



Whole genome characterization of strains belonging to the *Ralstonia solanacearum* species complex and *in silico* analysis of TaqMan assays for detection in this heterogenous species complex

Viola Kurm · Ilse Houwers · Claudia E. Coipan · Peter Bonants · Cees Waalwijk · Theo van der Lee · Balázs Brankovics · Jan van der Wolf

Accepted: 17 December 2020
© The Author(s) 2021

Abstract Identification and classification of members of the *Ralstonia solanacearum* species complex (RSSC) is challenging due to the heterogeneity of this complex. Whole genome sequence data of 225 strains were used to classify strains based on average nucleotide identity (ANI) and multilocus sequence analysis (MLSA). Based on the ANI score (>95%), 191 out of 192(99.5%) RSSC strains could be grouped into the three species *R. solanacearum*, *R. pseudosolanacearum*, and *R. syzygii*, and into the four phylotypes within the RSSC (I,II, III, and IV). *R. solanacearum* phylotype II could be split in two groups (IIA and IIB), from which IIB clustered in three subgroups (IIBa, IIBb and IIBc). This division by ANI was in accordance with MLSA. The IIB subgroups found by ANI and MLSA also differed in the number of SNPs in the primer and probe sites of various assays. An *in-silico* analysis of eight TaqMan and 11 conventional PCR assays was performed using the whole genome sequences. Based on this analysis several cases of potential false positives or false negatives can be expected upon the use of these assays for their intended target organisms. Two TaqMan assays and two PCR assays targeting the 16S rDNA sequence should be able to detect all phylotypes of the RSSC. We conclude

that the increasing availability of whole genome sequences is not only useful for classification of strains, but also shows potential for selection and evaluation of clade specific nucleic acid-based amplification methods within the RSSC.

Keywords MLSA · ANI · *in-silico* analysis · *Ralstonia solanacearum* species complex · Phylogenetic classification

Introduction

Bacteria belonging to the *Ralstonia solanacearum* species complex (RSSC) are the causative agents of diseases in plants of many different botanical families. It is estimated that the cumulative host range of the pathogen exceeds 250 plant species (Genin and Denny 2012). These include economically important crops such as banana, tomato, tobacco, strawberry, peanut and potato (Denny 2006). Its relatively long survival both in water and soil contributes to the spread of the pathogen, for example through irrigation or planting in infested soils (Schönfeld et al. 2003; van Elsas et al. 2000). The RSSC ranks second in the list of the economically most important bacterial plant pathogens (Mansfield et al. 2012) and detection of the pathogen in plant propagation material, soil and irrigation water is important to manage the disease. Recently, the RSSC underwent several instances of phylogenetic reclassification based on results using advanced molecular techniques. These reclassifications call for reinvestigation of species identity of bacterial strains deposited in collections.

V. Kurm (✉) · I. Houwers · P. Bonants · C. Waalwijk · T. van der Lee · B. Brankovics · J. van der Wolf
Wageningen University and Research, P.O. Box 16, 6700 AA Wageningen, the Netherlands
e-mail: viola.kurm@wur.nl

C. E. Coipan · C. Waalwijk
National Institute for Public Health and the Environment, Antonie van Leeuwenhoeklaan 9, 3721 MABilthoven, the Netherlands

Previous studies demonstrated a high phylogenetic complexity within the RSSC (Prior et al. 2016). For the phylogenetic classification of strains within the RSSC different principles have been used. Traditionally, the classification was based on differences in host range, resulting in a subdivision in five races (Buddenhagen 1962) and on biochemical properties resulting in six biovars or biotypes (Hayward 1964). More recently, molecular phylogenies based on ITS sequence, the *egl* and the *hrpB* genes differentiated four phylotypes (I, II, III and IV) and phylotype II was further differentiated into IIA and IIB. Each phylotype harbours predominantly strains from a specific geographic origin: phylotype I from Asia, phylotype III from Africa, and phylotype IV from Japan, the Philippines and Australia (Ailloud et al. 2015; Safni et al. 2014). In contrast, strains of phylotype II have a more global distribution and were retrieved from Europe, Africa, Asia, and South America. Grouping within the phylotypes was done on the basis of the endoglucanase gene (*egl*) sequences resulting in 53 sequevars (Albuquerque et al. 2014). Recently this overlapping classification has been resolved by Safni et al. (2014) who used a polyphasic approach to revise the species complex into three species: *R. solanacearum* (phylotypes IIA and IIB), *R. pseudosolanacearum* (phylotypes I and III) and *R. syzygii* (phylotype IV) with three subspecies, *R. syzygii* subsp. *syzygii*, *R. syzygii* subsp. *celebesensis* and *R. syzygii* subsp. *indonesiensis*. The robustness of this classification system was confirmed by whole genome sequencing (Ailloud et al. 2015). Furthermore sequence variation identified in single-copy orthologous genes ($n = 686$) among the genomes of 39 strains suggest that phylotypes IIA and IIB may represent different subspecies of *R. solanacearum* while phylotype I and III are suggested to represent subspecies of *R. pseudosolanacearum* (Zhang and Qiu 2016).

Due to recent developments in sequencing techniques as well as in bioinformatics, annotated whole genome sequences of a high number of RSSC strains have become available; 192 at the time of this study. These sequences include the chromosome as well as the megaplasmid that most strains of the RSSC possess (Salanoubat et al. 2002). These genomic sequences can be used to assess phylogenetic relationships, for instance by the calculation of average nucleotide identity (ANI), which measures the sequence distance between strains using all homologous genomic regions (Arahal et al. 2014). ANI analysis was previously successfully applied to confirm the phylogenetic identity of a small number of well-characterized *Ralstonia* strains

(Prior et al. 2016; Remenant et al. 2010). Classifying strains into species based on ANI analysis scores is the new gold standard for prokaryotes (Richter and Rosselló-Móra 2009). In addition, multilocus sequence analysis (MLSA) has been successfully exploited for the classification of RSSC strains. This method relies on the simultaneous analysis of not one but several loci and allows for a higher resolution than phylogeny deduced from 16S rRNA genes (Glaeser and Kämpfer 2015). Wicker et al. (2011) used seven housekeeping and two virulence genes to separate *Ralstonia* strains into the four phylotypes and identified eight clades within these phylotypes. An approach using a large number of complete genome sequences for phylogenetic analysis can be a powerful tool to re-assess the phylogeny of the RSSC and to classify novel strains.

Detection and classification of strains belonging to the RSSC is currently largely performed by PCR based assays. A large number of published PCR assays were designed to detect all or specific groups of the RSSC. One of the first PCR assays to be described is based on 16S rDNA sequences and enables the detection of all 50 pathogenic *Ralstonia* strains that were analysed and which belonged to the four different phylotypes (Seal et al. 1993). For detection purposes TaqMan assays were developed, such as the one targeting the endoglucanase gene (*egl*) of ten race 1 strains from China (Huang et al. 2009). Many assays were developed for specific phylotypes, such as phylotype IIB, the causative agent of potato brown rot in Europe (Guidot et al. 2009). Similarly, primers were developed by Pastrik et al. (2002) and Fegan and Prior (2005), that are able to distinguish subgroups of the RSSC based on the ITS sequences. Moreover, a TaqMan assay in a triplex format was designed that enabled detection of the regulated potato pathogens *R. solanacearum* and *R. pseudosolanacearum* as well as *Clavibacter sepedonicus*, and also included a potato amplification control (Vreeburg et al. 2018). Although the developed primers and probes reacted specifically with the panel of target strains involved in the studies, the number of strains tested was often limited and they were in many cases not classified according to the presently used phylotype system. Therefore, in this study, we test the primer and probe fit for several PCR and TaqMan assays on a large number of strains classified according to the newest system, in an *in-silico* approach.

The high number of available whole genome sequences creates new opportunities to review the phylogeny of the RSSC complex and the specificity of nucleic acid-based assays. In this study, we used 225 whole

genome sequences, 33 belonging to the species *R. insidiosa*, *R. mannitolilytica* and *R. picketti* and 192 belonging to the RSSC, with an emphasis on *R. solanacearum* and *R. pseudosolanacearum* strains, and characterized them by ANI and MLSA analysis. We compared the phylogenetic information generated by both methods. In addition, we used whole genome sequences and an *in-silico* analysis to support selection of PCR assays based on the predicted specificity of the 19 commonly used PCR based methods for detection and diagnostics.

Materials and methods

In total, 225 *Ralstonia* strains were analysed, 192 of which from the RSSC complex. Of these strains, 20 were kindly provided by the Dutch General Inspection Service (NAK, Emmeloord, The Netherlands). Whole genome sequencing data, originating from different countries and isolated from various hosts, were generated by HiSeq (137) or PacBio (10) sequencing or extracted from the NCBI Genbank (79). One strain was sequenced with both PacBio and HiSeq. The metadata of all these isolates are listed in Table S1.

Reference strains

For each phylotype and species, a reference was chosen for identification of the strains and determination of the closest relative in the MLSA. If available, the type strains of the respective phylotype were selected as a reference strain. For all reference strains see Table 1.

