Genome-enabled phylogeographic investigation of the quarantine pathogen

*Ralstonia solanacearum* race 3 biovar 2 and screening for sources of resistance against its core effectors

Christopher R. Clarke1*, David J. Studholme2*, Byron Hayes1, Brendan Runde1, Alexandra Weisberg1, Rongman Cai1, Tadeusz Wroblewski3, Marie Christine Daunay4, Emmanuel Wicker5, Jose A. Castillo6, Boris A. Vinatzer1

1 Department of Plant Pathology, Physiology, and Weed Science, Virginia Tech,
Latham Hall, Ag Quad Lane, Blacksburg, VA, USA;
2 Biosciences, University of Exeter, Exeter, Devon, UK
3 Genome Center and Department of Plant Sciences, University of California, Davis,
California 95616;
4 Unité de Genetique et Amelioration des Fruits et Legumes, INRA, Centre
d’Avignon, Montfavet, France;
5 CIRAD, UMR Peuplements Vegetaux et Bioagresseurs en Milieu Tropical (PVBMT), Saint Pierre, La Reunion, France
6 PROINPA Foundation, Cochabamba, Bolivia
* The first two authors provided equal contributions

Author for correspondence: Boris A. Vinatzer, Tel: +1 540 231 2126, Email:
vinatzer@vt.edu
Phylogeographic studies inform about routes of pathogen dissemination and are instrumental for improving import/export controls to limit pathogen spread. Genomes of seventeen isolates of the bacterial wilt and potato brown rot pathogen *Ralstonia solanacearum* race 3 biovar 2 (R3bv2), a select agent in the USA, were thus analyzed to get insight into the phylogeography of this pathogen. Most isolates were found to belong to the same genetic lineage that was probably imported from South America into Africa and Europe. At least one more isolate was imported into Europe in a separate event. Moreover, a conserved repertoire of 31 type III-secreted effector genes was identified in all R3bv2 isolates. These genes are excellent candidates to be targeted with plant resistance genes in breeding programs to develop durable disease resistance. Towards this goal, 17 core effectors were tested in eggplant, tomato, pepper, tobacco, and lettuce for induction of a hypersensitive-like response indicative of recognition by cognate resistance genes. Eight of these effectors triggered a response in one or more plant species. These genotypes may thus harbor resistance genes that could be identified, mapped, cloned and expressed in tomato or potato, for which sources of genetic resistance against R3bv2 are extremely limited.
INTRODUCTION

There is no better disease control than disease prevention. Therefore, keeping pathogens out of an unaffected geographic area or eradicating recently introduced pathogens is paramount. To do this efficiently, it is necessary to understand how pathogens are disseminated. Phylogeography, i.e., inferring past migration events by combining phylogenetic reconstruction with geographic information, is instrumental to gain this understanding. Once a pathogen is introduced into a geographic area, disease control becomes necessary. Unfortunately, many crops are affected by pathogens against which no high-yielding, high-quality cultivars are resistant. In addition, complex genetics and heterozygous status of clonally propagated crop species, such as potato, make it extremely difficult to introduce resistance genes by classical breeding from wild resistant relatives (33). Yet, utilization of resistance genes provides a very effective strategy for crop protection, particularly against bacterial plant pathogens for which chemical control and agricultural practices are often only temporary and ineffective solutions (59).

Today, genome sequencing in combination with biotechnology has the potential to greatly improve diagnostics and global tracking of high-risk pathogens and to accelerate the creation of durable disease resistant cultivars. Here we show progress using such an approach for Ralstonia solanacearum (Rs), a beta-proteobacterium that enters plant roots from the soil through natural openings and wounds, enters and colonizes the xylem where it blocks flow of water and nutrients, and finally causes wilting and eventual death of the host (13). Rs pathogenicity is based on a large repertoire of virulence genes encoding type II-secreted extracellular enzymes and type III-secreted effector (T3E) proteins (23), which in some plant genotypes trigger immunity and thus contribute to host range (51). Rs overall has a
very large host range, encompassing more than 54 plant families (29). Rs thus constitutes a major biotic constraint in tropical and subtropical areas throughout the world, both on subsistence crops (tomato, pepper, eggplant) and cash crops (potato, banana, ginger) (16).

Rs race 3 biovar 2 (R3bv2) is a sub-group of Rs that causes bacterial wilt on many solanaceous species. Because R3bv2 causes potato tubers to rot, the disease is also called “brown rot”. Based on phylogeny of the egl gene (17), most R3bv2 fall within phylootype IIB sequevar 1 (IIB-1) although a few isolates belong to other phylootype II sequevars (9). Based on multilocus sequence analysis, R3bv2 isolates fall in clade 5 (69). Contrary to most Rs strains, which are adapted to hot humid climates, R3bv2 strains are able to cause disease at cool temperatures (43). Consequently, this pathogen constitutes a major problem for subsistence potato farmers in many mountainous regions of Africa and South America. In the Mediterranean region, R3bv2 was first identified in the 1920s, in Egypt, Italy and Spain. From there it probably spread with contaminated early ware potatoes to northern Europe where it started to cause frequent outbreaks in the 1990s after being disseminated via contaminated seed potatoes (25, 32). It is now endemic in many waterways and on the Woody Nightshade (Solanum dulcamara) in the Netherlands (58), France (63), the UK (49), Spain (6, 36) and Portugal (11). R3bv2 is currently not present in North America although it was accidentally imported at least twice on geranium cuttings from Kenya and Guatemala, but in both cases it was possible to eradicate it before it became endemic (18, 70). Because of its potentially devastating effect on the US potato industry, R3bv2 is on the USDA – APHIS list of select agents (8, 36) and as such under stringent quarantine regulation.
The limited genetic diversity of R3bv2 strains identified by PCR-RFLP, AFLP, 16S rRNA, and 16S-23S suggests a recent evolutionary origin of R3bv2 (52, 62). However, pulsed field gel electrophoresis and rep-PCR (54), microarrays (9), and Multilocus Variable Number Tandem Repeat Analysis (MLVA) (46) provided enough resolution to distinguish R3bv2 isolates into separate populations, some of which were found to be associated with different geographic regions. Single nucleotide polymorphisms (SNPs) have been successful in reconstructing detailed phylogeographies of several human pathogens, for example of the plague pathogen Yersinia pestis (45), because they allow determination of ancestral versus derived relationships. SNPs have also been used with some success in resolving the phylogeography of plant pathogenic bacteria, for example of Pseudomonas syringae pv. actinidiae (42), but they have not been used yet for phylogeographic studies of Rs. Unfortunately, there is not a single potato cultivar resistant to any Rs strain (19). While some wild potato relatives are resistant, for example S. commersonii (53), resistance is difficult to introduce into cultivated potato because of different ploidy levels (33). Moreover, R3bv2 has been found virulent and very aggressive on most bacterial wilt-resistant tomato genotypes that contain genes for quantitative resistance (5, 67) (38, 68).

