J, Crumbach J, Descorps-Declere S, Hartland EL, Jarraud S, Dougan G, Schroeder GN, Frankel G, and Buchrieser C. 2019. More than 18,000 effectors in the Legionella genus genome provide multiple, independent combinations for replication in human cells. *Proc Natl Acad Sci U S A* 116:2265-2273. 10.1073/pnas.1808016116

Gonzalez A, Plener L, Restrepo S, Boucher C, and Genin S. 2011. Detection and functional characterization of a large genomic deletion resulting in decreased pathogenicity in Ralstonia solanacearum race 3 biovar 2 strains. *Environ Microbiol* 13:3172-3185. 10.1111/j.1462-2920.2011.02636.x

Lebeau A, Daunay MC, Frary A, Palloix A, Wang JF, Dintinger J, Chiroleu F, Wicker E, and Prior P. 2011. Bacterial wilt resistance in tomato, pepper, and eggplant: genetic resources respond to diverse strains in the Ralstonia solanacearum species complex. *Phytopathology* 101:154-165. 10.1094/PHYTO-02-10-0048

Liu Y, Tang Y, Qin X, Yang L, Jiang G, Li S, and Ding W. 2017. Genome Sequencing of Ralstonia solanacearum CQPS-1, a Phylotype I Strain Collected from a Highland Area with Continuous Cropping of Tobacco. *Front Microbiol* 8:974. 10.3389/fmicb.2017.00974

Lonjon F, Peeters N, Genin S, and Vailleau F. 2018. In Vitro and In Vivo Secretion/Translocation Assays to Identify Novel Ralstonia solanacearum Type 3 Effectors. *Methods Mol Biol* 1734:209-222. 10.1007/978-1-4939-7604-1_17

Lonjon F, Turner M, Henry C, Rengel D, Lohou D, van de Kerkhove Q, Cazale AC, Peeters N, Genin S, and Vailleau F. 2016. Comparative Secretome Analysis of Ralstonia solanacearum Type 3 Secretion-Associated Mutants Reveals a Fine Control of Effector Delivery, Essential for Bacterial Pathogenicity. *Mol Cell Proteomics* 15:598-613. 10.1074/mcp.M115.051078

Macho AP, Schwessinger B, Ntoukakis V, Brutus A, Segonzac C, Roy S, Kadota Y, Oh MH, Sklenar J, Derbyshire P, Lozano-Duran R, Malinovsky FG, Monaghan J, Menke FL, Huber SC, He SY, and Zipfel C. 2014. A bacterial tyrosine phosphatase inhibits plant pattern recognition receptor activation. *Science* 343:1509-1512. 10.1126/science.1248849

Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, Dow M, Verdier V, Beer SV, Machado MA, Toth I, Salmond G, and Foster GD. 2012. Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol Plant Pathol* 13:614-629. 10.1111/j.1364-3703.2012.00804.x

Mazo-Molina C, Mainiero S, Hind SR, Kraus CM, Vachev M, Maviane-Macia F, Lindeberg M, Saha S, Strickler S, Feder A, Giovannoni J, Smart C, Peeters N, and Martin G. 2019. The Ptr1 locus of Solanum lycopersicoides confers resistance to race 1 strains of Pseudomonas syringae pv. tomato and to Ralstonia pseudosolanacearum by recognizing the type III effectors AvrRpt2/RipBN. *Mol Plant Microbe Interact*. 10.1094/MPMI-01-19-0018-R

Mukaihara T, and Tamura N. 2009. Identification of novel Ralstonia solanacearum type III effector proteins through translocation analysis of hrpB-regulated gene products. *Microbiology* 155:2235-2244. 10.1099/mic.0.027763-0

Peeters N, Carrere S, Anisimova M, Plener L, Cazale AC, and Genin S. 2013a. Repertoire, unified nomenclature and evolution of the Type III effector gene set in the Ralstonia solanacearum species complex. *BMC Genomics* 14:859. 10.1186/1471-2164-14-859

Peeters N, Guidot A, Vailleau F, and Valls M. 2013b. Ralstonia solanacearum, a widespread bacterial plant pathogen in the post-genomic era. *Mol Plant Pathol* 14:651-662. 10.1111/mpp.12038

Perrier A, Barberis P, and Genin S. 2018. Introduction of Genetic Material in Ralstonia solanacearum Through Natural Transformation and Conjugation. *Methods Mol Biol* 1734:201-207. 10.1007/978-1-4939-7604-1_16

Ramesh R, Gaitonde S, Achari G, Asolkar T, Singh NP, Carrere S, Genin S, and Peeters N. 2014. Genome Sequencing of Ralstonia solanacearum Biovar 3, Phylotype I, Strains Rs-09-161 and Rs-10-244, Isolated from Eggplant and Chili in India. *Genome Announc* 2. 10.1128/genomeA.00323-14

Remenant B, Coupat-Goutaland B, Guidot A, Cellier G, Wicker E, Allen C, Fegan M, Pruvost O, Elbaz M, Calteau A, Salvignol G, Mornico D, Mangenot S, Barbe V, Medigue C, and Prior P. 2010. Genomes of three tomato pathogens within the Ralstonia solanacearum species complex reveal significant evolutionary divergence. *BMC Genomics* 11:379. 10.1186/1471-2164-11-379