DNA extraction

Strains were grown on yeast peptone glucose agar (YPGA) (per litre: yeast extract 5 g, peptone 5 g, glucose 10 g, agar 15 g) at 25 °C. DNA was extracted using the Wizard® Magnetic DNA Purification System for Food (Promega, Leiden, The Netherlands). DNA was extracted according to the manufacturer's protocol from approximately one full inoculation loop (10 µl) of bacterial slime collected from multiple colonies on plates. The elution of DNA was performed in two steps, in which subsequently 50 µl eluate and 75 µl elution buffer were used. For quantification, the DNA solutions were analysed with the Infinite® 200 PRO (Tecan, Männedorf, Switzerland) using the Quant-iT™

Table 1 Reference strains for the five phylotypes and the other *Ralstonia* spp. used in this study

Collection name	Species	Phylotype	Host	Origin	GenBank	Reference
GMI1000	<i>Ralstonia pseudosolanacearum</i>	I	<i>Lycopersicon esculentum</i>	French Guyana	GCA_000009125.1	Salanoubat et al. (2002)
NCPPB325 ^T	<i>R. solanacearum</i>	IIA	<i>Lycopersicon esculentum</i>	USA	De novo assembly	This study
PD2673	<i>R. solanacearum</i>	IIB	<i>Solanum tuberosum</i>	Netherlands	De novo assembly	This study
LMG9673 ^T	<i>R. pseudosolanacearum</i>	III	<i>Pelargonium capitatum</i>	Réunion	De novo assembly	This study
PSI07	<i>R. syzygii</i>	IV	<i>Lycopersicon esculentum</i>	Indonesia	GCA_000283475.1	Remenant et al. (2010)
AU12-08	<i>R. insidiosa</i>	NA	<i>Homo sapiens</i>	Australia	GCA_000442475.1	Zhang et al. (2014)
NCTC10894	<i>R. mannitolilytica</i>	NA	<i>Homo sapiens</i>	Japan	GCA_900455575.1	Suzuki et al. (2015)
ATCC 27511 ^T	<i>R. pickettii</i>	NA	<i>Homo sapiens</i>	USA	GCA_000743455.1	Daligault et al. (2014)

NA not applicable

^T - type strain of the given species

PicoGreen™ dsDNA Assay Kit (ThermoFisher, Waltham, USA) according to the manufacturer's protocol. DNA samples (5 µl) were run on 1% agarose gels to check for quality of high molecular weight DNA.

Next generation sequencing (NGS)

136 of the strains used in this study were sequenced using only Illumina HiSeq (Table S1). A random sheared shotgun library preparation was performed using the Truseq Nano DNA Library Prep kit (dual indexing) following the manufacturer's protocol. The samples were loaded on a paired end flowcell, using the Hiseq PE cluster kit V4. A cBot was used for cluster generation (Illumina). The procedure was continued with sequencing on a Hiseq 2500 (Illumina, San Diego, USA) on which 125 bp paired-end sequences were generated.

Ten of the strains were sequenced only with the PacBio Platform (Pacific BioSciences, Menlo Park, USA) (Table S1). One strain (PD2673) was sequenced with both Illumina HiSeq and PacBio sequencing. For the PacBio sequencing, DNA was sheared to 6 Kb and two SMRT bell libraries were prepared using the kit Barcoded Adapters for Multiplex SMRT sequencing in combination with the Sequel Binding Kit V2.0 and the Sequel Polymerase 2.0 Kit. Per library, a pool with sheared DNA of all strains was used as input. The library preparation was completely performed according to the manufacturer's protocol. Sequencing was done on a Sequel system. For coverage of the PacBio sequencing see Table S2.

Both sequencing procedures were operated at the services of Business Unit Bioscience, Wageningen Plant Research (Wageningen, The Netherlands).

Sequence data processing

Whole genome assembly was performed for all Illumina HiSeq sequenced strains using SPAdes 3.11.1 (Bankevich et al. 2012). The strain PD2763 was sequenced using both the PacBio and the Illumina HiSeq platforms. For the PacBio reads, de-multiplexing was performed by aligning the barcodes to the sub-reads with pyPaSWAS version 3.0 (<https://doi.org/10.1371/journal.pone.0190279>) and selecting the best matching barcodes for binning (<https://github.com/swarris/pyPaSWAS/wiki/Demultiplexing>). Canu version 1.6 (Koren et al. 2017) was used to assemble the PacBio reads, using a minimum read length of 1 kb and a minimum overlap of 500 bp. After this, ARCS

(Yeo et al. 2017) was used to create scaffolds of these assemblies using the 10x Illumina read sets. Genome sequences were deposited at NCBI (Table S1). In addition, publicly available genomes were downloaded from NCBI (Table S1).

All whole genome assemblies belonging to *Ralstonia* strains were downloaded from NCBI. All of the genomes were compared to the reference strain collection using FastANI (Jain et al. 2018) and genomes that did not have at least 95% identity to a reference strain were left out of the analyses, because they did not belong to *Ralstonia* spp.

A three-step quality check was performed for all the strains before including them in the analyses: presence of the 16S rRNA gene, the maximum sequence divergence between 16S rRNA copies, and BUSCO scores. All 16S rRNA sequences (both completed and partial) were extracted from each genome. All genomes that contained no 16S rRNA sequences were excluded from the study. Per genome, all 16S rRNA sequences were aligned using MUSCLE (Edgar 2004). Subsequently the divergence between each sequence entry within the alignment was calculated. Divergence was calculated by identifying whether there is an overlap between the two sequence entries in the same multiple sequence alignment (divergence was only calculated for overlapping sequences), then divergence was calculated by dividing the number of variable positions in the overlapping region by the length of the overlapping region. Genomes that had greater than 0.05 16S rRNA sequence divergence were excluded. The remaining genomes should either be single strain assemblies or contain closely related strains. To rule out contamination by closely related strains we calculated the BUSCO scores for each genome based on the prokka protein predictions. The protein sequences were analyzed using BUSCO v3.0.2 (Seppey et al. 2019) and betaproteobacteria_odb9 lineage dataset. Only the genomes that had at least 95% BUSCO score and less than 10% duplicated, fragmented or missing genes were kept.

ANI

Average nucleotide identity (ANI) analysis was performed using FastANI (Jain et al. 2018) with default settings comparing all strains to each other. The species demarcation was set at an ANI threshold value of 95% identity from the reference strains (the type strain when it was present) (Goris et al. 2007). A neighbour joining tree

was generated from a distance matrix derived from the pairwise sequence identity matrix using *BioPerl* module *Bio::Tree::DistanceFactory*. A heat map was constructed in *R* studio using the *heatmap* function from the package ‘Complex Heatmap’ (Schlesner et al. 2016).

MLSA

A multilocus sequence analysis (MLSA) was performed based on nine protein coding genes (*adk*, *egl*, *fliC*, *gdhA*, *gyrB*, *leuS*, *mutS*, *ppsA* and *rplB*). *Egl* and *fliC* are located on the megaplasmid, whereas the remaining genes are located on the chromosome. These have been used previously in the study of the RSSC phylogeny (Wicker et al. 2011). The sequences were aligned using the MUSCLE alignment program (Edgar 2004) and concatenated. For seven strains, the *fliC* gene could not be located and was replaced by a sequence of gap values (‘-’) in the concatenated sequence. A maximum likelihood phylogenetic tree was constructed using IQ-Tree 1.6.19 using edge-proportional partition model (–spp) and 1000 bootstrap. Eight strains belonging to the species *R. insidiosa* (Ri), 15 strains belonging to *R. pickettii* (Rp) and 10 strains of *R. mannitolilytica* (Rm) obtained from NCBI Genbank were used as outgroups (Table S1). Alignments and tree construction were performed as described above for each of the nine genes separately as well. Furthermore, for each strain, the nine genes were compared to the reference sequences representing the respective phylotypes or species (Table 1). For each gene, the closest reference strain sequence was identified and the SNP distance was determined between the two.

In silico analysis TaqMan assays

An *in-silico* analysis was performed in order to determine the presence of SNPs between the concatenated primers and probes of 11 different conventional PCR and 8 real-time TaqMan PCR assays and the target sequences present in the *Ralstonia* strains (Table 2). An in-house pipeline was used for the analysis. First, primers and probes were mapped to genomes using *numcr* with at least 6 bp long kmer exact match (–*maxmatch* –*nosimplify* –*l* 6 –*c* 1 –*b* 6) (Delcher et al. 2002). Then, hits were extended to completely include the entire primer/probe sequence. Hits with more than 5 mismatches in either primer or probe were removed. This was a conservative choice as it has been shown

previously than assays with ≥ 4 SNPs in one primer or 3 SNPs in one and 2 in the other primer is unlikely to result in amplification (Lefever et al. 2013). Amplicon regions were identified by either forward primer matched to positive strand and reverse primer matched to reverse strand or reverse primer matched to positive strand and forward primer matched to reverse strand. Amplicons longer than 2000 bp were removed. For each assay, the amplicon with the least mismatches were kept per assembly. Mismatches (and gaps) were summed for both primers and probe (when present).

To estimate the primer binding efficiency, and therefore the consequences of SNPs, the ΔG value was calculated for all individual primers and probes and the five reference strains using the Oligo Analyzer tool from IDT (Integrated DNA Technologies, Skokie, Illinois, U.S.A.).

All primer/probe sets were checked for a potential reaction with non-target organisms by blasting the primer and probe sequences at NCBI with *blastn*. Primers and probes were concatenated with 5–10 N as spacers in between, the nucleotide collection was selected as database and restricted to bacteria. The word size was set to 7, the expect threshold to 1000 and no low complexity filter was used. The *blastn* standards for Match/Mismatch (2,-3) and gap costs (Existence: 5, Extension: 2) were left at the default. The *blastn* search was conducted on 01.07.2020.

Data analysis

All data analysis was done in R Studio (RStudio 2016).

To assess if the number of SNPs differed between the primer/probe sets, negative binomial linear regressions were used for PCR assays and TaqMan assays separately with the total number of SNPs between the primer sequences and the target sequences as the dependent and the primer set as the explanatory variable. Here, only those strains were included in which the respective target sequence was present with ≤ 5 mismatches per primer/probe. The overall effect of primer/probe set was determined with the *Anova* function from the package ‘*car*’ (Fox and Weisberg 2011). Pairwise differences between the sets were calculated with the *emmeans* function from the package ‘*emmeans*’ (Lenth et al. 2018). The same procedure was used to assess if the number of SNPs differed between primer/probe sets and phylotypes with primer set and phylotype as the explanatory variables in the regression.