The best-characterized gene that provides resistance to Rs is the Arabidopsis thaliana RRS1-R gene, which recognizes the T3E PopP2 (14, 15), recently renamed RipP2 (50). In crop species, the only dominant genes for resistance to Rs have so far been mapped in eggplant (37, 72) and indication for a resistance gene to an Rs effector has also been found in the wild eggplant relative Solanum torvum (48). Moreover, Lebeau and colleagues (38) found eggplant germplasm to encompass some remarkable genotypes with resistance to multiple Rs strains, including R3bv2.
Because tomato, potato, and eggplant all belong to the same botanical family, expression of disease resistance genes from eggplant in potato or tomato will likely provide resistance, due to compatibility of signal transduction components (31, 61). Therefore, identifying the underlying resistance genes in these eggplant genotypes is paramount.

We have shown in the past that using Agrobacterium-mediated transient assays to express individual T3Es in plants and screening for a hypersensitive response (HR) – like plant reaction is an efficient way to detect the presence of potential resistance genes (66, 71). These genes could then be genetically mapped, cloned and expressed in crop cultivars. The problem is that pathogens can overcome a single resistance gene relatively easily by losing or mutating the effector that is being recognized (22, 35). However, Bart and colleagues used the example of Xanthomonas axonopodis pv. manihotis to show that by sequencing a large number of strains it is possible to identify the core set of T3Es of a pathogen (2). The assumption is that effectors present in most strains of a pathogen play an essential role in pathogen virulence and thus cannot be easily lost by the pathogen after a new resistance gene is employed in the host. Resistance genes that recognize such conserved effectors are thus expected to be more durable than resistance genes that recognize dispensable effectors.

Here, we analyzed the genomes of seventeen R3bv2 isolates collected over 58 years on different continents to get insight into R3bv2 phylogeography and to identify the R3bv2 core effector repertoire to target in crop breeding. We previously tested putative R3bv2 effectors in transient assays on lettuce, tomato, pepper, and tobacco genotypes to survey them for potential resistance genes (71). Here we integrate these results with new results obtained with two eggplant genotypes, one of which is
resistant to R3bv2 (37), to prioritize effectors to target in breeding for bacterial
tilt/brown rot resistance

MATERIALS AND METHODS

Bacterial Growth and DNA extraction
Bacterial strains were stored in Cryobank™ beads at -80°C. They were grown first on
Nutrient Broth overnight at 28°C, then streaked (50 µL) on Kelman’s triphenyl
tetrazolium chloride (TZC) agar medium (34) supplemented with 1% yeast extract,
and sub-cultured two days at 28°C. DNA was extracted from fresh cultures (~ 1.0 to
2.0 x 10^9 cells), using the QIAamp “DNA Blood and Tissue Kit” (QIAGEN), and
checked for titer and quality using a Nanodrop device.

Genome Sequencing
Paired-end Illumina sequencing was performed for 76 cycles on an Illumina GAIIX
with PE module at the Center for the Analysis of Genome Evolution and Function,
University of Toronto, Toronto, Ontario, Canada as previously described (42). 454
sequencing was performed on a GS FLX sequencer with Titanium chemistry and
sequencing reads were assembled with Newbler v. 2.5 at the Virginia Bioinformatics
Institute, Blacksburg, Virginia, USA. Sequencing details are listed in Table 2.

SNP Identification
To facilitate identification of single-nucleotide polymorphisms, we aligned the
sequence reads (454 or Illumina – see Table 2) against the reference genome sequence
of R. solanacearum strain Po82 (73) using BWA-mem version 0.7.5a-r405 (39) with
default parameter values and excluding any reads that did not map uniquely to a
single site on the reference genome. From the resulting alignments, we generated pileup files using SAMtools version 0.1.19-96b5f2294a (40). We then parsed the pileup-formatted alignments to examine the polymorphism status of each single-nucleotide site in the entire 5,430,263-bp reference genome. For each single-nucleotide site we categorized it as either ambiguous or unambiguous. A site was considered to be unambiguous only if there was at least 5X coverage by genomic sequence reads from each and every bacterial isolate and only if for each and every bacterial isolate, at least 95% of the aligned reads were in agreement. Any sites that did not satisfy these criteria were considered to be ambiguous and excluded from further analysis. Over the remaining unambiguous sites, we could be very confident in the genotype for all the sequenced isolates.

The sequence reads for several genomes that had been previously published by other laboratories were not publicly available at the time of this study. Therefore, we generated 50X coverage of simulated ‘pseudo-reads’ and aligned these to the reference genome by the same method. The genomes for which we had to resort to using pseudo-reads were the genomes of the R3bv2 isolates UW551 and IPO1609 and the following non-R3bv2 genomes: IPO1609, CMR15, FJAT-1458, FJAT-91, FQY_4, GMI1000, MolK2, NCPPB909, P673, Po82, PSI07, R229, R24, Rs-09-161, Rs-10-244, SD54, UW551 and Y45. The pseudo-reads comprised randomly selected 50-bp substrings from the genome sequence; substitution errors were introduced randomly into the pseudo-reads at a rate of 0.1% per base-pair. Clearly there is a significantly higher risk of false-positive variants being called based on these pseudo-reads since the use of pseudo-reads derived from a single consensus genome assembly does not benefit from the alignment of multiple independent sequence reads as is the case for the Illumina and 454 reads. The accuracy of these previously published
genome assemblies (and hence the pseudo-reads derived from them) is limited by the
depth and reliability of the Sanger sequence reads used to assemble them and by the
accuracy of the assembly process. Therefore, we might expect to observe an elevated
number of private variants in these bacterial isolates due to the inevitable presence of
effects in their genome sequence assemblies.

Construction of phylogenetic trees
We used the results of the genome-wide SNP-calling to investigate the phylogenetic
relationships among the sequenced bacterial isolates. The SNP sites were
concatenated together to generate a multiple sequence alignment, on which was based
a maximum likelihood (ML) phylogenetic tree, built using RAxML (57). Twenty ML
searches were performed for each along with 1200 non-parametric bootstrap
replicates using the ASC_GTRGAMMA model. Bootstrap values as a percentage of
1200 replicates were mapped to branches on the best ML tree, and branches with less
than 50% bootstrap support were collapsed into polytomies using TreeCollapseCL4
(30).