Remenant B, de Cambiaire JC, Cellier G, Jacobs JM, Mangenot S, Barbe V, Lajus A, Vallenet D, Medigue C, Fegan M, Allen C, and Prior P. 2011. Ralstonia syzygii, the Blood

Disease Bacterium and some Asian R. solanacearum strains form a single genomic species despite divergent lifestyles. *PLoS One* 6:e24356. 10.1371/journal.pone.0024356

Safni I, Cleenwerck I, De Vos P, Fegan M, Sly L, and Kappler U. 2014. Polyphasic taxonomic revision of the Ralstonia solanacearum species complex: proposal to emend the descriptions of Ralstonia solanacearum and Ralstonia syzygii and reclassify current R. syzygii strains as Ralstonia syzygii subsp. syzygii subsp. nov., R. solanacearum phylotype IV strains as Ralstonia syzygii subsp. indonesiensis subsp. nov., banana blood disease bacterium strains as Ralstonia syzygii subsp. celebesensis subsp. nov. and R. solanacearum phylotype I and III strains as Ralstonia pseudosolanacearum sp. nov. *Int J Syst Evol Microbiol* 64:3087-3103. 10.1099/ijs.0.066712-0

Salanoubat M, Genin S, Artiguenave F, Gouzy J, Mangenot S, Arlat M, Billault A, Brottier P, Camus JC, Cattolico L, Chandler M, Choisne N, Claudel-Renard C, Cunnac S, Demange N, Gaspin C, Lavie M, Moisan A, Robert C, Saurin W, Schiex T, Siguier P, Thebault P, Whalen M, Wincker P, Levy M, Weissenbach J, and Boucher CA. 2002. Genome sequence of the plant pathogen Ralstonia solanacearum. *Nature* 415:497-502. 10.1038/415497a

Schwartz AR, Potnis N, Timilsina S, Wilson M, Patane J, Martins J, Jr., Minsavage GV, Dahlbeck D, Akhunova A, Almeida N, Vallad GE, Barak JD, White FF, Miller SA, Ritchie D, Goss E, Bart RS, Setubal JC, Jones JB, and Staskawicz BJ. 2015. Phylogenomics of Xanthomonas field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. *Front Microbiol* 6:535. 10.3389/fmicb.2015.00535

She X, Tang Y, He Z, and Lan G. 2015. Genome Sequencing of Ralstonia solanacearum Race 4, Biovar 4, and Phylotype I, Strain YC45, Isolated from Rhizoma kaempferiae in Southern China. *Genome Announc* 3. 10.1128/genomeA.01110-15

She X, Yu L, Lan G, Tang Y, and He Z. 2017. Identification and Genetic Characterization of Ralstonia solanacearum Species Complex Isolates from Cucurbita maxima in China. *Front Plant Sci* 8:1794. 10.3389/fpls.2017.01794

Vailleau F, Sartorel E, Jardinaud MF, Chardon F, Genin S, Huguet T, Gentzbittel L, and Petitprez M. 2007. Characterization of the interaction between the bacterial wilt pathogen Ralstonia solanacearum and the model legume plant Medicago truncatula. *Mol Plant Microbe Interact* 20:159-167. 10.1094/MPMI-20-2-0159

Wang K, Remigi P, Anisimova M, Lonjon F, Kars I, Kajava A, Li CH, Cheng CP, Vailleau F, Genin S, and Peeters N. 2016. Functional assignment to positively selected sites in the

core type III effector RipG7 from Ralstonia solanacearum. *Mol Plant Pathol* 17:553-564. 10.1111/mpp.12302

Wei HL, Chakravarthy S, Mathieu J, Helmann TC, Stodghill P, Swingle B, Martin GB, and Collmer A. 2015. Pseudomonas syringae pv. tomato DC3000 Type III Secretion Effector Polymutants Reveal an Interplay between HopAD1 and AvrPtoB. *Cell Host Microbe* 17:752-762. 10.1016/j.chom.2015.05.007

Wicker E, Lefeuvre P, de Cambiaire JC, Lemaire C, Poussier S, and Prior P. 2012. Contrasting recombination patterns and demographic histories of the plant pathogen Ralstonia solanacearum inferred from MLSA. *ISME J* 6:961-974. 10.1038/ismej.2011.160

Xu J, Zheng HJ, Liu L, Pan ZC, Prior P, Tang B, Xu JS, Zhang H, Tian Q, Zhang LQ, and Feng J. 2011. Complete genome sequence of the plant pathogen Ralstonia solanacearum strain Po82. *J Bacteriol* 193:4261-4262. 10.1128/JB.05384-11

Zumaquero A, Macho AP, Rufian JS, and Beuzon CR. 2010. Analysis of the role of the type III effector inventory of Pseudomonas syringae pv. phaseolicola 1448a in interaction with the plant. *J Bacteriol* 192:4474-4488. 10.1128/JB.00260-10

549 **Figure legends**

550

551 **Figure 1. New T3E prediction pipeline.**

552

553 **Figure 2. RipBM and RipBO are secreted through the T3SS.**

554 The wild-type strain and the *hrcV* mutant were transformed to express a $RipBM_{CMR15}$-3HA,

555 $RipBM_{Psi07}$-3HA **(A)** or a RipBO-3HA **(B)** fusion protein. Secretion assays were performed and

556 total proteins from bacterial pellets and proteins in the supernatants were detected by Western-

557 Blot.

558

559 **Figure 3. *mutS* alignment and phylogenetic tree on the set of 84 different strains.** A

560 neighbor-joining tree was build using the *mutS* from *Ralstonia pickettii* as an outgroup. Bootstrap

561 were performed on 100 replicates, only support higher than 50% displayed in the consensus tree.