Table 2 PCR and TaqMan assays for the RSSC species complex used in this study

Name	Target region	Forward primer (5'-3')	Reverse primer (5'-3')	Probe (5'-3')	Reference
Chen	intergenic UDP-3-O-acyl-N-acetylglucosamine deacetylase & Peroxiredoxin	GTGCTGCCTCCAAAACGACT	GAGGCACCCCGCATCCCTC		Chen et al. (2010)
Guidot1	hypothetical protein with signalpeptide	CTGAATTCGAAAAAGGATAGA GCA	CTCGACAAAACCTTTGCCAACTGAC		Guidot et al. (2009)
Guido2	DUF3987 super family	CAATAGAAAATTGCCGAGGTG ATA	CCTTCATAAGGATCTTCAACGAC		Guidot et al. (2009)
Guido3	DUF1018 super family	GCATGAAAAATGCTACGTTC CTC	ATGGTGATAGTCCGGAAATGAC		Guidot et al. (2009)
Huang	Endoglucanase gene	CCGACACCACGACCCCTGAA	GCGGACGGATAGATGTAGTTGC	ACCTGGCATACCTT GGCGACACC	Huang et al. (2009)
Körner	16S	CATGCCCTACACATGCAAGTC	CACGTTCCGATGTATTACTCA	AGCTTGCTACCTGCCGG	Körner et al. (2017)
Massart	ITS	CGCGGAGCAITGTATGAGAT	TCGTAATACTGGTTGATACA ATCACAAAC	CTCGCAAAAACGC	Massart et al. (2014)
Nmult1	ITS	CGTTGATGAGGGCGGCAATT	TCGCTTGACCCTATAACGAGTA		(Fegan and Prior 2005)
Nmult2	ITS	AAGTTATGGACGGTGGAAATC	TCGCTTGACCCTATAACGAGTA		(Fegan and Prior 2005)
Nmult3	ITS	ATTACSAAGCAATCGAAAAG ATT	TCGCTTGACCCTATAACGAGTA		(Fegan and Prior 2005)
Nmult4	ITS	ATTGCCAAGACGAGAGAAGTA	TCGCTTGACCCTATAACGAGTA		(Fegan and Prior 2005)
Nytor	Endoglucanase gene	CGCGAAACGAGCTGTC	TCACGTTGCCGTAGTAG	P1: CGGGTTTCGTC AACGCCGTGA C P2: CGGGTTTGTCAAACG CCGTGAC	Vreesburg et al. (2018)
Opina	intergenic UDP-3-O-acyl-N-acetylglucosamine deacetylase & Peroxiredoxin	GTCCGCCGTCAACTCACTTCC	GTCCGCCGTCAGCAATGCCGGAATCG		Opina et al. (1997)
Ozakman	Phage_tail_S super family	TTCACCCGAAACAGCG	TAGCCCCAGCAGATG	TTCGCCGATGCTTCCCA	Ozakman and Schaad (2003)
Pastrik	16S-23S ITS to within 23S rRNA	ACTAACGAAGCAGAGATGCA TTA	CCCAGTCACGGCAGAGACT		Pastrik et al. (2002)
Seal	16S	GGGGTAGCTTGTACCTACCTGCC	CCCACCTGCTGCCCTCCCGTAGGAGT		Seal et al. (1993)
Weller	16S	GCATGCTTACACATGCAAG TC	GGCAGTTCGATGTATTACTCA	AGCTTGCTACCTGCCGGCGA GTG	Weller et al. (2000)
WellerII	Phage_tail_S super family	TGGCGCACTGCACCTCAAC	AATCACATGCAATTCGCCTACG	TTCAAAGCCGAACACCTGCTG CAAG	Weller et al. (2000)
Wu	16S	CAACGGGAAGAACCTTACC	ACGTATCCCCACCTTCC	ACGACAGCCATGCAGCACCT	(Wu et al. 2008)

Results

Average nucleotide identity

The ANI analysis clustered the RSSC strains into three main groups with a high (>98%) intragroup average

similarity, and a relatively low (<89%) cross-group average similarity (Fig. 1). These three groups correspond to the previously described three species *R. pseudosolanacearum* (Rps), *R. solanacearum* (Rs) and *R. syzygii* (Rsy). In addition, the strains can be clustered into five groups that represent the five different

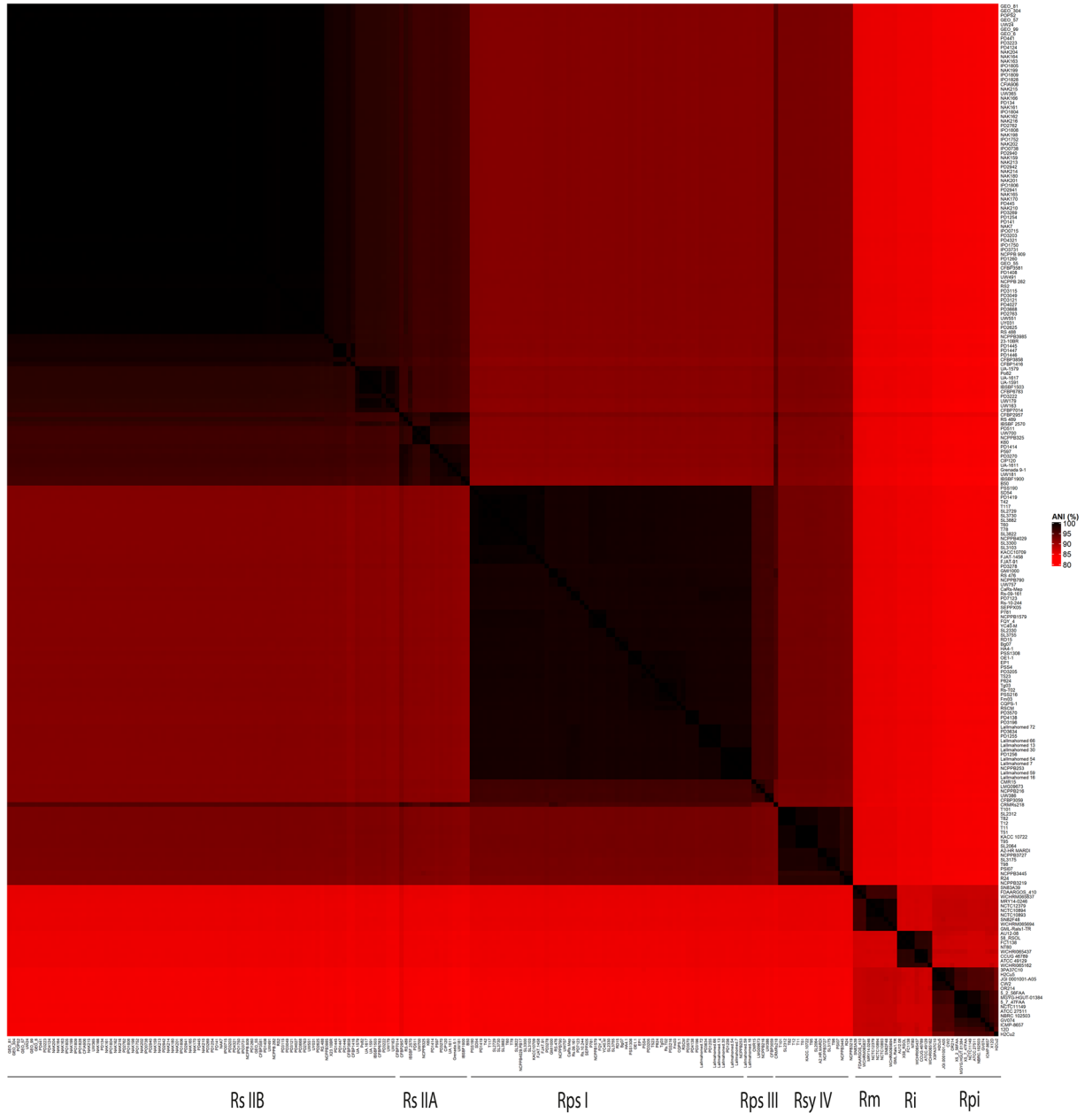


Fig. 1 Heatmap of ANI % identity values between all individual strains. A darker colouring represents a higher percentage of identity. Clusters with a high percentage of identity, representing

the three species and five phylotypes, are indicated on the y-axis and the x-axis respectively

Table 3 Mean Average nucleotide identity (%) between the phylotypes; min- and max-values in brackets

Species	Ri	Rm	Rp	Rps I	Rps III	Rs IIA	Rs IIB	Rsy IV
<i>Ralstonia insidiosa</i>	98.3 (96.8–100)							
<i>Ralstonia mannitolilytica</i>	86.8 (86.6–87.1)	98.0 (95.9–100)						
<i>Ralstonia pickettii</i>	87.2 (86.7–88.0)	87.8 (87.4–88.4)	96.7 (92.6–100)					
<i>Ralstonia pseudosolanacearum</i> (I)	85.2 (85.1–85.5)	85.6 (85.3–85.9)	84.8 (84.5–85.3)	98.9 (95.0–100)				
<i>Ralstonia pseudosolanacearum</i> (III)	85.2 (85.1–85.3)	85.6 (85.4–85.9)	84.8 (84.5–85.0)	95.9 (93.0–96.1)	97.6 (96.3–100)			
<i>Ralstonia solanacearum</i> (IIA)	85.1 (84.8–86.0)	85.4 (85.0–86.5)	84.6 (84.2–85.4)	91.4 (90.9–96.9)	91.4 (91.0–92.2)	98.1 (95.5–100)		
<i>Ralstonia solanacearum</i> (IIB)	85.2 (84.9–85.4)	85.5 (85.1–85.8)	84.7 (84.2–85.0)	91.6 (91.0–96.2)	91.6 (91.1–91.8)	96.1 (95.2–98.2)	99.2 (95.9–100)	
<i>Ralstonia syzygii</i> (IV)	85.2 (85.1–85.4)	85.6 (85.4–85.9)	84.7 (84.4–84.9)	92.7 (91.8–92.5)	92.3 (92.0–92.5)	91.9 (91.3–92.9)	92.3 (91.5–92.5)	98.5 (97.1–100)

phylotypes. Strains belonging to phylotype I also showed a higher ANI similarity to phylotype III than the commonly applied 95% for species demarcation. Strains belonging to phylotype IIA showed an ANI similarity above 95% to strains belonging to phylotype IIB, while strains belonging to phylotype IV only showed similarity above 95% to strains belonging to the same phylotype (Table 3). The tree created from a pairwise distance matrix for all 192 strains and outgroups also showed a separation of the five phylotypes. In addition, it demonstrates that for most phylotypes, a high genetic variation was found (Fig. 2). Phylotype IIB contained three groups, named group a, b and c. In addition, by comparison with the reference strains, it was shown that three strains deviated from the phylotype they have been assigned to. The strains RS 489 and IBSBF 2570, identified as phylotype IIB, belong to phylotype IIA according to the ANI. They grouped with phylotype IIA with an average similarity of 97.9% and 97.4% respectively compared to a similarity of 97.1% and 96.8% with IIB. Strain CRMRS218 was identified as phylotype I, but had an average similarity of 95.9% with phylotype IIA compared to 95.4% with phylotype I. Strain CRMRS218 matched the identification criteria for both *R. pseudosolanacearum* and *R. solanacearum*, and could therefore not be assigned to one of the species described.