Identification of effector repertoires
We checked for presence of each of the 60 predicted effector genes previously
catalogued (50) in Rs UW551 (21) by aligning sequence reads against the gene
sequences using BWA-mem version 0.7.5a-r405 (39) with default parameter values
(allowing non-uniquely mapping reads to match multiple genomic sites). From the
resulting alignments, we generated pileup files using SAMtools version 0.1.19-
96b5f2294a (40) and parsed these pileup files to determine the proportion of the gene
that was covered by aligned sequence reads (that is, the breadth of coverage for each
gene). To check for frame-shifts and premature stop-codons, we generated consensus sequences for each gene for each sequenced isolate (based on the pileup files). We then checked whether each consensus effector gene sequence by translating from nucleotide into amino acid sequence using BioPerl (56).

Agrobacterium-mediated transient expression of effectors

Eggplant seeds were sowed in Sunshine 1 Soil mix (company, location) and grown at 22°C on a 16 hour day cycle in Conviron (Henderson, North Carolina) PGC20 chamber. Other plant species and cultivars tested were grown in the greenhouse condition as described previously (71). Five to seven weeks old eggplants, two to three weeks old lettuce plants, and four to six weeks old tomato, pepper and *N. benthamiana* plants were used for transient assays.

Thirty-two effectors from *R. solanacearum* strain UW551 that were previously cloned into the vector pBAV150 (66) under control of a dexamethasone-inducible promoter were transformed via conjugation into *Agrobacterium tumefaciens* strain C58 (28). *Agrobacterium*-mediated transient expression was performed similar to previously described protocols (71). Briefly, *Agrobacterium* was inoculated into liquid Luria-Bertani Broth with antibiotics and 200uM Acetosyringone and incubated at 28°C at 200RPM for 18 hours in 50ml conical tubes. Cells were harvested via centrifugation at 4000 RPM. Cells were resuspended in sterile H2O and centrifuged to pellet a second time. Cells were resuspended in sterile H2O again and adjusted to an approximate optical density at 600nm (OD600) of 0.5-1 and acetosyringone was added to a final concentration of 200uM and incubated at 28°C at 200RPM for an additional 4-6 hours. The OD600 was adjusted to 0.3 and the adaxial side of leaves were infiltrated with a 1ml blunt-end syringe. Two days after infiltration, infiltrated
leaves were sprayed with 30uM Dexamethasone. Five days after infiltration, the severity of induced HR-like symptoms was scored.

RESULTS

A genome-based phylogeny and estimation of the time since the most recent common ancestor for R3bv2

To investigate phylogeny and phylogeography of R3bv2, we chose to sequence the genomes of fourteen R3bv2 isolates (Table 1). The isolates were chosen to maximize the time span of isolation (from 1950 to 2008), geographic location (South America, Africa, Europe, and Asia), and genetic diversity based on previously obtained microarray analysis (9). The publicly available genomes of the R3bv2 isolates UW551 (21), IPO1609 (27), and NCPPB909 (GenBank: JNGD0000000.1, unpublished) were also included in the analysis. Details of the employed sequencing technology and the obtained genome coverage for each genome are listed in Table 2.

To start investigating the phylogeny of R3bv2 strains, we first placed them into a larger phylogenetic context by building a phylogenetic tree including publicly available representatives of the entire Rs species complex. The genome of strain Po82 isolated from potato but also pathogenic to other Solanaceae and to banana (73) was used as reference since Po82 is the strain most closely related to R3bv2 whose genome sequence is complete (rather than of draft quality). A total of 1,108,724 base-pairs (bp) of this genome were covered at least 5 times by sequencing reads with consistent base calls (or assembled genome sequences) from each of the sequenced genomes. Therefore, the tree (Figure 1A) was built based on this 20% of the genome for which the genotype could be unambiguously determined for all sequenced isolates. As expected, all newly sequenced R3bv2 isolates cluster closely together in a
clade shared with previously sequenced known members of phylotype IIB: the R3bv2 isolates UW551 (21), IPO1609 (27), and NCPPB909, strain Molk2 (causative agent of banana Moko disease), strain P673 isolated from the ornamental plant *Epipremnum aureum* (pothos) and also able to infect tomato and, to a lesser degree, potato at 18°C (4) and the reference strain Po82 (73).

Next, we built a tree of only R3bv2 genomes. For this tree, 2,006,873 positions were unambiguous (37% of the total Po82 genome), providing greater phylogenetic resolution than the previous tree. Moreover, SNPs were distributed quite evenly over both the chromosome and the megaplasmid (17,852 SNPs within the main chromosome and 7,136 SNPs within the mega plasmid). Importantly, the tree built on all SNPs (Figure 1B) and the trees built on SNPs from only the main chromosome had the same branching order and the tree built only on SNPs from the mega plasmid had a very similar topology (see Figure S1A and S1B) showing that the main chromosome and megaplasmid have similar evolutionary histories and, thus, the megaplasmid was not frequently exchanged between strains.

The South American isolates 23-10BR, NCPPB282 and Rs2 (isolated in Brazil, Colombia, and Bolivia, respectively), and the Dutch isolate CFBP3858 are on relatively long branches and are distinct from all other R3bv2 isolates, which instead form a tight cluster of highly similar isolates (See supplementary Table S1 for the exact number of pairwise SNPs for all R3bv2 genomes). Within this cluster of 13 isolates, 11 isolates collected in Europe, Africa, and Asia only differ at one to seven SNP loci from each other although they were collected over a time period of 50 years. We thus used the two most similar isolates collected the furthest apart in time (CFBP4808 isolated in 1954 and Pa1 isolated in 2004) to infer a minimal yearly mutation rate by dividing the number of mutations that distinguished these two
genomes by the difference in years of isolation divided by two (since both strains mutated subsequent to diverging from their most recent common ancestor, MRCA). Because only four mutations distinguished the two isolates over the entire length of the 2,006,873 bp that were compared, the inferred yearly mutation rate was $1.99 \times 10^{-8}$ mutations bp$^{-1}$ year$^{-1}$. Applying this rate to the two most divergent R3bv2 strains, 23-10BR and CFBP3858, which are separated by 5370 SNPs - the MRCA of R3bv2 was inferred to have possibly existed as long ago as 67,125 years ago.

