OSIEL SILVA GONÇALVES

ON THE MOVE: MOBILOME OF THE *Ralstonia solanacearum* SPECIES COMPLEX AND ITS IMPACT ON FITNESS, PATHOGENICITY AND GENOME EVOLUTION

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Microbiologia Agrícola, para obtenção do título de *Magister Scientiae*.

Orientador: Mateus Ferreira Santana

Coorientadoras: Marisa Vieira de Queiroz
Denise Mara Soares Bazolli

VIÇOSA - MINAS GERAIS
2019

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T

G635o
2019

Gonçalves, Osiel Silva, 1994-
On the move: mobilome of the *Ralstonia solanacearum* species complex and its impact on fitness, pathogenicity and genome evolution / Osiel Silva Gonçalves. - Viçosa, MG, 2019.
91 f. : il. (algumas color.) ; 29 cm.

Texto em inglês.
Orientador: Mateus Ferreira Santana.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Elementos genéticos móveis. 2. Elementos de DNA transponíveis. 3. Bactérias fitopatogênicas - Genética. I. Universidade Federal de Viçosa. Departamento de Microbiologia. Programa de Pós-Graduação em Microbiologia Agrícola. II. Título.

CDD 22. ed. 579.332

OSIEL SILVA GONÇALVES

ON THE MOVE: MOBILOME OF THE *Ralstonia solanacearum* SPECIES COMPLEX AND ITS IMPACT ON FITNESS, PATHOGENICITY AND GENOME EVOLUTION

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Microbiologia Agrícola, para obtenção do título de *Magister Scientiae*.

APROVADA: 26 de julho de 2019.

Assentimento:

Osiel Silva Gonçalves
Autor

Mateus Ferreira Santana
Orientador

Dedico essa obra àqueles que desde o princípio acreditaram que ela seria possível, minha tão amada mãe Selma, meu querido pai Carlos, meu inseparável irmão Carlos (Dudu). A vocês e por vocês.

ACKNOWLEDGMENT

A Deus e aos Mikisi pela fé, força, dedicação, e todo coragem para trilhar essa jornada;

À minha família, meus três pilares. Minha mãe Selma, por ter sido e continuar sendo meu pilar de força, acalanto, que abrira sua mão para que seu colibri voasse e conquistasse novos ares. Que apesar da distância, se fez presente a cada instante. Ao meu pilar de perseverança, meu querido Pai Carlos, aquele que sempre comprou minhas loucuras e nunca hesitou em fazer o possível para minha educação e do meu irmão, meu exemplo de honestidade e de homem. E ao meu pilar de irmandade, meu caçula Carlos (Dudu), difícil de falar dele, mas representa um sentimento puro de bondade, amor; o meu orgulho!

A meu tio Maurício e Val, meu casal preferido, pela força, torcida, nossos momentos de descontração e aconchego.

Ao Professor Mateus Ferreira Santana, pela confiança depositada para executar esse trabalho, dedicação e comprometimento em compartilhar seu conhecimento e pela sua participação efetivamente nessa importante etapa de minha formação profissional.

À Professora Marisa Viera de Queiroz, pelos sábios conselhos de biologia molecular, pela acolhida no Laboratório de Genética Molecular de Fungos (LGMF), e pelo exemplo de profissional dedicada à pesquisa.

Aos meus ilustres e eternizados amigos do LGMF. Deixaram mais leve o peso do meu coração. Fizeram cada segundo no laboratório valer a pena. A minha querida amiga, Mirele (*miaD*), que iniciou essa jornada comigo e logo tratou de se instalar no aqui no lado esquerdo do peito. Ao Leandro (Léo), pensa num menino chato! Mas é grande em generosidade, o que compensa os inúmeros defeitos que ele tem. Ao Felipe, uma pessoa que gostamos de graça, nossa fábrica de memes, “dona da minha vida” (ele vai entender a referência), voa longe, menino. A “ala internacional do LGMF”, Ângela e Hanna, pelas trocas culturais, apoio e todo carinho. Lucas (Luquinhas) pela recepção desde o princípio no LGMF e pela aprendizagem na

área de bioinformática. À Kiara, pela experiência de coorientá-la durante sua iniciação científica, o que foi uma experiência linda! E pela colaboração nesta obra. A meu amigo Ediones, pelas nossas prosas sobre qualquer assunto – um repertório enorme. A Thamires e Luís, pela experiência de orientá-los e pela contribuição nesse trabalho. Vocês foram, são e serão os meus melhores.

Aos meus colegas das Pós-Graduação, Marlon, Isabel e Adriele, pela oportunidade de troca e consolo nas horas mais difíceis das disciplinas. Assim como a Jessica, e o Giarlã, do Laboratório de Genética Molecular de Bactérias, pelas risadas e conhecimentos compartilhados.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e a Fundação Arthur Bernardes (Funarbe);

À Universidade Federal de Viçosa (UFV), pela oportunidade de realização deste curso;

À coordenação administrativa do Programa de Pós-Graduação em Microbiologia Agrícola;

Ao corpo técnico-administrativo do Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO);

A todos vocês, meus mais singelos agradecimentos!