Multilocus sequence analysis (MLSA)

The phylogenetic tree based on the concatenated loci split the RSSC strains into four clusters supported by bootstrap values that corresponded to phylotype I, II, III and IV and within phylotype II the two groups IIA and IIB were clustered (Fig. 3). Also, in the MLSA tree three groups of phylotype IIB could be detected. Two groups comprised mainly strains isolated from plants of the order Zingiberales (designated group b and c), including strains from the genera *Musa* and *Heliconia* all originated from Latin- and Central America. The other group (designated group a) included strains isolated from different plant species or from various other substrates and various geographic origins including Europe, Asia, Africa, and Latin-America.

Several strains that were grouped into one phylotype according to the concatenated sequences of all used genes, contained one or two genes that were assigned to another phylotype. The 89 phylotype IIB strains contained nine strains with genes resembling another

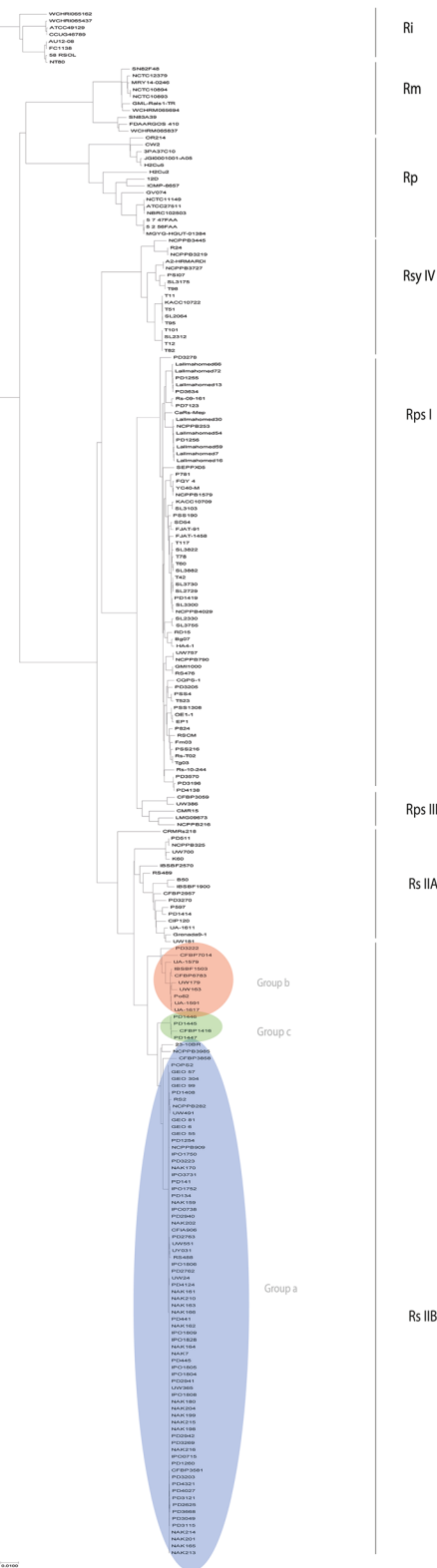


Fig. 2 Neighbour joining tree of ANI pairwise distances

phylotype. Eight strains of group b from phylotype IIB contained *fliC* genes that were grouped with phylotype IIA and two strains from group b contained *gyrB* genes grouped with IIA. Majority (13 out of 16) of strains grouped into phylotype IIA contained *fliC*, *gdhA* and *gyrB* genes that grouped to phylotype IIB. One strain, CMRMs218 (the unclassified strain) that was first identified as phylotype I, contains two genes (*egl* and *mutS*) that grouped to phylotype I and two genes that grouped to phylotype IIB (*fliC* and *gdhA*). The remaining genes were most similar to phylotype IIA and consequently this strain clusters with IIA in the tree. For the number of SNPs with the most closely related reference strain for each gene, see Table S3. Out of the five strains of phylotype III, three strains contained *mutS* and *ppsA* genes more related to phylotype I, while the 64 phylotype I strains contained in 30 cases *fliC* genes and 20 *gyrB* genes grouping to phylotype III. In one instance, a *gyrB* gene grouping to phylotype IV was found. In seven cases, no *fliC* gene was identified in a strain based on annotation that was validated by homology search by exonerate.

Trees generated based on the individual loci *adk*, *egl*, *fliC*, *gdhA*, *gyrB*, *leuS*, *mutS*, *ppsA* and *rplB* largely supported the phylogeny predicted by the concatenated loci, but showed some differences as well (Fig. S1). In the tree based on the *egl* gene, three strains from phylotype IIA were split off from the main cluster (Fig. S1 b). In the trees based on *fliC* and *gyrB*, phylotype IIA was grouped within phylotype IIB (Fig. S1 c,e). Phylotype I and III were virtually undistinguishable with respect to the *fliC*-based tree. In the *gyrB* based tree, phylotype I was split in several groups. In addition, several members of phylotype IIA are grouped close to phylotype IIB. In the *ppsA* tree, group IIA was again split (Fig. S1 h).

While the individual trees might show some differences from each other, the classification into species and phylotypes based on the concatenated sequences is similar to the ANI.

Evaluation of PCR assay

In total, seven regions were targeted by the PCR and TaqMan assays analysed in this study (Table 4). The intergenic region between UDP-3-O-acyl-N-acetylglucosamine deacetylase & Peroxiredoxin, the *egl* gene and the rRNA genes (including the ITS region) were present in all strains of the RSSC (Table 5). In contrast, the gene belonging to the Phage-tail-S super

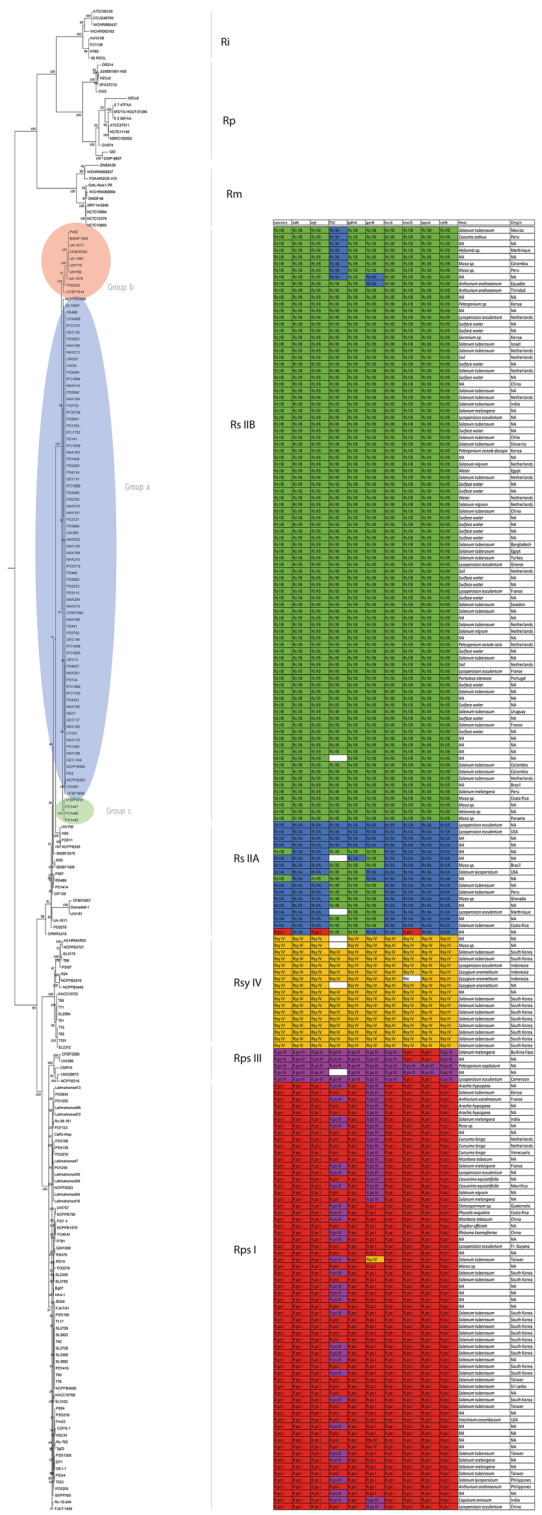


Fig. 3 Maximum likelihood tree generated from the concatenated sequences used in the MLSA. The table on the right indicates plants host, country of origin and which phylotype single genes

of each strain resemble most closely; empty fields indicate that the gene was not detected

Table 4 List of assays and corresponding target regions, numbered 1–7

Assay	Target	Region
Guidot1	hypothetical protein with signalpeptide	1
Guidot2	DUF3987 super family	2
Guidot3	DUF1018 super family	3
WellerII	Phage_tail_S super family	4
Ozakman	Phage_tail_S super family	4
Opina	intergenic	5
	UDP-3-O-acyl-N-acetylglucosamine deacetylase & Peroxiredoxin	
Chen	intergenic	5
	UDP-3-O-acyl-N-acetylglucosamine deacetylase & Peroxiredoxin	
Nytor	egl	6
Huang	egl	6
Nmult	rRNA	7
Massart	rRNA	7
Pastrik	rRNA	7
Wu	rRNA	7
Seal	rRNA	7
Weller	rRNA	7
Körner	rRNA	7

family, targeted by the Ozakman and Weller II TaqMan assay were present in a large part of phylotype IIB strains, but only in a small percentage of strains belonging to the other phylotypes. Also, the coding region of hypothetical proteins, that are targeted by the Guidot assays are present in the majority of phylotype IIB strains (78.02%, 83.52% and 79.12% for Guidot1, Guidot2 and Guidot3, respectively), but only sporadically in other phylotypes.