562

563 **Figure 4. Venn diagram of conserved T3Es among different sets of "host-of-isolation"**

564 **defined strains. (A)** comparisons of conserved T3Es among "TOM" (host of isolation : tomato),

565 "EGG" (Host of isolation: eggplant) and "POT" (host of isolation : potato). **(B)** Comparison

566 between "TOM", "POT" and "BAN" (host of isolation: banana). The lists of compared T3Es are

567 visible in Table 2.

568

569 **Figure 5. Venn diagram of conserved T3Es among the different phylogenetic clades of**

570 **strains.** Comparison of conserved T3Es between *R. pseudosolanacearum* (phylotypes I and III),

571 *R. solanacearum* (phylotypes IIA and IIB), and *R. syzygii* (Phylotype IV) strains. The lists of

572 compared T3Es are visible in Table 3.

573

574 **Supplemental Figure S1. *mutS* alignment and phylogenetic tree on the set of 140 different**

575 **strains.** A neighbor-joining tree was build using the *mutS* from *Ralstonia pickettii* as an

576 outgroup. Bootstrap were performed on 100 replicates, only support higher than 50% displayed

577 in the consensus tree.

578

579 **Supplemental Figure S2. Uncropped western-blot version of Figure 2.**

580

Table 1(on next page)

8 new T3E and 2 new Hypothetical T3E identified

Table 1. 8 new T3E and 2 new Hypothetical T3E identified

Proposed T3E family name	Representative gene member	Hop/Xop homologues	Functional domain	Evidence for T3SS-dependent secretion or translocation
RipBJ	GMI1000 RSp0213	none		Lonjon et al. 2016
RipBK	YC45_c025370	HopAM1		Chang et al. 2005
RipBL	YC45_m001910	HopAO1	Protein-tyrosine phosphatase	Chang et al. 2005
RipBM	Psi07 RSPsi07_1860 (former Hyp15)		Protein-Ser/Thr kinase	This work
RipBN	CMR15v4_30917	AvrRpt2	cysteine protease	Eschen-Lippold et al. 2016
RipBO	GMI1000 RSc3174 (former Hyp16)	none		This work
RipBP	OE1-1_24290	HopW1 + homologs in Xantho	N-term domain=HopW1 and C-term= uncharacterized protein ABJ99_3552 [Pseudomonas syringae pv. cilantro]	Zumaquero et al. 2010
RipBQ	KACC10722_38580	HopK1/XopAK		Chang et al. 2005
Hyp17	RS244_m000380	none		This work
Hyp18	CMR15v4_mp10535	none		This work

Table 2(on next page)

List of Type III effectors conserved according to the host of isolation.

The # strains, indicate the total number of strains analysed and the 5% tolerance. The numbers in each cell indicate how many strains actually harbor the cognate effector. Grey scale according to conservation between columns.

Table 2. List of Type III effectors conserved according to the host of isolation. The # strains, indicate the total number of strains analysed and the 5% tolerance. The numbers in each cell indicate how many strains actually harbor the cognate effector. Grey scale according to conservation between columns.

	Eggplant	Tomato	Banana	Potato	Solanacea
# strains	9	14-15	14-15	28-30	55-58
RipA2	9	15		30	58
RipA3		15		28	55
RipA4		14			
RipA5		14	14	30	55
RipB	9	15	15	30	57
RipC1		14	15		
RipD			15		
RipE1	9	14	15		
RipE2			15		
RipF1		15	14	29	
RipG2	9	14			
RipG3			14		
RipG4	9				
RipG5	9	15	14	30	58
RipG6	9	14	15	30	57
RipG7		15		28	55
RipH1	9	14	14		
RipH2	9	15	15	30	58
RipH3	9	15		28	56
RipI		14	15		
RipJ	9				
RipL	9				
RipM	9	15		28	55
RipN	9	15			55
RipO1	9				
RipQ	9				
RipR	9	15		30	58
RipS1	9				
RipS2		15			
RipS3		15			

RipS4		**14**			
RipS5	**9**	**14**	**14**	**29**	**55**
RipS6	**9**				
RipU	**9**	**15**		**30**	**58**
RipV1	**9**	**15**	**15**	**28**	**56**
RipW	**9**	**15**	**15**	**29**	**57**
RipX	**9**	**15**		**29**	**57**
RipY	**9**	**14**			
RipZ	**9**	**15**		**29**	**57**
RipAA	**9**	**14**		**29**	**55**
RipAB	**9**	**15**	**14**	**30**	**58**
RipAC	**9**	**15**	**15**		
RipAD		**14**		**29**	
RipAE	**9**	**15**			
RipAF1	**9**				
RipAI	**9**	**15**	**15**	**30**	**58**
RipAJ	**9**	**15**	**15**	**30**	**58**
RipAL				**30**	
RipAM	**9**	**15**		**30**	**58**
RipAN		**15**	**15**	**30**	**57**
RipAO	**9**	**15**	**15**	**29**	**57**
RipAP	**9**				
RipAQ	**9**	**15**		**28**	**56**
RipAR		**14**	**14**		
RipAS	**9**				
RipAT			**15**		
RipAU			**14**		
RipAV	**9**				
RipAW		**14**			
RipAX1			**14**		
RipAY		**14**	**15**	**29**	
RipAZ1	**9**				
RipBA	**9**				
RipBM	**9**				
RipTAL	**9**				
RipTPS	**9**	**14**		**29**	**55**
Total CoreT3Es	**44**	**44**	**27**	**30**	**27**

1

Table 3(on next page)

List of Type III effectors conserved according to the phylogenetic origin.