Inferred Phylogeography of R3bv2

Within the tree of R3bv2 isolates (Figure 1B), all South American isolates (23-10BR, Rs2, and NCPPB282) are on three of the four most basal branches. Therefore, these isolates can be considered more ancestral than all other R3bv2 isolates, which is typical of strains at the geographic origin of a pathogen (24). Moreover, the relatively large genetic distance between the South American isolates (see Supplementary Table 1) is suggestive of relatively high diversity of R3bv2 in South America, also expected for the geographic origin of a pathogen (26, 60). Therefore, the analyzed genome sequences here are in agreement with the previously assumed South American origin of R3bv2 (13) and migration of R3bv2 out of South America to Africa, Europe, and Asia (Figure 2).

The next objective of this study was to infer the direction of dissemination between Europe, Africa, and Asia. Since most strains isolated in Europe, Africa, and Asia were highly similar to each other and are part of a single clade with a bootstrap value of 100 % (Figure 1B), a single ancestor of these isolates may have been exported from South America some time before 1954 (the year of isolation of the earliest strain, CFBP4808) and then exchanged between Europe, Asia, and Africa.
Interestingly, a small number of SNPs within this clade are informative and can provide some indication about the possible order of events during these exchanges (see also Figure 2). First, CFBP4808 (isolated in Israel in 1954) is located on the most basal branch of the clade because at one SNP locus it retains the ancestral (Po82-like) allele whereas all other isolates in the same clade carry the derived allele. CFBP4808 may thus represent a very early derivative of the R3bv2 strain exported from South America and from which most other European, African, and Asian isolates are derived from. Second, the two isolates from Reunion Island (LNPV28.23 and JT516) share alleles at one SNP locus with the Egyptian isolate CFBP4578, possibly suggesting a Mediterranean origin of the Reunion outbreak. Third and finally, the Dutch isolate IPO1609 and the Swedish isolate CFBP3884 share alleles at two SNP loci and may thus be both derived from a recent ancestor imported into Northern Europe.

The European isolate IPO1609 and the Kenyan isolate UW551 (imported into the USA on geranium) are the only two isolates within the major European, African, and Asian clade having more than 100 private single-nucleotide variants. However, there is 100% bootstrap support for their placement within the same clade, strongly suggesting that they share the same common ancestor exported out of South America with the other isolates in the clade. Possible explanations for this high number of variants private to these two isolates are discussed below.

The Dutch isolate CFBP3858 differs at more than 4000 SNP loci from all other European isolates and lies clearly outside of the main European, African, and Asian clade (Figure 1B). Its ancestor was thus most likely imported into Europe independently from the ancestor of all other analyzed European isolates.
R. solanacearum R3bv2 core effector repertoire

Repertoires of type III-secreted effectors are known to be important determinants of the host range of plant pathogens including Rs (50). We therefore sought to define the core effector repertoire of the R3bv2 group. To do this, we aligned the R3bv2 genomes sequenced here (Table 2) and the previously sequenced genomes of IPO1609 and NCPPB909 against 58 R. solanacearum effectors catalogued in the R3bv2 genome UW551 by Peeters and colleagues (50). Results are summarized in Table 3 and detailed coverage information for each effector in each genome is available in Table S2. We considered an effector to be present in a genome when sequencing reads from that genome covered more than 80% of the length of the UW551 ortholog. Using this cut-off, 31 effectors were found to be present in all analyzed R3bv2 genomes and not to have any frame shift or early stop codon in the gene regions covered by reads. We thus defined these 31 effectors as R3bv2 core effector repertoire. However, we may have slightly underestimated the core effector repertoire because some effectors may have less than 80% coverage due to sequence divergence or biased sequencing coverage of the genome and not because of actual absence of genes. On the other hand, some of the effector genes with sequencing coverage between 80% and 100% may contain stop codons or frame-shift in the regions of the effectors not covered by reads and therefore not detected. However, only two (out of 915) of the effector sequences with over 80% coverage had any observable frame-shift mutation or early stop codons in any of the 17 analyzed genomes: RipAX1, which is truncated in NCPPB282, and RipS3, which is truncated in Rs2. We therefore have confidence that most of the 31 core effectors are truly present and functional in the corresponding genomes.
The R3bv2 core effectors were then compared against the effector repertoires predicted for non-R3bv2 sequenced Rs genomes at the RalstoT3E database (50) revealing that 11 of the 31 R3bv2 core effectors represent core effectors of the entire Rs species complex: RipE1, RipF1, RipG6, RipH1, RipW, RipAB, RipAC, RipAI, RipAN, RipAR, and RipAY. Interestingly, none of the effectors was present exclusively in R3bv2.

Screening for resistance genes against R3bv2 effectors

Agrobacterium-mediated transient expression of T3E proteins in the leaves of plants carrying corresponding resistance gene results in the induction of defense responses usually associated with programmed cell death known also as hypersensitive response (HR) (3). Therefore, 27 R3bv2 effector genes, including 17 of the R3bv2 core effectors, were expressed in two eggplant genotypes, 20 tomato genotypes, 5 pepper genotypes, and 22 lettuce genotypes to determine the possible presence of resistance genes in these plant genotypes. An additional five genes that had been previously predicted to encode effectors but were not identified as such by Peeters and colleagues (50) were also included: Ucd163 (a predicted membrane-bend lytic murein transglycosylase B); Ucd190 (a predicted PopC-like putative type III effector similar to RipAC); Ucd191 (HpaP, a putative type III secretion substrate specificity switch protein which was recently found not to be secreted nor translocated (41); Ucd197 (a putative Type III effector); Ucd203 (a conserved protein of unknown function). All effectors had been cloned from an R3bv2 strain from Bolivia previously (71). While the results for ten of the analyzed effectors for tomato, pepper, N. benthamiana, and lettuce were already reported (71) and have been accessible in the CHARGE database
these results were analyzed here for the first time in the context of the R3bv2 core effectorome.

The overall severity of the cell death elicited by transient expression of the effectors across many replicates was demarcated into the broad categories of “Strong”, “Intermediate” and “None” (see Figure 3 for examples). Specifically, fifteen of the screened effectors elicited an HR-like plant response on at least one genotype tested suggesting the presence of underlying effector-detecting R genes (Table 4, see detailed results in Table S3). Eight of these fifteen effectors were core R3bv2 effectors: RipB1, RipE1, RipG6, RipH2, RipX, RipV1, RipV2, and RipAB. Therefore, the genotypes on which these effectors elicited cell death potentially harbor R genes capable of possibly providing durable resistance to R3bv2.