Assay sensitivity was determined by identifying assay/target combinations with more than five mismatches in one or more primer and probe (Table S3). These combinations were removed from the SNP analysis.

For both the TaqMan and the conventional PCR assays there were significant differences between the assays in the average number of SNPs with the 192 RSSC sequences (TaqMan assays: $\chi^2 = 3429.1$, $p < 0.01$; PCR assays: $\chi^2 = 1945.6$, $p < 0.01$).

The highest average number of SNPs for the TaqMan assays was found for the primer probe set Ozakman (6.9 ± 4.1) and the lowest for Körner (0.01 ± 0.07) and Weller (0.01 ± 0.07) (Fig. 4a). For the PCR assay, the highest average number of SNPs was found for the primer set Chen (3.3 ± 2.1), and the lowest for the set Seal (0.01 ± 0.2), Guidot 1, Guidot2 and Guidot 3 (0.0 ± 0.0) and Nmult 4 (0.0 ± 0.0) (Fig. 4b). In addition, SNPs for the different primer sets differed between the different phylotypes (TaqMan: $\chi^2 = 1162.3$, $p < 0.01$; PCR: $\chi^2 = 1217.3$, $p < 0.01$). For both assays the fit of most primer sets, except for those targeting the 16S gene, showed a high variability between the phylotypes (Fig. 5). Interestingly, for some assays the number SNPs also differed within phylotype IIB. For the primer set Chen strains belonging to the groups b and c and for the primer set Huang, the strains belonging to group c had on average less SNPs than group a (Fig. S2). For the sets Opina and Ozakman, most of the strains in group b and c had more SNPs in comparison to the main group a. In an overview, for each assay the number of SNPs with each phylotype and the percentage of strains with >5 mismatches is shown (Table 6).

In order to estimate the consequences of SNPs for primer-target binding, ΔG values were calculated for all

Table 5 Percentage of strains within a phylotype that contained the respective target region

Species	Phylotype	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	Region 7
<i>R. pseudosolanacearum</i>	I	1.54	9.23	3.08	3.08	100	100	100
<i>R. pseudosolanacearum</i>	III	0	0	20	20	100	100	100
<i>R. solanacearum</i>	IIA	0	57.14	0	7.14	100	100	100
<i>R. solanacearum</i>	IIB	78.02	83.52	79.12	81.32	100	100	100
<i>R. syzygii</i>	IV	0	0	11.76	17.65	100	100	100
<i>R. insidiosa</i>		0	0	0	0	0	0	100
<i>R. mammitolytica</i>		0	0	0	0	0	0	100
<i>R. pickettii</i>		0	0	0	0	0	0	100

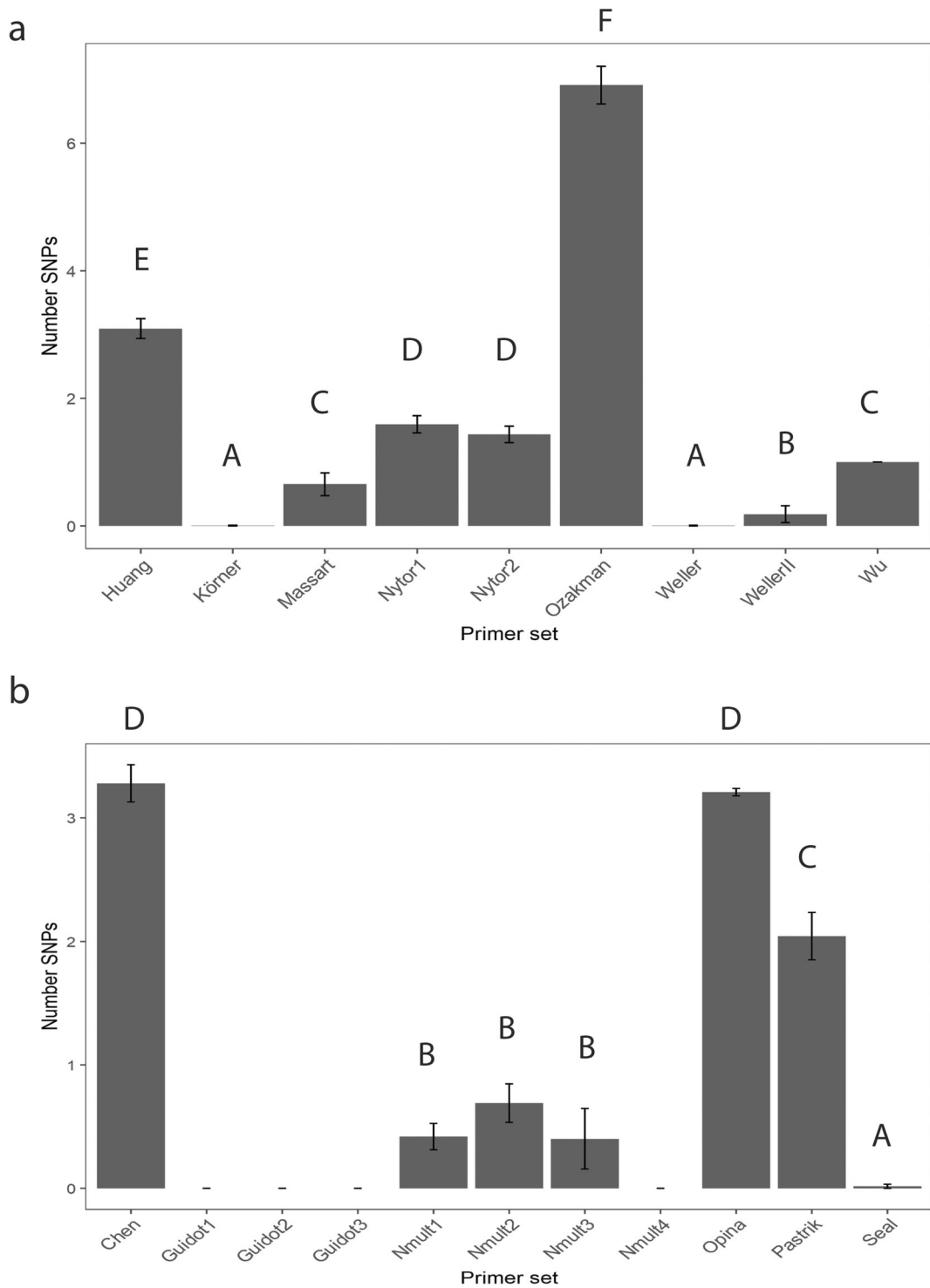


Fig. 4 Average number of SNPs between the concatenated primer and probes of the different sets and the target regions; **a** TaqMan assays, **b** conventional PCR assays; error bars indicate the standard error

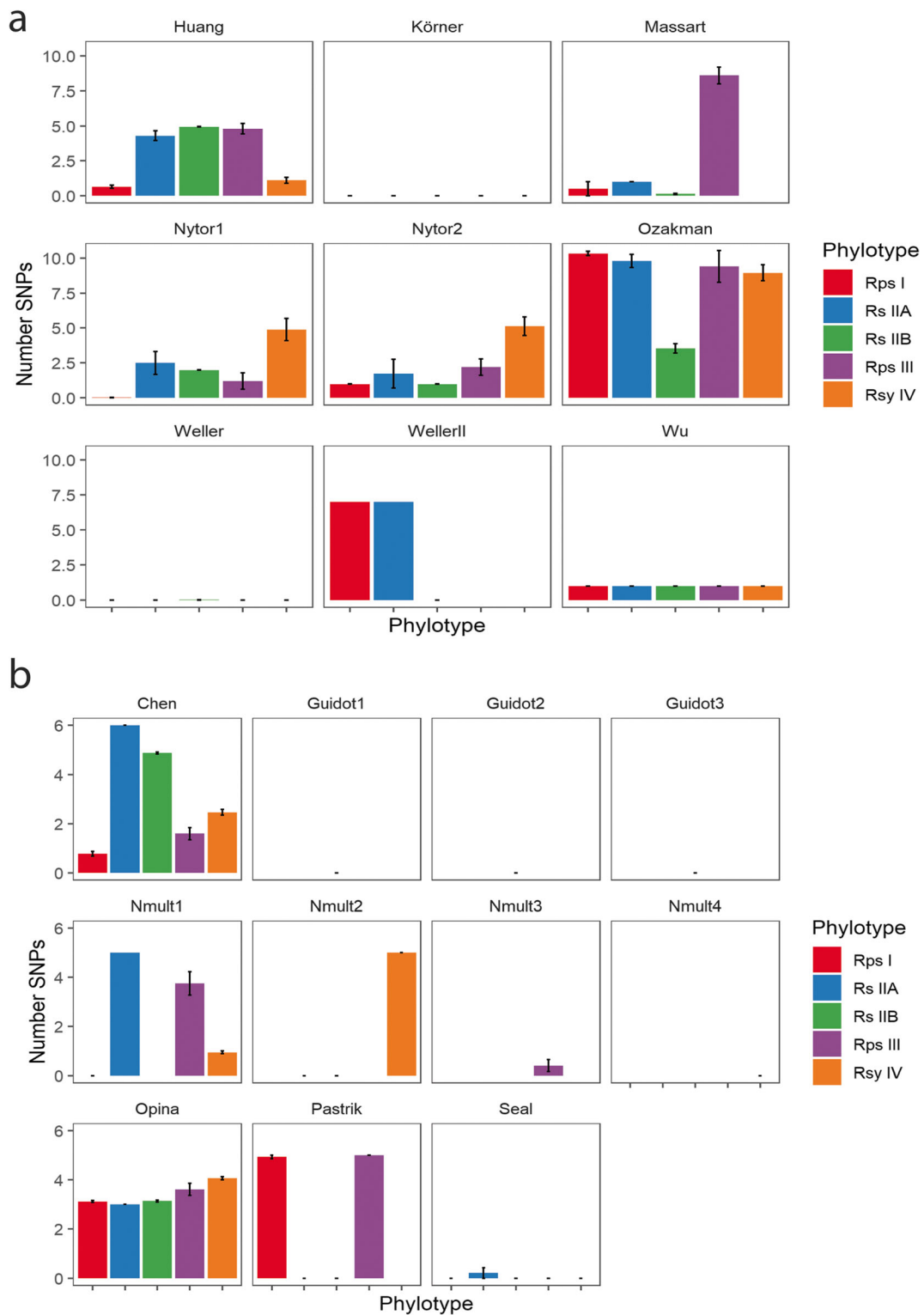


Fig. 5 Average number of SNPs between the concatenated primer and probes of the different sets and the target regions of the 5 different phylotypes; **a** TaqMan assays, **b** conventional PCR

assays. Red = phylotype I, blue = phylotype IIB, green = phylotype IIB, purple = phylotype III, orange = phylotype IV; error bars indicate the standard error

primers and probes separately for the five reference strains and compared to the number of SNPs with the target region (Fig. S3).