The # strains, indicate the total number of strains analysed and the 5% tolerance. The numbers in each cell indicate how many strains actually harbor the cognate effector. Grey scale according to conservation between columns.

Table 3. List of Type III effectors conserved according to the phylogenetic origin. The # strains, indicate the total number of strains analysed and the 5% tolerance. The numbers in each cell indicate how many strains actually harbor the cognate effector. Grey scale according to conservation between columns.

	Phylotype I and III	Phylotype IIA and IIB	Phylotype IV	"stringency 2"
# strains	**36-38**	**24-25**	**20-21**	**80-84**
RipA2	**38**	(21)	**21**	**80**
RipA3	**36**		**21**	
RipA5			**21**	
RipB	**37**	**25**	**21**	**83**
RipC1		**25**		
RipD			**20**	
RipE1		**24**		
RipE2		**24**		
RipF1		**25**	**20**	
RipF2		**24**		
RipG2	**36**			
RipG4	**38**	**24**		
RipG5	**38**	(23)	**21**	**82**
RipG6	**37**	(23)	**21**	**81**
RipG7	**37**	**24**		
RipH1			**20**	
RipH2	**36**	**25**	**21**	**82**
RipH3	**36**			
RipH4			**20**	
RipI		**25**		
RipL	**38**			
RipM			**21**	
RipN			**20**	
RipO1		**24**		
RipQ	**38**			
RipR	**38**	**24**	**20**	**82**
RipS2	**36**			
RipS4	**38**			
RipS5			**21**	
RipS6	**37**			

RipU	**38**	**25**	(18)	**81**
RipV1	**38**	**25**	(19)	**82**
RipV2			**20**	
RipW	**37**	**25**	**21**	**83**
RipX			**21**	
RipY			**20**	
RipZ	**38**		**21**	
RipAB	**37**	**24**	**21**	**82**
RipAC		**25**		
RipAD		**25**		
RipAE		**24**		
RipAF1	**36**			
RipAI	**36**	**25**	**21**	**82**
RipAJ	**38**	**25**	**21**	**84**
RipAK	**36**			
RipAM	**38**	(23)	**21**	**82**
RipAN	(35)	**25**	**21**	**81**
RipAO	**36**	**25**	**21**	**82**
RipAP	**36**	**25**		
RipAQ			**21**	
RipAS	**36**			
RipAU			**20**	
RipAV	**36**			
RipAY	(35)	**25**	**20**	**80**
RipAZ1	**36**		**20**	
RipBF			**20**	
T3E_Hyp1			**20**	
Total Core T3Es	**31**	**25**	**32**	**16**