Of special interest, six effectors, including five core effectors (RipE1, RipH2, RipV1, RipV2, and RipX), elicited substantially stronger cell death in eggplant genotype MM853 than in genotype MM738 (Table 4). Because MM853 is resistant to R3bv2 whereas MM738 is susceptible (38), the eggplant genotype MM853 possibly has R genes that recognize one or more of these core effectors.

Discussion

Because of its impact on world-wide potato production (16) and because of its status as select agent in the USA (8), R3bv2 has been the focus of many studies. For example, the phylogenetic relationship between R3bv2 and other Rs strains has been determined (7, 9, 69), the genomes of a small number of R3bv2 strains have been sequenced (21, 27), and these genome sequences allowed the prediction of virulence genes (50). However, the geographic origin and intercontinental routes of dissemination of R3bv2 are not well known (9) and, even more importantly, the
identification of resistance genes against R3bv2 is still in its infancy (5, 37, 53). Here we made progress in these two areas.

An estimated timescale for R3bv2 emergence and diversification

While yearly mutation rates and divergence times for human pathogen species have recently been inferred, for example (12, 44, 47), our knowledge of divergence times for bacterial crop pathogens is preliminary at best (65) making it impossible to determine the time scale of bacterial crop pathogen evolution.

Mutation rates can only be inferred reliably if there is a statistically significant correlation between years of isolation and the genetic distance from the MRCA (47). Therefore, genomes of a number of isolates collected over many years need to be obtained whereby the collected isolates need to be closely related so that it can be assumed that the collected isolates in fact evolved from each other during the time frame of collection (instead of being derived from a distant ancestor with most of the mutations having occurred long before the isolates were collected, which would make it impossible to infer the number of mutations that accumulated per year).

Surprisingly, eleven closely related R3bv2 isolates collected over a time span of 50 years did not accumulate more than 7 mutations (in the 36% of the genome that could be compared with high confidence). Importantly, isolates collected in similar years were as different from each other as isolates collected 50 years apart (see Table 1 and Table S1). Therefore, a mutation rate could not be inferred based on the accumulation of mutations over time. Instead, a minimal mutation rate based on the most similar isolates collected 50 years apart was inferred and found to be as low as 1.99 x 10^-8 bp^-1 year^-1. This is lower than the mutation rates recently calculated for human pathogens (mostly in the range from 3x10^-5 to 3x10^-7; (44) and references
The reason could be that *Rs* bacteria may grow — and thus mutate - very slowly in soil and surface water populations, which often constitute the inoculum sources of outbreaks (1, 64) and most of the mutations that get fixed during fast pathogen replication during plant infection may constitute a sink rather than a source according to the model of Sokurenko and colleagues (55).

However, at this point we cannot exclude the possibility that we underestimated the mutation rate because of ascertainment bias arising from considering only the 36% of the genome for which we could be confident of the genotype in all our sequenced genomes. By definition, this 36% that is represented in all the sequenced genomes represents part of the core genome rather than the variable genome. Examining this portion of the genome has the advantage that it is less likely to have undergone horizontal transfer and thus reflects the ‘true’ phylogeny in addition to the advantage of high confidence in the nucleotide sequence in all isolates. However, if there is an over-representation of vertically inherited SNPs in the non-core (variable) portion of the genome, then conversely this implies SNPs are under-represented in the core-genome, including the 36% on which we based our phylogenetic analysis.

The high number of private variants in the UW551 and IPO1609 genomes may be due to sequencing errors in these two draft-quality genomes that were obtained by relatively low coverage Sanger sequencing (21, 27). This is in contrast to the situation for the genomes that were sequenced to greater depth by next-generation sequencing methods and for which accuracy benefited from consensus of multiple overlapping independent sequencing reads. Alternatively, UW551 and IPO1609 may be derived from outbreaks during which rates of bacterial replication were high.
The phylogeographic studies undertaken with the plague pathogen *Yersinia pestis* are an impressive example of how whole genome sequencing of bacteria allows reconstruction of historic routes of pathogen dissemination (45). While the number of analyzed R3bv2 isolates in our study is much lower than the number of *Y. pestis* isolates used in the above study, we can still make some conclusions by combining the microbial forensics evidence derived from our data with previous results about the origin and dissemination of R3bv2.

R3bv2 is generally assumed to have first emerged in the Andes of South America (13). In particular, French (20) reports that he found biovar 2 in virgin soil in the Amazon basin and he suggests that potatoes brought from the Peruvian highlands and cultivated in these soils centuries ago got infected there and that infected potatoes were brought back later to the Peruvian Highlands leading to the emergence of brown rot. From South America, R3bv2 is assumed to have been exported with potatoes to the rest of the world and molecular data in agreement with this hypothesis were obtained previously by pulsed field gel electrophoresis and rep-PCR (54), microarray analysis (9), and multilocus sequence analysis (69). However, these previous studies were either based on markers that do not allow determination of ancestral versus derived relationships (9, 54) or were based on a small number of loci (69). Here instead we analyzed over one third of the entire length of R3bv2 genomes using SNPs, which are informative of ancestral versus derived relationships. This approach identified much higher diversity of genomes of South American origin compared to genomes of European origin together with the phylogenetic tree in Figure 1B, which clearly shows that all but one of the European, African, and Asian isolates are derived from more ancestral South American isolates, is so far the strongest evidence of a
South American origin of R3bv2. Moreover, our estimated time since the MRCA of R3bv2 (over 50,000 years ago) clearly places the origin of R3bv2 before the domestication of potato and suggests that the ancestor of R3bv2 was a pathogen of wild potatoes or other Solanaceae.

In regard to the phylogeography of R3bv2 outside of South America, Cellier and colleagues (9) had found that R3bv2 strains were clustered geographically based on a phylogeny and population structure. One R3bv2 group was found to be associated with African and the Indian Ocean region, one group with the Mediterranean regions, and one group with Northern Europe. The northern European isolates IPO1609 (Netherlands) and CFBP3884 (Sweden) cluster together also based on SNPs. However, for the rest our data do not support geographic differentiation of populations. In particular, the Dutch isolates Pa1 and KRZ-5 do not share any alleles exclusively with IPO1609 or CFBP3884 supporting a Northern European population. We thus conclude frequent exchange of R3bv2 between regions, which is in agreement with the pulsed field gel electrophoresis and rep-PCR results that also revealed almost identical isolates from Northern Europe, the Mediterranean region, and Africa (54).