The maximal ΔG of each set, which should determine the reaction efficiency, shows qualitative differences between the reference strains (Fig. S4).

With a BLAST search, matches with non-target strains were found for several sets (Table S4). Some showed no or only one SNP between primer and non-target organism. The Wu assay matched with a high number of species of which only few are listed as it was designed for Gram-negative bacteria. Most frequently found non-target matches were found for the taxa *Ralstonia pickettii*, *Ralstonia mannitolilytica* and *Ralstonia insidiosa*. Therefore, these species were more closely investigated. In total, nine assays gave positive results for *R. insidiosa* and *R. pickettii* (i.e. less than 15 mismatches) and 13 assays gave positive results for *R. mannitolilytica* (Table S3). For some assays, there were only 0–3 SNPs between the assay and the target gene present in the non-target strain (Fig. S5). For example, there was only one mismatch for the Wu assay and the strains used here of either of the three species and assays targeting the 16S rRNA gene, such as Weller, Körner and Seal had only few mismatches with *R. mannitolilytica*.

Table 6: For each assay showing the percentage of strains detected <5 SNPs per primer and the average number of SNPs per phylotype (%detected, nr. SNPs); the first row indicates for which phylotype the assay was designed; values that are in agreement with the designed specificity are marked in green; a) PCR assays, b) TaqMan assays.

Discussion

In this study, we demonstrated the application of whole genome sequences for classification of the *Ralstonia solanacearum* species complex. By generating and collecting sequences of 192 strains of the RSSC, we were able to confirm current species classification using an ANI and MLSA approach. In addition, we showed that both average nucleotide identity (ANI) and multi-locus sequencing analysis (MLSA) can reveal grouping of strains into species and phylotypes whereas single locus analysis might lead to misidentification. New strains can be quickly and reliably placed into this classification based on their whole genome sequences,

which will be valuable in forensic studies. Furthermore, we showed that whole genome sequences can be used to evaluate the specificity of PCR based diagnostic tools, including qPCR-assays developed for the detection and quantification of RSSC. This *in-silico* approach can give an indication for the choice of an assay for general detection of the RSSC or only of certain groups. Further studies will have to show if whole genome sequencing data can also be used to assess phenotypic traits, in particular those involved in host-pathogen interactions, as for example by the identification of virulence-related genes (Ailloud et al. 2015; Didelot et al. 2012).

Comparing whole genome sequences of 192 strains by means of an average nucleotide analysis (ANI) verified a clustering into three different species, *R. solanacearum*, *R. pseudosolanacearum* and *R. syzygii* as proposed by Safni et al. (2014), confirming current taxonomy. Nevertheless, some variation with respect to genome similarity within the species could be observed, especially within *R. solanacearum* and *R. pseudosolanacearum*. This variation reflects the phylotypes IIA and IIB of *R. solanacearum*, and I and III of *R. pseudosolanacearum*, respectively and is consistent with previous results obtained by ANI studies (Prior et al. 2016; Remenant et al. 2010). Prior et al. (2016) were able to identify the three different species using ANI in their set of 29 strains. It has been proposed that the phylotypes IIA and IIB, as well as I and III, should be defined as different subspecies of *R. solanacearum* and *R. pseudosolanacearum*, respectively (Zhang and Qiu 2016). The ANI analysis performed in the present study also showed three clusters within phylotype IIB. This heterogeneity is in accordance with several previous studies that find variation within the phylotype (Ailloud et al. 2015; Prior et al. 2016). It remains to be determined which genetic features are linked with the clustering in subgroups by ANI and if there is a relation between subgroups and (important) phenotypic properties. In addition, the ANI also revealed that two strains previously classified according to their similarity with a reference strain were misidentified as IIB, while they show a higher identity to phylotype IIA. It is unclear whether this finding is due to recombination events or the close relatedness of the two groups. Still our results emphasized the ability of ANI to correctly classify strains based on their whole genome sequence.

One strain, CRMRS218 could not be classified in one of the RSSC species described. It had a > 95% similarity with the two type strains of *R. pseudosolanacearum* and

R. solanacearum. This is in accordance with a recent study by Etmiani et al. (2020) who were not able to classify the strain into a phylotype using conserved signature indels either. It is possible that this strain might be a hybrid or the product of recombination events. However, it remains unclear which processes led to the observed genetic similarities. Phylogenomic or pangenomic analysis would be needed to clarify the mechanism behind this finding.

MLSA separated the 192 RSSC strains into three species and four phylotypes (I, II, III, and IV), which is in accordance with earlier studies and with ANI analysis (Prior et al. 2016; Wicker et al. 2011). This reproducibility between studies using variable sets of genes, and between different methods, shows that the reliability of MLSA for the classification of RSSC into phylotypes is less dependent on the number and choice of the barcoding genes. For all strains, *adk*, *leuS* and *rplB* sequences adequately predicted the RSSC phylotype. For *fliC*, *gyrB*, *gdhA*, *egl*, *mutS* and *ppsA*, it was found that several strains contained genes that were more similar to the respective loci of another phylotype. Especially, for strains classified within phylotypes IIA and IIB, and I and III respectively, these genes frequently resembled the other phylotype. This could be due to recombination, but also ancestral diversity that has not yet been completely characterized by researchers. It should be kept in mind that we might not yet have completely resolved the diversity of the RSSC, which might also influence detection using assays specific to a phylotype or species.

Genes found to be the most variable included *fliC* and *gyrB*. The virulence gene *fliC*, which is located on a megaplasmid, and also the housekeeping gene *gyrB* might have been subjected to horizontal gene transfer as reported previously (Wicker et al. 2011). Alternatively, the close relatedness of the phylotypes IIA and IIB, and I and III might lead to genes being classified as belonging to the other phylotype respectively if differences are due to single SNPs. Interestingly, Wicker et al. (2011) and Castillo and Greenberg (2007) found recombination mostly occurring in strains belonging to the phylotypes III and IV, while we could not find evidence of such recombination events. We did find genes that are more similar to the phylotypes IIB and IIA, respectively, which may be due to the relative high number of phylotype IIA and IIB strains compared to the phylotype III and IV strains used in our study. It is notable that the *fliC* gene was not detected in several

strains. Since this gene plays an important role in flagellar motility (Tans-Kersten et al. 2001), it is unlikely that it was missing from the respective strains. It is possible that the gene was not detected by alignment due to substantial sequence differences with the reference strains. Alternatively, the respective strains might possess another gene encoding for a protein similar to *fliC*. Still, while the grouping of strains based on individual loci might yield different results, MLSA based on five or more concatenated sequences enables a reliable grouping into the four phylotypes and thus, yields results that are in accordance with ANI.

Similar to ANI analysis, MLSA indicated a further clustering of several phylotypes into subgroups. The clustering of phylotype IIB into three groups is similar to the pattern revealed by ANI, indicating that the MLSA has sufficient resolution to reflect differences between the three groups. Notably, the strains in the subgroups b and c of IIB originated predominantly from banana and Heliconia, indicating that the genetic differences between these subgroups and the other strains of IIB might be connected to host specificity. This is further supported by the observation that all strains in subgroup b have *fliC* genes closer related to phylotype IIA, *fliC* being associated with virulence. In addition, the strains in these subgroups originated from South- and Mesoamerica, which also suggests genetic similarity due to evolution in close proximity under similar environmental conditions. Also other phylotypes, like phylotype IIA show some heterogeneity, but no clustering as clear as in phylotype IIB.

As in the ANI, a special case was represented by strain CRMRS218 that has formerly been classified as *R. pseudosolanacearum* by Albuquerque et al. (2017) and by comparison with reference sequences from five strains belonging to the different phylotypes. The MLSA revealed that while two genes were most closely related to phylotype I, the remaining genes grouped with phylotype IIA. This is in accordance with the ANI, which showed the highest similarity of CRMRS218 with phylotype IIA as well. It is not known which method was used by Albuquerque et al. (2017) for classification. Still, the ambiguity of this specific strain shows that the use of a single gene for classification, as has been done in the past, or even comparison to a limited number of references, can be misleading due to the heterogeneity of the RSSC. In contrast, MLSA with several genes as well as ANI that considers the whole genome, especially when combined with a large

number of genomes from all phylotypes are likely to perform better in classifying previously ambiguous strain.

Despite similarities between phylogenetic analyses of the RSSC over different studies, their comparability suffers from severe restrictions. Different studies tend to use different sets of strains with limited overlap. The strains used in the present study mostly belonged to the phylotypes IIB and I, whereas others included higher percentages of the phylotypes III or IIA (Castillo and Greenberg 2007; Safni et al. 2014). However, the number of whole genome sequences that can be used for comparison of strains is growing steadily and also this study adds a considerable number of available genomes to the public resources. Such increase in the availability of complete whole genome sequences will allow a wide range of phylogenetic and diagnostic analyses, while avoiding the need to choose specific genes for classification.

Selection of assays for the detection in the RSSC

Analysis of the presence of the target regions in all 225 strains revealed that while the rRNA region, the *egl* gene and the intergenic region between UDP-3-O-acyl-N-acetylglucosamine deacetylase & Peroxiredoxin were present in all strains, some targets were only present in a subset of strains, such as the regions targeted by the Guidot assays and the Phage-tail-S super family gene targeted by the Ozakman and WellerII assays. The percentage of strains per assay that were kept for analysis largely complied with the percentage of strains in which the gene is present. For several assays it could be observed that, although the target region was present, more than five mismatches in one primer were detected for some strains, which make amplification unlikely. Only for the Ozakman assay a fit with less than five mismatches per primer could be observed with all strains, while the target region was only present in a subset of strains. This might be due to primer binding to a non-target region, which is supported by the high number of SNPs for the whole assay that can be observed with most non-phylotype IIB strains.