1

Figure 1

T3E prediction pipeline

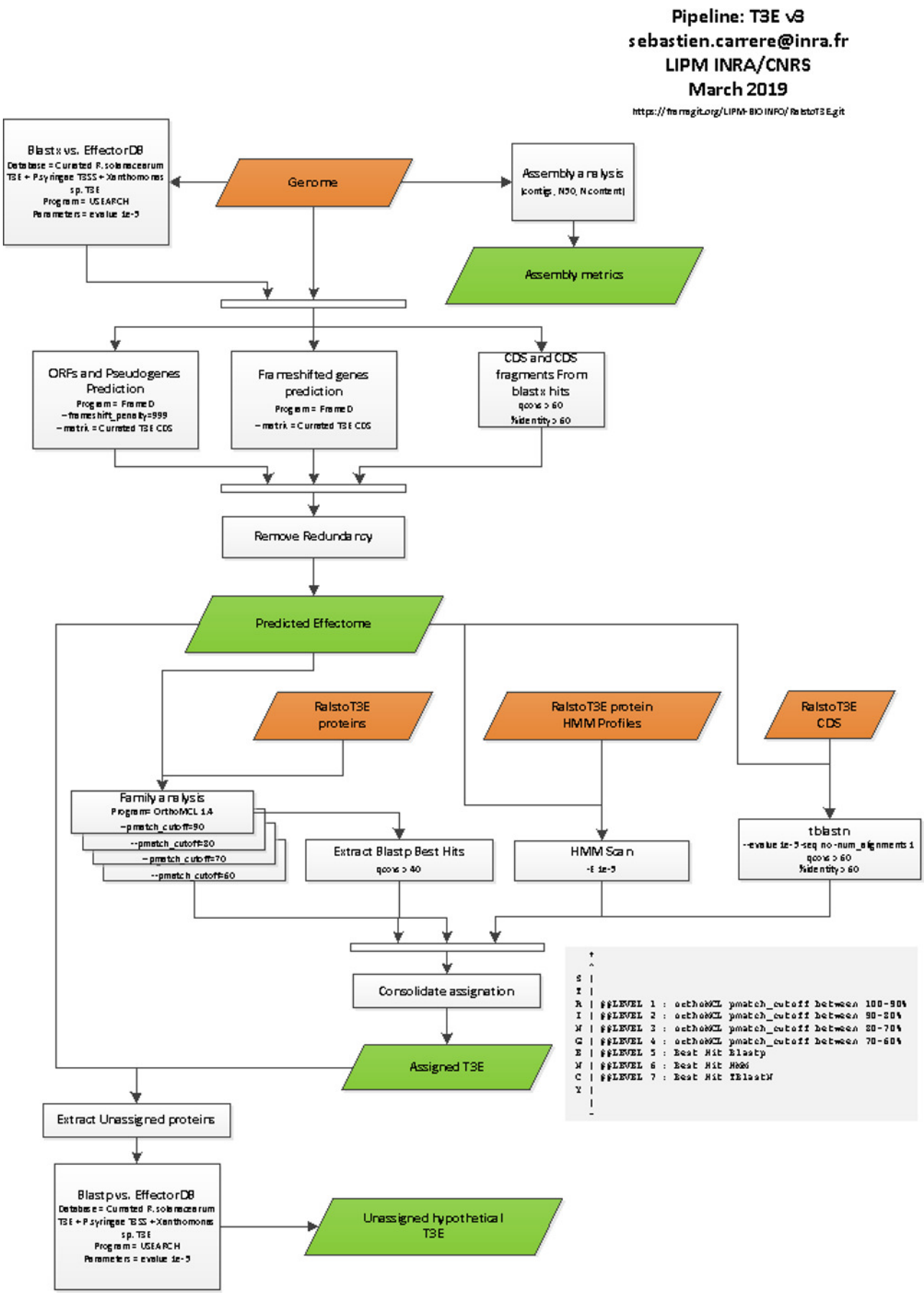

Pipeline: T3E v3
sebastien.carrere@inra.fr
LIPM INRA/CNRS
March 2019
https://forgemia.inra.fr/LIPM-BIOINFO/RalstoT3E.git
Blastx vs. EffectorDB
Database = Curated R.solanacearum T3E + P.syringae T3SS + Xanthomonas sp. T3E
Program = USEARCH
Parameters = evalue 1e-5
Genome
Assembly analysis
(contigs, N50, N content)
Assembly metrics
ORFs and Pseudogenes Prediction
Program = FrameD
–frameshift_penalty=999
–matrix = Curated T3E CDS
Frameshifted genes prediction
Program = FrameD
–matrix = Curated T3E CDS
CDS and CDS fragments From blastx hits
qcovs > 60
%identity > 60
Remove Redundancy
Predicted Effectome
RalstoT3E proteins
RalstoT3E protein HMM Profiles
RalstoT3E CDS
Family analysis
Program= OrthoMCL 1.4
–pmatch_cutoff=90
--pmatch_cutoff=80
–pmatch_cutoff=70
--pmatch_cutoff=60
Extract Blastp Best Hits
qcovs > 40
HMM Scan
-E 1e-5
tblastn
--evalue 1e-5 -seg no -num_alignments 1
qcovs > 60
%identity > 60
Consolidate assignation
Assigned T3E
$$LEVEL 1 : orthoMCL pmatch_cutoff between 100-90%
$$LEVEL 2 : orthoMCL pmatch_cutoff between 90-80%
$$LEVEL 3 : orthoMCL pmatch_cutoff between 80-70%
$$LEVEL 4 : orthoMCL pmatch_cutoff between 70-60%
$$LEVEL 5 : Best Hit Blastp
$$LEVEL 6 : Best Hit HMM
$$LEVEL 7 : Best Hit TBlastN
Extract Unassigned proteins
Blastp vs. EffectorDB
Database = Curated R.solanacearum T3E + P.syringae T3SS + Xanthomonas sp. T3E
Program = USEARCH
Parameters = evalue 1e-5
Unassigned hypothetical T3E

Figure 2

RipBM and RipBO are secreted through the T3SS

The wild-type strain and the *hrcV*mutant were transformed to express a $RipBM_{CMR15}$-3HA, $RipBM_{Psi07}$-3HA **(A)** or a RipBO-3HA **(B)**fusion protein. Secretion assays were performed and total proteins from bacterial pellets and proteins in the supernatants were detected by Western-Blot. Uncropped western-blot are displayed on Fig. S2