As for the temporal order of transfers, it is interesting to note that isolate CFBP4808 was collected in the Mediterranean region in 1954 (the earliest isolate outside of South America) and is at the same time also the most ancestral isolate of the main clade outside of South America. This agrees with the fact that R3bv2 caused outbreaks in the Mediterranean region starting in the 1920s and that R3bv2-contaminated potatoes were imported from that region into the rest of Europe, possibly being the source of the outbreaks observed in northern Europe starting in the 1980s (25, 32). Our data also support a Mediterranean origin of the Reunion Island
population. Since the only Chinese and the only Kenyan isolates sequenced here are also nested within the European clade, either a Mediterranean or Northern European origin of the R3bv2 populations in these geographic areas is likely as well (see Figure 2).

The core effector repertoire of R3bv2

Bart and colleagues suggested that presence of an effector across all isolates may indicate that the effector is essential for the pathogen to cause disease and that resistance genes against such pathogen core effectors may provide durable resistance (2). While this hypothesis still needs to be empirically tested, it is plausible. We thus determined here the core effector repertoire R3bv2 consisting of 31 effectors and the core effector repertoire of Rs overall consisting of 13 effectors. These core effectors can now be prioritized to identify corresponding resistance genes for breeding potato, tomato, and pepper cultivars with durable disease resistance. We expect the members of the R3bv2 core effector repertoire to be particularly important to cause disease on the common R3bv2 hosts potato and tomato while the Rs core effector repertoire should consist of the basic tool set used by Rs to cause disease on all its hosts.

Several plant genotypes contain putative resistance genes against core R3bv2 effectors

Unfortunately, most Rs effectors for which sources of resistance had been previously identified are missing from R3bv2, for example: RipAA and RipP1 (51)), RipP2 (15), and Rip36 alias RipAX2 (48). Therefore, the R3bv2 core effectors identified here need to be given priority in identifying new sources of resistance for breeding/engineering resistance to R3bv2.
Towards this goal, a collection of cloned R3bv2 effectors was previously used in Agrobacterium-mediated expression to identify potential sources of resistance among tomato, pepper, and lettuce accessions and *N. benthamiana* (71). We extended the search here to two eggplant genotypes: MM853 found to be resistant to most *Rs* strains, including R3bv2 strains, and the susceptible MM738 genotype (38).

Combining all results, an interesting picture arose. First of all, none of the five tested *Rs* core effectors triggered strong cell death in any of the tested genotypes: two *Rs* core effectors did not trigger any HR-like response in any genotype (RipAY and RipW), two triggered an intermediate response in only one plant genotype each (RipAB in the tomato accession M82 and RipG6 in the tomato accession V121), and RipE1 triggered an intermediate response in two genotypes (the resistant eggplant genotype MM853 and the tomato genotype VF36). This almost complete absence of HR-like responses to *Rs* core effectors suggests that these effectors have been maintained in all *Rs* strains because plants have not evolved resistance against them.

Six of the 12 tested R3bv2 core effectors did not trigger any response in any host genotype and two triggered only intermediate responses (RipV2 and RipX). However, RipH2, RipAT, RipV1, and RipBI triggered intermediate or even strong HR-like responses in multiple genotypes of several plant species, revealing that promising sources of resistance may be available for these R3bv2 core effectors (see Table S3 for the complete list).

Particularly promising is the result that the R3bv2 core effector RipH2 triggered a strong HR-like cell death response in the resistant genotype MM853 but not in the susceptible genotype MM738. RipH2 is a member of the HLK effector family and ripH1-2-3 alleles from the phylotype I strain OE1-1 were demonstrated to be functionally redundant and mildly involved in virulence on tomato and eggplant.
RipH2 is thus an excellent candidate effector that is possibly recognized by a cognate R gene in genotype MM853, which could be mapped in a cross with the susceptible genotype MM738, cloned, and expressed in tomato and/or potato to confer resistance to R3bv2 strains. The multiple lettuce and tomato accessions that also reacted with a strong cell death response to the R3bv2 core effectors RipH2, RipV1, and RipBI may be very promising resistance sources as well.

Interestingly, the two effectors that triggered strong cell death responses in the largest number of plant accessions are the non-core effectors RipAT and RipAV. One possible explanation is that these effectors are in fact missing from some R3bv2 isolates because ancestors of these isolates were exposed to plants carrying cognate resistance genes and loss of these effectors thus increased fitness on such plants.

Finally, the putative effector Ucd197, translocation of which has not been determined, did trigger a strong cell death response in both eggplant accessions suggesting that it may in fact be an effector.

In summary, we have found several promising resistance sources for R3bv2, the resistance genes of which could be mapped and cloned. Importantly, Agrobacterium-mediated transient expression could be used for mapping the underlying resistance genes in segregating progenies. This would be very useful since employing actual R3bv2 isolates for this purpose would be very difficult in the USA due to the regulatory restrictions and the risk of accidental spread of R3bv2.

In conclusion, we have shown here how genome sequencing of multiple isolates of a select agent and transient expression assays of its effector genes can give new insight into pathogen evolution and dissemination and can constitute a tool for identifying new promising sources of disease resistance. Importantly, this research
was done without having to actually do any research with the select agent R3bv2 itself, avoiding regulatory hurdles and minimizing the risk of accidental pathogen spread.
### Table 1: R3bv2 strains used in this study with references to papers in which strains were described previously.