In-silico analysis of the number of SNPs between primers and primer/probe sets showed that false negatives and false positives can be expected for several conventional PCR and TaqMan assays developed for the detection of all or parts of the RSSC. Among the PCR assays, the primer pairs by Chen, Pastrik, Opina, and Seal have been developed to target all groups of the RSSC. While the Seal assay had a low number of SNPs for all

phylotypes, for the Pastrik assay it could be seen that a number of strains from phylotype II and III and all strains of phylotype IV had more than 5 mismatches per primer with the target gene making an amplification unlikely (Lefever et al. 2013). Also, the Opina assay had a number of SNPs with all phylotypes, which are likely to reduce the efficiency of the assay, and the Chen primers show 4–6 SNPs with phylotype II. In contrast, the TaqMan assays developed for all phylotypes (Körner, Weller and Wu) showed few mismatches with most strains, suggesting that only few false negatives will occur. However, it should be noted that these assays might also yield false positive results. The primers by Wu et al. (2008) were originally developed for a broad range of Gram-negative bacteria and while the Seal, Körner and Weller assays were developed for the RSSC, they are likely to amplify the closely related species *R. insidiosa*, *R. mannitolilytica* and *R. pickettii* as well. Therefore, these assays should preferably be used when no background of these species can be expected.

Also, assays that were developed for specific groups of the RSSC might suffer from false negative or false positive results. With respect to PCR assays, the three primer sets developed by Guidot et al. (2009) were designed for the detection of phylotype IIB. However, here we observed that only strains from phylotype IIB group a, but not group b and c are likely to be detected due to more than five mismatches in one primer for these groups. Furthermore, for the series of Nmult primers, developed by Fegan and Prior (2005) in order to specifically detect the phylotypes I-IV, false negative results are not expected based on the number of SNPs detected in this study. Nevertheless, false positives might occur as Nmult1 also shows a low number of SNPs with most phylotype I and IV strains. Moreover, Nmult 2 and 3 could potentially also react with *R. insidiosa*, *R. mannitolilytica* and *R. pickettii*. Among the TaqMan assays, the Huang assay was designed for the phylotypes I, IIA and III, but the number of mismatches suggested that it is rather specific to phylotype I and might in addition detects phylotype IV. On the other hand, the Nytor assay had only few SNPs on average with all phylotypes except IV, which is in accordance with Vreeburg et al. (2018), indicating that few, if any false negatives will occur. No false positives can be expected with closely related species as the non-target strains analysed in this study did not possess the target gene. The assays Ozakman and Weller II, both targeting a bacteriophage mu-g-like protein sequence, are specific for phylotype IIB, which we also found in

the *in-silico* analysis, although only phylotype IIB group a is likely to be detected. The Massart assay targets the whole of phylotype II, but this assay as well is likely to miss phylotype IIB group c due to more than five mismatches. These differences in primer fit for the groups a, b and c of phylotype IIB further suggest that these subgroups differ substantially from one another, which should be taken into account while choosing or designing an assay.

It is likely that the number of SNPs does not perfectly predict primer fit and amplification. However, the suitability of SNPs to predict primer fit is supported by the finding that the average number of SNPs found between the phylotypes and the different assays is largely congruent with the ΔG values calculated for the respective reference strains. Only in a few cases did a few mismatches already lead to a strong increase in ΔG , suggesting that the effectiveness of the assay might be compromised. This was the case for the Chen assay and phylotype I. Therefore, based on our *in-silico* results, we could make suggestions for the most suitable assay for either general detection of the RSSC or of specific phylotypes. General detection of all phylotypes will most likely be accomplished by the PCR assay by Seal or the TaqMan assays by Weller and Körner, all based on the 16S rRNA gene as a target. While only few false negatives can be expected, false positives might occur in the presence of closely related species. If the presence of phylotype IV is unlikely, also the Nytor assay can be considered due to its lack of false positive results. For detection of phylotype I specifically, the Nmult1 and Huang assay both appeared to be suitable, but both might also detect phylotype IV. For all strains within phylotype II, Nmult 2 can be recommended as no mismatches were detected with any strain. From the number of phylotype IIB specific assays, none is likely to detect all groups of this phylotype. The largest group a can either be detected by the PCR assays by Guidot or the TaqMan assays by Ozakman and Weller. The only assay specific for phylotype III and phylotype IV in this study were Nmult3 and Nmult4, respectively.

In conclusion, accurate classification of strains can be done using ANI or MLSA, because single locus methods may give an inaccurate result due to the reticulate evolution of the target gene. Whole genomes have already been shown to be invaluable in phylogenetic, evolutionary, and niche adaptation studies. In this paper,

we demonstrated that they can also be used for evaluating and designing detection assays. All genome sequences used for this study are publicly available and that will foster reproducibility of the current results and comparability with future (phylogenetic) studies.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10658-020-02190-8>.

Acknowledgements We are very grateful to Miriam Kooman and Robert Vreeburg (NAK) and to Maria Bergsma and Leon Tjou-Tam-Sin (NVWA), for fruitful discussions and sharing bacterial strains used in this study.

We thank Elio Schijlen, Jan Cordewener and Sven Warris (BU Bioscience WUR) for their contribution in the sequencing of the isolates.

Funding information This research was funded by the Dutch Ministry of Economic Affairs in the Topsector Program “Horticulture and Starting Materials” under theme “Plant Health” in the projects Diagnostics (no. 1406–062), Fundament Phytosanitary Research (no. 1406–023 / 1410–023) and Ralstonia (no. 1605–082).

Data availability The datasets generated during the current study are not yet publicly available, but are available from the corresponding author on reasonable request. Sequencing reads and genome assemblies generated in this study are available at NCBI under PRJNA663270.

Compliance with ethical standards

Conflict of interest The authors are not aware of any conflicting interests.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Code availability Used software and functions are described in Material and Methods.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory

regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ailloud, F., Lowe, T., Cellier, G., Roche, D., Allen, C., & Prior, P. (2015). Comparative genomic analysis of *Ralstonia solanacearum* reveals candidate genes for host specificity. *BMC Genomics*, *16*(1), 270. <https://doi.org/10.1186/s12864-015-1474-8>.
- Albuquerque, G. M. R., Santos, L. A., Felix, K. C. S., Rollemberg, C. L., Silva, A. M. F., Souza, E. B., Cellier, G., Prior, P., & Mariano, R. L. R. (2014). Moko disease-causing strains of *Ralstonia solanacearum* from Brazil extend known diversity in paraphyletic phylotype II. *Phytopathology*, *104*(11), 1175–1182. <https://doi.org/10.1094/phyto-12-13-0334-r>.
- Albuquerque, G. M. R., Souza, E. B., Silva, A. M. F., Lopes, C. A., Boiteux, L. S., & Fonseca, M. E. D. N. (2017). Genome sequence of *Ralstonia pseudosolanacearum* strains with compatible and incompatible interactions with the major tomato resistance source Hawaii 7996. *Genome Announcements*, *5*(36), e00982–e00917. <https://doi.org/10.1128/genomeA.00982-17>.
- Arahal, D. R., Goodfellow, M., Sutcliffe, I. and Chun, J. (2014). Chapter 6 - whole-genome analyses: average nucleotide identity. In (Ed.), *Methods in Microbiology* (pp. 103-122). Academic Press. <https://doi.org/10.1016/bs.mim.2014.07.002>.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477.
- Buddenhagen, I. (1962). Designations of races in *Pseudomonas solanacearum*. *Phytopathology*, *52*, 726.
- Castillo, J. A., & Greenberg, J. T. (2007). Evolutionary dynamics of *Ralstonia solanacearum*. *Applied and Environmental Microbiology*, *73*(4), 1225–1238. <https://doi.org/10.1128/aem.01253-06>.
- Chen, Y., Zhang, W. Z., Liu, X., Ma, Z. H., Li, B., Allen, C., & Guo, J. H. (2010). A real-time PCR assay for the quantitative detection of *Ralstonia solanacearum* in the horticultural soil and plant tissues. *Journal of Microbiology and Biotechnology*, *20*(1), 193–201.
- Daligault, H. E., Davenport, K. W., Minogue, T. D., Broomall, S. M., Bruce, D. C., Chain, P. S., Coyne, S. R., Gibbons, H. S., Jaissle, J., Lo, C. C., Meincke, L., Munk, A. C., Rosenzweig, C. N., & Johnson, S. L. (2014). Draft genome assembly of *Ralstonia picketti* type strain K-288 (ATCC 27853). *Genome Announcements*, *2*(5), e00973–e00914. <https://doi.org/10.1128/genomeA.00973-14>.
- Delcher, A. L., Phillippy, A., Carlton, J., & Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Research*, *30*(11), 2478–2483. <https://doi.org/10.1093/nar/30.11.2478>.
- Denny, T. (2006). Plant pathogenic *Ralstonia* species. In S. S. Gnanamanickam (Ed.), *Plant-associated Bacteria* (pp. 573–644). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-4538-7_16.
- Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A., & Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nature Reviews Genetics*, *13*, 601. <https://doi.org/10.1038/nrg3226>.
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>.
- Etmnani, F., Yousefvand, M., & Harighi, B. (2020). Phylogenetic analysis and molecular signatures specific to the *Ralstonia solanacearum* species complex. *European Journal of Plant Pathology*, *158*(1), 261–279. <https://doi.org/10.1007/s10658-020-02073-y>.
- Fegan, M., & Prior, P. (2005). How complex is the *Ralstonia solanacearum* species complex. In C. Allen, P. Prior, & A. Hayward (Eds.), *Bacterial wilt disease and the Ralstonia Solanacearum species complex* (pp. 449–461). Minnesota: APS press.
- Fox, J., & Weisberg, S. (2011). *An {R} companion to applied regression (trans)*. Thousand Oaks: Sage.
- Genin, S., & Denny, T. P. (2012). Pathogenomics of the *Ralstonia solanacearum* species complex. *Annual Review of Phytopathology*, *50*(1), 67–89. <https://doi.org/10.1146/annurev-phyto-081211-173000>.
- Glaeser, S. P., & Kämpfer, P. (2015). Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Systematic and Applied Microbiology*, *38*(4), 237–245. <https://doi.org/10.1016/j.syam.2015.03.007>.
- Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, *57*(1), 81–91. <https://doi.org/10.1099/ijs.0.64483-0>.
- Guidot, A., Elbaz, M., Carrère, S., Siri, M. I., Pianzola, M. J., Prior, P., & Boucher, C. (2009). Specific genes from the potato brown rot strains of *Ralstonia solanacearum* and their potential use for strain detection. *Phytopathology*, *99*(9), 1105–1112.
- Hayward, A. C. (1964). Characteristics of *Pseudomonas solanacearum*. *Journal of Applied Bacteriology*, *27*(2), 265–277. <https://doi.org/10.1111/j.1365-2672.1964.tb04912.x>.
- Huang, J., Wu, J., Li, C., Xiao, C., & Wang, G. (2009). Specific and sensitive detection of *Ralstonia solanacearum* in soil with quantitative, real-time PCR assays. *Journal of Applied Microbiology*, *107*(5), 1729–1739. <https://doi.org/10.1111/j.1365-2672.2009.04364.x>.
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 5114. <https://doi.org/10.1038/s41467-018-07641-9>.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736.
- Körner, U. A., Kroll, U., & Hilgert, S. (2017). False positive *Ralstonia solanacearum* real-time PCR results in routine