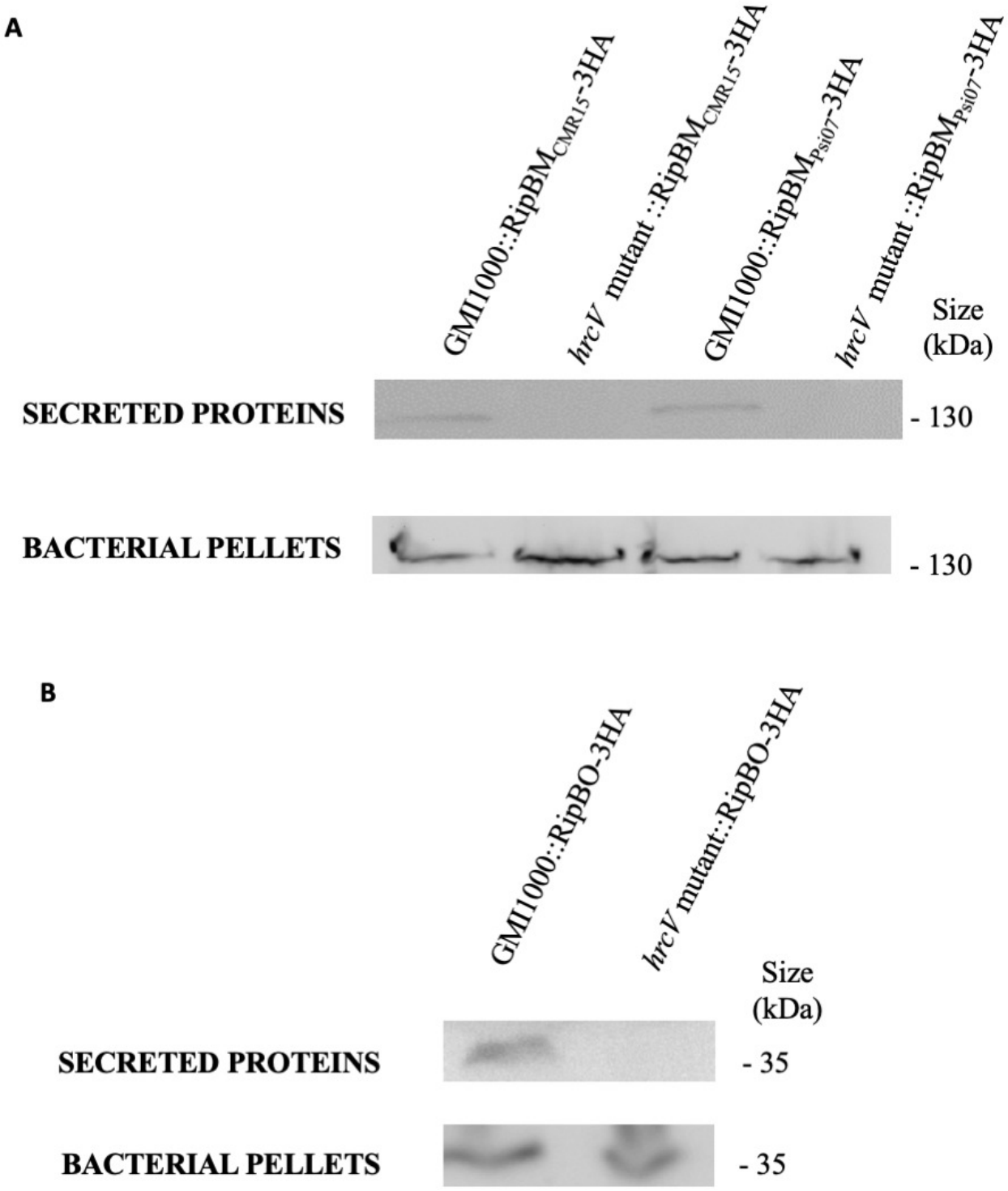

A
GMI1000::RipBM$_{CMR15}$-3HA
hrcV mutant ::RipBM$_{CMR15}$-3HA
GMI1000::RipBM$_{Psi07}$-3HA
hrcV mutant ::RipBM$_{Psi07}$-3HA
Size (kDa)
SECRETED PROTEINS
- 130
BACTERIAL PELLETS
- 130
B
GMI1000::RipBO-3HA
hrcV mutant::RipBO-3HA
Size (kDa)
SECRETED PROTEINS
- 35
BACTERIAL PELLETS
- 35

Figure 3

mutS alignment and phylogenetic tree on the set of 84 different strains.

A neighbor-joining tree was build using the *mutS* from *Ralstonia pickettii* as an outgroup. Bootstrap were performed on 100 replicates, only support higher than 50% displayed in the consensus tree.