<table>
<thead>
<tr>
<th>Strain Name</th>
<th>sequevar</th>
<th>Year of isolation</th>
<th>Geographic location of isolation</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPO1609&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>1995</td>
<td>Netherlands</td>
<td>(27)</td>
</tr>
<tr>
<td>UW551&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>2003</td>
<td>Kenya/Wisconsin</td>
<td>(21)</td>
</tr>
<tr>
<td>NCPPB909&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>1961</td>
<td>Egypt</td>
<td>(54)</td>
</tr>
<tr>
<td>23-10BR&lt;sup&gt;b&lt;/sup&gt;</td>
<td>27</td>
<td>1981</td>
<td>Brazil</td>
<td>this work</td>
</tr>
<tr>
<td>NCPPB282&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2</td>
<td>1950</td>
<td>Colombia</td>
<td>(9)</td>
</tr>
<tr>
<td>POPS2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1</td>
<td>1980</td>
<td>China</td>
<td>this work</td>
</tr>
<tr>
<td>Rs2</td>
<td>1</td>
<td>2008</td>
<td>Bolivia</td>
<td>this work</td>
</tr>
<tr>
<td>CFBP4808</td>
<td>1</td>
<td>1954</td>
<td>Israel</td>
<td>(9)</td>
</tr>
<tr>
<td>CFBP4578</td>
<td>1</td>
<td>1966</td>
<td>Egypt</td>
<td>(9)</td>
</tr>
<tr>
<td>CFBP3927</td>
<td>1</td>
<td>1968</td>
<td>Greece</td>
<td>(9)</td>
</tr>
<tr>
<td>CFBP3884</td>
<td>1</td>
<td>1984</td>
<td>Sweden</td>
<td>(9)</td>
</tr>
<tr>
<td>CFBP3873</td>
<td>1</td>
<td>1992</td>
<td>Belgium</td>
<td>(9)</td>
</tr>
<tr>
<td>JT516</td>
<td>1</td>
<td>1993</td>
<td>Reunion</td>
<td>(9)</td>
</tr>
<tr>
<td>CFBP3858</td>
<td>25</td>
<td>1995</td>
<td>Netherlands</td>
<td>(9)</td>
</tr>
<tr>
<td>LNPV28.23</td>
<td>1</td>
<td>2004</td>
<td>Reunion</td>
<td>(9)</td>
</tr>
<tr>
<td>KZR-5&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1</td>
<td>2004</td>
<td>Netherlands</td>
<td>(58)</td>
</tr>
<tr>
<td>PA1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1</td>
<td>2004</td>
<td>Netherlands</td>
<td>(58)</td>
</tr>
</tbody>
</table>

<sup>a</sup> genome sequences were not obtained as part of this study but were publicly available

<sup>b</sup> DNA of isolate kindly provided by Dr. Allen (University of Wisconsin, USA)

<sup>c</sup> DNA kindly provided by Dr. Stevens (University of Groningen, The Netherlands)
<table>
<thead>
<tr>
<th>Strain</th>
<th>SRA accession numbers</th>
<th>Average Read length (bp)</th>
<th>Number of reads</th>
<th>Genome coverage depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>23-10BR&lt;sup&gt;d&lt;/sup&gt;</td>
<td>SRX702422</td>
<td>433.59&lt;sup&gt;a&lt;/sup&gt;</td>
<td>271,396</td>
<td>20 x</td>
</tr>
<tr>
<td>NCPPB282&lt;sup&gt;d&lt;/sup&gt;</td>
<td>SRX703654</td>
<td>431.73&lt;sup&gt;a&lt;/sup&gt;</td>
<td>290,941</td>
<td>22 x</td>
</tr>
<tr>
<td>POPS2&lt;sup&gt;d&lt;/sup&gt;</td>
<td>SRX703652</td>
<td>422.94&lt;sup&gt;a&lt;/sup&gt;</td>
<td>230,546</td>
<td>16 x</td>
</tr>
<tr>
<td>Rs2</td>
<td>SRX703653</td>
<td>427.76&lt;sup&gt;a&lt;/sup&gt;</td>
<td>277,564</td>
<td>23 x</td>
</tr>
<tr>
<td>CFBP4808</td>
<td>SRX701225</td>
<td>2 x 70&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5,065,895&lt;sup&gt;c&lt;/sup&gt;</td>
<td>102 x</td>
</tr>
<tr>
<td>CFBP4578</td>
<td>SRX701307</td>
<td>2 x 70&lt;sup&gt;b&lt;/sup&gt;</td>
<td>755,467&lt;sup&gt;c&lt;/sup&gt;</td>
<td>13 x</td>
</tr>
<tr>
<td>CFBP3927</td>
<td>SRX701324</td>
<td>2 x 70&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1,511,314&lt;sup&gt;c&lt;/sup&gt;</td>
<td>24 x</td>
</tr>
<tr>
<td>CFBP3884</td>
<td>SRX701786</td>
<td>2 x 70&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2,711,618&lt;sup&gt;c&lt;/sup&gt;</td>
<td>45 x</td>
</tr>
<tr>
<td>CFBP3873</td>
<td>SRX701787</td>
<td>2 x 70&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4,035,125&lt;sup&gt;c&lt;/sup&gt;</td>
<td>84 x</td>
</tr>
<tr>
<td>JT516</td>
<td>SRX701788</td>
<td>2 x 70&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3,214,532&lt;sup&gt;c&lt;/sup&gt;</td>
<td>54 x</td>
</tr>
<tr>
<td>CFBP3858</td>
<td>SRX701790</td>
<td>2 x 70&lt;sup&gt;b&lt;/sup&gt;</td>
<td>9,365,297&lt;sup&gt;c&lt;/sup&gt;</td>
<td>140 x</td>
</tr>
<tr>
<td>LNPV28.23</td>
<td>SRX701791</td>
<td>2 x 70&lt;sup&gt;b&lt;/sup&gt;</td>
<td>6,619,254&lt;sup&gt;c&lt;/sup&gt;</td>
<td>108 x</td>
</tr>
<tr>
<td>KZR-5</td>
<td>SRX701792</td>
<td>2 x 70&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4,111,225&lt;sup&gt;c&lt;/sup&gt;</td>
<td>81 x</td>
</tr>
<tr>
<td>PA1</td>
<td>SRX701793</td>
<td>2 x 70&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4,145,998&lt;sup&gt;c&lt;/sup&gt;</td>
<td>85 x</td>
</tr>
</tbody>
</table>

<sup>a</sup> Genomes of strains 23-10BR, NCPPB282, POPS2 and Rs2 were sequenced by 454.

<sup>b</sup> Genomes of these strains were sequenced by Illumina HiSeq, generating paired reads.

<sup>c</sup> These numbers refer to the numbers of read-pairs (rather than individual reads).

<sup>d</sup> Genome assemblies have been deposited in GenBank under accession numbers JQOI00000000.1, JQSH00000000.1 and JQSI00000000.1.
Table 3: Summary of R3bv2 core effector analysis

<table>
<thead>
<tr>
<th>R3bv2 Core effectors</th>
<th>R3bv2 non-core effectors</th>
</tr>
</thead>
</table>

* Core effectors of *Rs* overall are underlined.
Table 4: Elicitation of cell death by transient expression of *Rs* effectors on eggplant (MM738 and MM853), *Nicotiana benthamiana*, tomato (20 genotypes), pepper (5 genotypes), and lettuce (22 genotypes).