- potato tuber samples caused by *Advenella* species: adaptations to avoid these cross-reactions. *EPPO Bulletin*, 47(1), 33–40. <https://doi.org/10.1111/epp.12366>.
- Lefever, S., Pattyn, F., Hellemans, J., & Vandesompele, J. (2013). Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays. *Clinical Chemistry*, 59(10), 1470–1480. <https://doi.org/10.1373/clinchem.2013.203653>.
- Lenth, R., Singmann, H., Love, J., Buerkner, P. and Herve, M. (2018) 'emmeans: estimated marginal means, aka least-square means' 1.2.3. p.^pp. Available at: <https://github.com/rvlenh/emmeans> (Accessed).
- Mansfield, J., et al. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology*, 13(6), 614–629. <https://doi.org/10.1111/j.1364-3703.2012.00804.x>.
- Massart, S., Nagy, C., & Jijakli, M. H. (2014). Development of the simultaneous detection of *Ralstonia solanacearum* race 3 and *Clavibacter michiganensis* subsp. *sepedonicus* in potato tubers by a multiplex real-time PCR assay (journal article). *European Journal of Plant Pathology*, 138(1), 29–37. <https://doi.org/10.1007/s10658-013-0294-4>.
- Opina, N., et al. (1997). A novel method for development of species and strain-specific DNA probes and PCR primers for identifying *Burkholderia solanacearum* (formerly *Pseudomonas solanacearum*). *Asia-Pacific Journal of Molecular Biology and Biotechnology*, 5(1), 19–30.
- Ozakman, M., & Schaad, N. W. (2003). A real-time BIO-PCR assay for detection of *Ralstonia solanacearum* race 3, biovar 2, in asymptomatic potato tubers. *Canadian Journal of Plant Pathology*, 25(3), 232–239. <https://doi.org/10.1080/07060660309507075>.
- Patrik, K.-H., Elphinstone, J. G., & Pukall, R. (2002). Sequence analysis and detection of *Ralstonia solanacearum* by multiplex PCR amplification of 16S–23S ribosomal intergenic spacer region with internal positive control (journal article). *European Journal of Plant Pathology*, 108(9), 831–842. <https://doi.org/10.1023/a:1021218201771>.
- Prior, P., Ailloud, F., Dalsing, B. L., Remenant, B., Sanchez, B., & Allen, C. (2016). Genomic and proteomic evidence supporting the division of the plant pathogen *Ralstonia solanacearum* into three species (journal article). *BMC Genomics*, 17(1), 90. <https://doi.org/10.1186/s12864-016-2413-z>.
- Remenant, B., Coupat-Goutaland, B., Guidot, A., Cellier, G., Wicker, E., Allen, C., Fegan, M., Pruvost, O., Elbaz, M., Calteau, A., Salvignol, G., Mormico, D., Mangenot, S., Barbe, V., Médigue, C., & Prior, P. (2010). Genomes of three tomato pathogens within the *Ralstonia solanacearum* species complex reveal significant evolutionary divergence. *BMC Genomics*, 11(1), 379. <https://doi.org/10.1186/1471-2164-11-379>.
- Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences*, 106(45), 19126–19131. <https://doi.org/10.1073/pnas.0906412106>.
- Safni, I., Cleenwerck, I., De Vos, P., Fegan, M., Sly, L., & 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 *Ralstoniapseudosolanacearum* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 64(9), 3087–3103. <https://doi.org/10.1099/ijs.0.066712-0>.
- Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brottier, P., Camus, J.C., Cattolico, L., Chandler, M., Choisine, N., Claudel-Renard, C., Cunmac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T., Siguier, P., Thébault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J., Boucher, C.A. (2002). Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature*, 415, 497. <https://doi.org/10.1038/415497a>. <https://www.nature.com/articles/415497a#supplementary-information>.
- Schlesner, M., Gu, Z., & Eils, R. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18), 2847–2849. <https://doi.org/10.1093/bioinformatics/btw313>.
- Schönfeld, J., Gelsomino, A., van Overbeek, L. S., Gorissen, A., Smalla, K., & van Elsas, J. D. (2003). Effects of compost addition and simulated solarisation on the fate of *Ralstonia solanacearum* biovar 2 and indigenous bacteria in soil. *FEMS Microbiology Ecology*, 43(1), 63–74. <https://doi.org/10.1111/j.1574-6941.2003.tb01046.x>.
- Seal, S. E., Jackson, L. A., Young, J. P. W., & Daniels, M. J. (1993). Differentiation of *Pseudomonas solanacearum*, *Pseudomonas syzygii*, *pseudomonas pickettii* and the blood disease bacterium by partial 16S rRNA sequencing: construction of oligonucleotide primers for sensitive detection by polymerase chain reaction. *Microbiology*, 139(7), 1587–1594. <https://doi.org/10.1099/00221287-139-7-1587>.
- Sepey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: assessing genome assembly and annotation completeness. In M. Kollmar (Ed.), *Gene prediction: Methods and protocols* (pp. 227–245). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-9173-0_14.
- Suzuki, M., Nishio, H., Asagoe, K., Kida, K., Suzuki, S., Matsui, M., & Shibayama, K. (2015). Genome sequence of a carbapenem-resistant strain of *Ralstonia mannitolilytica*. *Genome Announcements*, 3(3). <https://doi.org/10.1128/genomeA.00405-15>.
- T. RStudio (2016) 'RStudio: Integrated Development Environment for R'. Boston, MA: RStudio Inc., p.^pp. Available at: <http://www.rstudio.com/> (Accessed).
- Tans-Kersten, J., Huang, H., & Allen, C. (2001). *Ralstonia solanacearum* needs motility for invasive virulence on tomato. *Journal of Bacteriology*, 183(12), 3597–3605. <https://doi.org/10.1128/jb.183.12.3597-3605.2001>.
- van Elsas, J. D., Kastelein, P., van Bekkum, P., van der Wolf, J. M., de Vries, P. M., & van Overbeek, L. S. (2000). Survival of *Ralstonia solanacearum* biovar 2, the causative agent of potato brown rot, in field and microcosm soils in temperate climates. *Phytopathology*, 90(12), 1358–1366. <https://doi.org/10.1094/phyto.2000.90.12.1358>.
- Vreeburg, R. A. M., Zendman, A. J. W., Pol, A., Verheij, E., Nas, M., & Kooman-Gersmann, M. (2018). Validation of four real-time TaqMan PCRs for the detection of *Ralstonia solanacearum* and/or *Ralstonia pseudosolanacearum* and/or

- Clavibacter michiganensis* subsp. *sepedonicus* in potato tubers using a statistical regression approach. *EPPO Bulletin*, 48(1), 86–96. <https://doi.org/10.1111/epp.12444>.
- Weller, S. A., Elphinstone, J. G., Smith, N. C., Boonham, N., & Stead, D. E. (2000). Detection of *Ralstonia solanacearum* strains with a quantitative, multiplex, real-time, fluorogenic PCR (TaqMan) assay. *Applied and Environmental Microbiology*, 66(7), 2853–2858. <https://doi.org/10.1128/aem.66.7.2853-2858.2000>.
- Wicker, E., Lefeuvre, P., de Cambiaire, J.-C., Lemaire, C., Poussier, S. and Prior, P. (2011). Contrasting recombination patterns and demographic histories of the plant pathogen *Ralstonia solanacearum* inferred from MLSA. *The ISME Journal*, 6, 961. <https://doi.org/10.1038/ismej.2011.160>, <https://www.nature.com/articles/ismej2011160#supplementary-information>.
- Wu, Y.-D., Chen, L. H., Wu, X. J., Shang, S. Q., Lou, J. T., du, L. Z., & Zhao, Z. Y. (2008). Gram stain-specific-probe-based real-time PCR for diagnosis and discrimination of bacterial neonatal sepsis. *Journal of Clinical Microbiology*, 46(8), 2613–2619. <https://doi.org/10.1128/jcm.02237-07>.
- Yeo, S., Coombe, L., Warren, R. L., Chu, J., & Birol, I. (2017). ARCS: Scaffolding genome drafts with linked reads. *Bioinformatics*, 34(5), 725–731. <https://doi.org/10.1093/bioinformatics/btx675>.
- Zhang, Y., & Qiu, S. (2016). Phylogenomic analysis of the genus *Ralstonia* based on 686 single-copy genes. *Antonie Van Leeuwenhoek*, 109(1), 71–82. <https://doi.org/10.1007/s10482-015-0610-4>.
- Zhang, L., Morrison, M., & Rickard, C. M. (2014). Draft genome sequence of *Ralstonia pickettii* AU12-08, isolated from an intravascular catheter in Australia. *Genome Announcements*, 2(1), e00027–e00014. <https://doi.org/10.1128/genomeA.00027-14>.