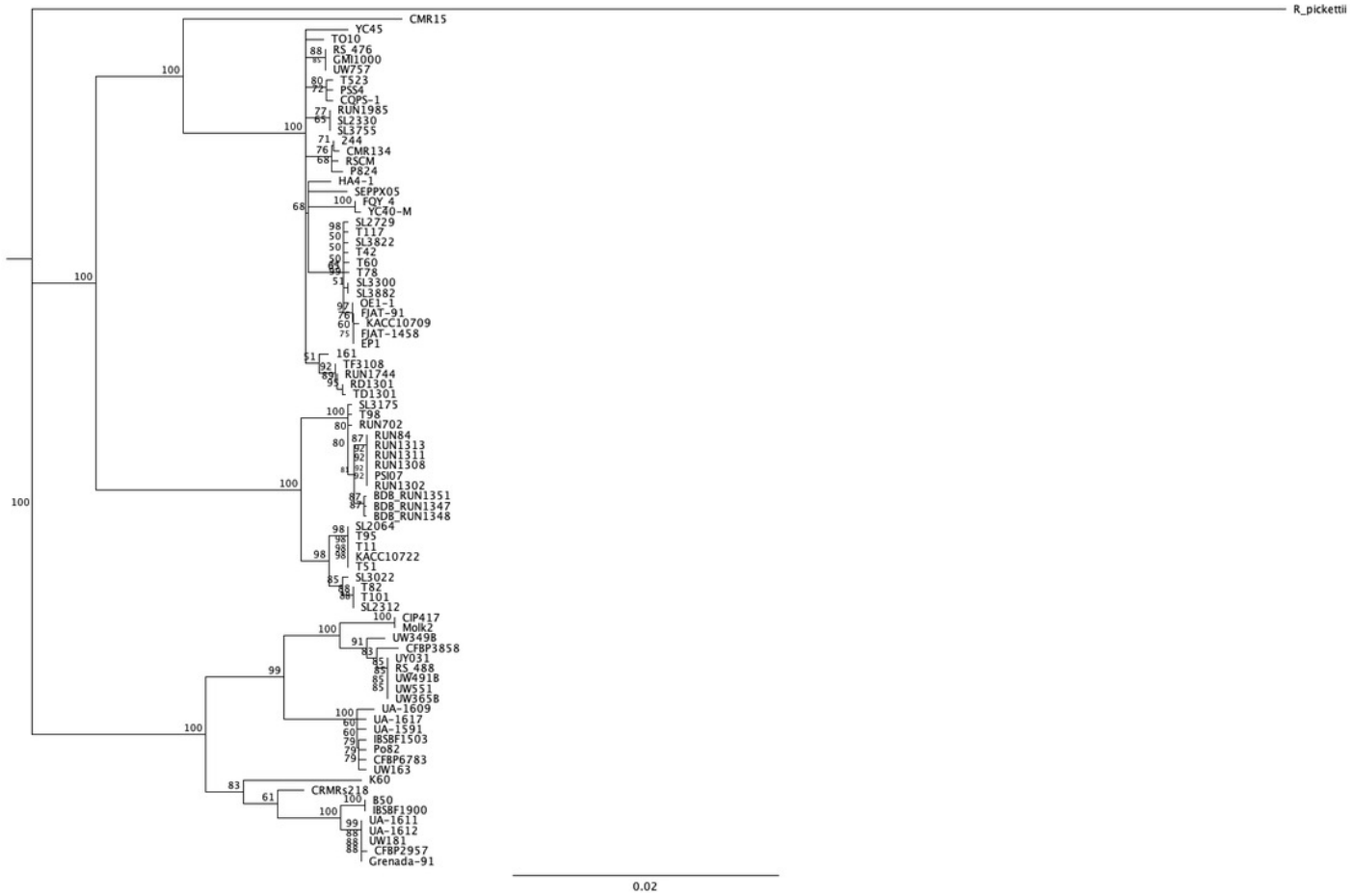


Figure 4

Venn diagram of conserved T3Es among different sets of “host-of-isolation” defined strains

(A) comparisons of conserved T3Es among “TOM” (host of isolation : tomato), “EGG” (Host of isolation: eggplant) and “POT” (host of isolation : potato). **(B)**Comparison between “TOM”, “POT” and “BAN” (host of isolation: banana). The lists of compared T3Es are visible in Table 2.