<table>
<thead>
<tr>
<th>Effector – Ral&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Effector – Rip&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Lettuce&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Nicotiana&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Pepper&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Tomato&lt;sup&gt;c&lt;/sup&gt;</th>
<th>MM853</th>
<th>MM738</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ral009t</td>
<td>not a Rip</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral010t</td>
<td>RipD</td>
<td>Inter.</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral012t</td>
<td>RipAA</td>
<td>Inter.</td>
<td>Strong</td>
<td>Strong</td>
<td>Inter.</td>
<td>Inter.</td>
<td>Inter.</td>
</tr>
<tr>
<td>Ral013t</td>
<td>RipW</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral014t</td>
<td>RipV2</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Inter.</td>
<td>Inter.</td>
<td>None</td>
</tr>
<tr>
<td>Ral015t</td>
<td>RipAY</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral016t</td>
<td>RipO1</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral017t</td>
<td>RipA4</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral018t</td>
<td>RipG5</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral019t</td>
<td>RipAB</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral021t</td>
<td>RipG7</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral022t</td>
<td>RipH2</td>
<td>Strong</td>
<td>None</td>
<td>None</td>
<td>Inter.</td>
<td>Strong</td>
<td>None</td>
</tr>
<tr>
<td>Ral023t</td>
<td>RipAJ</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Inter.</td>
</tr>
<tr>
<td>Ral024t</td>
<td>RipG4</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral025t</td>
<td>RipE2</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral027t</td>
<td>RipF2</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral028t</td>
<td>RipAT</td>
<td>Strong</td>
<td>None</td>
<td>Strong</td>
<td>Strong</td>
<td>Strong</td>
<td>Strong</td>
</tr>
<tr>
<td>Ral030t</td>
<td>RipG6</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Inter.</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral031t</td>
<td>RipAP</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral033t</td>
<td>RipAV</td>
<td>Strong</td>
<td>None</td>
<td>Strong</td>
<td>Strong</td>
<td>Strong</td>
<td>None</td>
</tr>
<tr>
<td>Ral034t</td>
<td>RipV1</td>
<td>Strong</td>
<td>None</td>
<td>Inter.</td>
<td>None</td>
<td>Inter.</td>
<td>None</td>
</tr>
<tr>
<td>Ral036t</td>
<td>not a RIP</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral037t</td>
<td>not a Rip</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral038t</td>
<td>RipAO</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral040t</td>
<td>RipB1</td>
<td>Inter.</td>
<td>None</td>
<td>Inter.</td>
<td>Strong</td>
<td>Inter.</td>
<td>Strong</td>
</tr>
<tr>
<td>Ral041t</td>
<td>RipE1</td>
<td>None</td>
<td>None</td>
<td>Inter.</td>
<td>Inter.</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral042t</td>
<td>RipAS</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral043t</td>
<td>not a Rip</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Strong</td>
<td>Strong</td>
</tr>
<tr>
<td>Ral044t</td>
<td>RipG3</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral045t</td>
<td>RipN</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Ral048t</td>
<td>RipX</td>
<td>Inter.</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Inter.</td>
<td>None</td>
</tr>
<tr>
<td>Ral049t</td>
<td>not a Rip</td>
<td>None</td>
<td>Inter.</td>
<td>None</td>
<td>Inter.</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

<sup>a</sup> Ral designations are from UC Davis Charge database (http://charge.ucdavis.edu/charge_db/chargedb_index.php).

<sup>b</sup> Bolded effectors are part of the R3bv2 core effectorome.

<sup>c</sup> Intermediate and strong results for tomato, pepper, and lettuce indicate that cell death was observed in at least one of the tested genotypes.
Supplementary Table Legends

Table S1. Number of SNPs distinguishing R3bv2 genomes from each other and from reference strain 23-10BR over 2,259,494bp.

Table S2. Coverage of 58 UW551 effectors in the 17 other analyzed genomes.

Table S3. Detailed results of cell death resulting from Agrobacterium transient expression of 33 cloned effector genes in all screened plant genotypes.
Figure Legends

Figure 1: A. Maximum likelihood tree of all newly sequenced and a selection of published *R. solanacearum* genomes. The tree is based on SNPs identified by aligning reads against the genome of *R. solanacearum* Po82. The tree is mid-point rooted. Bootstrap values based on 1200 non-parametric replicates are indicated at nodes. Branch lengths are proportional. The clade corresponding to phylotype IIB is highlighted using a higher weight. B. Maximum likelihood tree of R3bv2 isolates based on SNPs when aligning reads against the genome of *R. solanacearum* Po82. Bootstrap values based on 1200 non-parametric replicates are indicated at nodes. The tree is rooted on the genome of Po82. Because branch lengths are very different from each other, a cladogram is shown and the number of SNPs is indicated underneath each branch. The clade containing most European/African/Asian isolates is highlighted using a higher weight.

Figure 2: Inferred routes of dissemination of R3bv2 strains. Strain details are listed in Table 1. European isolates are not placed exactly over their country of isolation because of SPACE limitations. Map from Luke Roberts ([http://lukeroberts.deviantart.com/](http://lukeroberts.deviantart.com/)) released under a Creative Commons Attribution-Noncommercial 3.0 license.

Figure 3: Examples of the broad categories listed in Table 3 that were used for classification of cell death elicited by agrobacterium-mediated transient expression of effectors in example plant genotypes tested.
Figure S1: Maximum Likelihood tree of R3bv2 isolates based on SNPs identified in the main chromosome only (A) or in the mega plasmid only (B) using the genome of Po82 as reference for SNP calling and as root for tree construction.
References


8. CDC, and APHIS. National Select Agent Registry.


56. Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A.,
    Dagdigian, C., Fuellen, G., Gilbert, J. G. R., Korf, I., Lapp, H., Lehväslaiho,
    H., Matsalla, C., Mungall, C. J., Osborne, B. I., Pocock, M. R., Schattner, P.,
    The Bioperl Toolkit: Perl Modules for the Life Sciences. Genome Research

57. Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and
    post-analysis of large phylogenies. Bioinformatics 30(9):1312-1313

58. Stevens, P., and van Elsas, J. 2010. Genetic and phenotypic diversity of
    Ralstonia solanacearum biovar 2 strains obtained from Dutch waterways.


    Pathogens in Agro-Ecosystems. Annual Review of Phytopathology 46(1):75-100

61. Tai, T. H., Dahlbeck, D., Clark, E. T., Gajiwala, P., Pasion, R., Whalen, M. C.,
    confers resistance to bacterial spot disease in tomato. Proceedings of the
    National Academy of Sciences 96(24):14153-14158

    characterization and 16S-23S probe development for differentiating
geographically dispersed isolates of the phytopathogen Ralstonia solanacearum. Environmental Microbiology 3(12):785-797