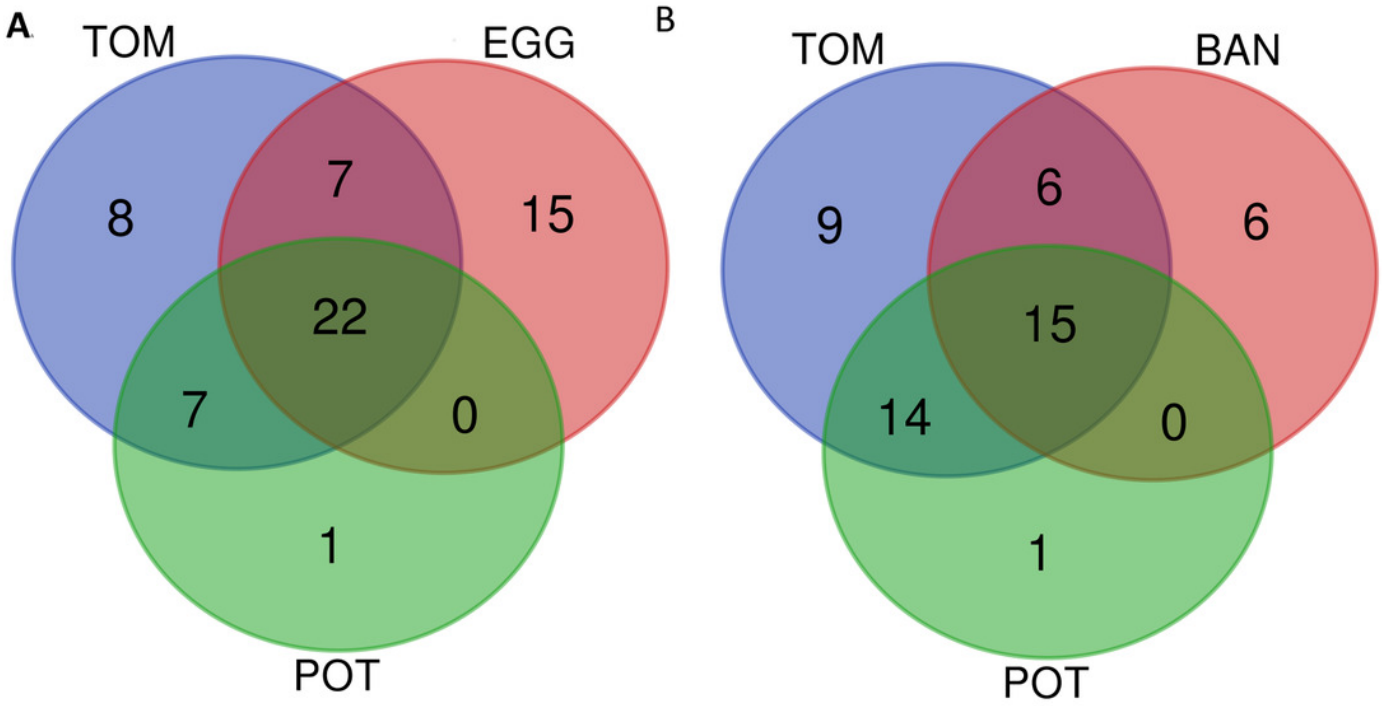
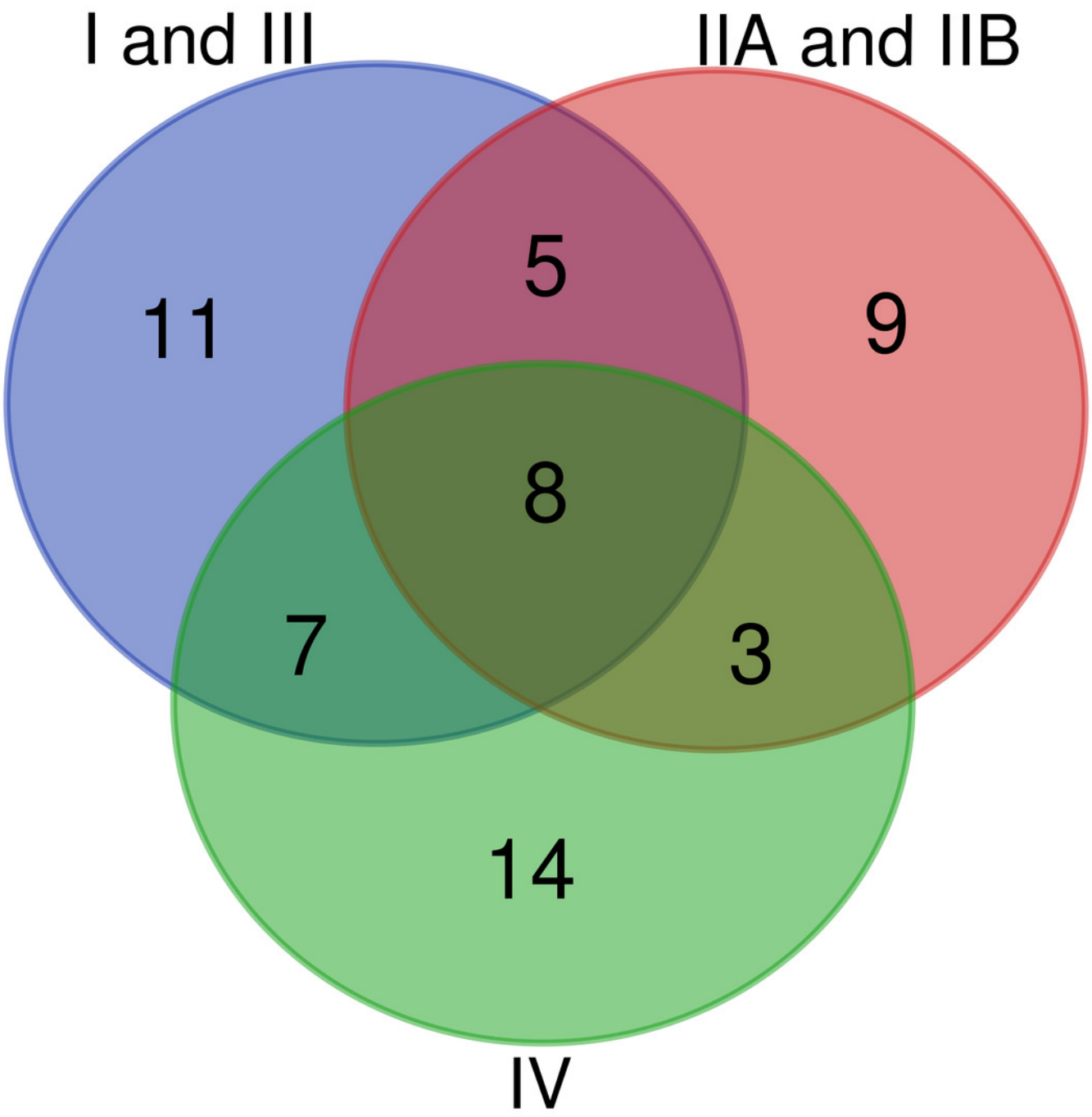


Figure 5

Venn diagram of conserved T3Es among the different phylogenetic clades of strains.

Comparison of conserved T3Es between *R. pseudosolanacearum*(phylotypes I and III), *R. solanacearum* (phylotypes IIA and IIB), and *R. syzygii* (Phylotype IV) strains. The lists of compared T3Es are visible in Table 3.


I and III
IIA and IIB
11
5
9
8
7
3
14
IV