ABSTRACT

GONÇALVES, Osiel Silva, M.Sc., Universidade Federal de Viçosa, July, 2019. **On the move: Mobilome of the *Ralstonia solanacearum* Species Complex and its role in the fitness, pathogenicity and genome evolution**. Adviser: Mateus Ferreira Santana. Co-advisers: Marisa Vieira de Queiroz and Denise Mara Soares Bazzolli

Ralstonia solanacearum is one of the most devastating plant pathogenic bacteria found all over the world, capable of infecting a wide diversity of hosts. This soil-borne pathogen is composed of a large-scale group of strains varying in geographical distribution and pathogenic behavior known as *R. solanacearum* Species Complex (RSSC). In this context, mobile genetic elements may play a significant role in shaping the genetic, and these elements may be important to the bacteria expand their ecological niche or colonize a new host since they promote genetic variability and often carry a cargo gene. In this context, we performed a genome mining in 106 RSSC and 15 *Ralstonia* spp. genomes and comparative genomic analysis based on Integrative and Conjugative Elements (ICEs), Genomic Islands (GIs), Insertion Sequences (ISs) and Transposons. Our results provided a collective dataset of 41 GIs and 12 ICEs in the RSSC. These elements represent an important fraction of the *Ralstonia* genomes (1-5%). Phylogenetic analysis showed that the elements are related to geographic origin and species, but there is evidence of genetic flux between isolates belonging to different countries from the Americas and Horizontal Gene Transfer between strains from Asia. ICEs and GIs carry a repertoire of genes with potential impact on *Ralstonia* fitness and pathogenicity such as stress response and candidate virulence genes as several type III effector proteins which may enable the bacteria to evolve rapidly and infect a wide range of hosts. In addition, we identified 2,957 IS elements in the genome of 59 representative RSSC strains and closely related *Ralstonia* spp. A unique set of 13 IS families were found highly conserved across the strains, suggesting their ancestral acquisition. We found four novel transposons sequences belong to Tn3 family carrying genes related to antibiotic resistance and avirulence protein as passenger genes. Numerous internal rearrangements events with a subset being associated with IS were demonstrated, indicating the impact of these elements on RSSC genome evolution. We also mapped numerous IS elements interrupting avirulence genes, which provides evidence that IS may be one of the driving forces of RSSC pathogenicity evolution. In conclusion, here we provide a dataset of the mobile genetic elements in one of the most important plant pathogens found all over the world. These data provide novel insights into the RSSC diversified adaptation, opening new paths to a better

understanding of the impact of these elements on the genome evolution and pathogenicity of *R. solanacearum*.

Keywords: Genome evolution. Phytopathogen. ICEs. Genomic islands. Insertion Sequence

RESUMO

GONÇALVES, Osiel Silva, M.Sc., Universidade Federal de Viçosa, julho de 2019. **Em movimento: Mobiloma do complexo de espécies de *Ralstonia solanacearum* e sua importância na adaptação, patogenicidade e evolução do genoma**. Orientador: Mateus Ferreira Santana. Coorientadoras: Marisa Vieira de Queiroz e Denise Mara Soares Bazzolli.

Ralstonia solanacearum é uma das bactérias fitopatogênicas mais devastadoras a nível mundial, capaz de infectar uma grande diversidade de hospedeiros. É transmitida pelo solo e apresenta linhagens heterogêneas, que variam em distribuição geográfica e patogenicidade, conhecido como Complexo de Espécies de *R. solanacearum* (RSSC). Considerando esta plasticidade genética, os elementos genéticos móveis (MGEs) podem desempenhar um importante papel na diversificação genômica do fitopatógeno. Esses elementos podem ser importantes para a bactéria expandir nichos ecológicos ou colonizar novos hospedeiros. Neste contexto, este estudo teve como objetivo identificar os elementos de Ilhas genômicas (GIs), Elementos integrativos e conjugativos (ICEs), Sequências de inserção (ISs) e Transposons nos genomas de 106 linhagens do RSSC e 15 *Ralstonia* spp. Um conjunto de dados de 40 GIs e 12 ICEs foram identificados para o RSSC. Esses elementos representam uma fração importante dos genomas de *Ralstonia* (1-5%). A análise filogenética mostrou que estes elementos são associados à origem geográfica e/ou às espécies de *Ralstonia* spp. Evidências de fluxo genético entre isolados pertencentes a diferentes países das Américas e transferência horizontal de genes entre linhagens da Ásia foram relatadas. ICEs e GIs carregam um repertório de genes com potencial impacto na adaptação e patogenicidade de *Ralstonia*, incluindo genes de resposta ao estresse e genes de virulência, como várias proteínas efetoras do tipo III que podem permitir a bactéria uma rápida adaptação e infecte uma ampla gama de hospedeiros. Além disso, foram identificados 2.957 elementos de IS no genoma de 59 linhagens representativas do RSSC e *Ralstonia* spp. Um conjunto de 13 famílias IS foi encontrado altamente conservado através das linhagens, sugerindo sua aquisição ancestral. Foram identificadas quatro novas sequências de transposons pertencentes à família Tn3 portadores de genes relacionados à resistência a antibióticos e proteínas de avirulência. Numerosos eventos de rearranjos internos associado ao ISs foram demonstrados, indicando o impacto destes elementos na evolução do genoma da RSSC. Também foi mapeado vários elementos IS interrompendo gene de avirulência, o que fornece evidências que ISs pode ser uma das forças motrizes da evolução da patogenicidade do RSSC. Em conclusão, foi reportado um conjunto de dados associados a elementos genéticos

móveis de um dos mais importantes fitopatógenos. Esses dados fornecem novas perspectivas sobre a adaptação do RSSC, abrindo novos caminhos para uma melhor compreender o impacto desses elementos na evolução do genoma e na patogenicidade de *R. solanacearum*.

Palavras-chave: Genômica evolutiva. Fitopatógeno. ICEs. Ilhas genômicas. Sequência de inserção

SUMMARY

GENERAL INTRODUCTION

Due to the next generation of sequencing technologies, substantial growth in complete genomes has revolutionized Genomics (Metzker, 2009). These genomes are publicly available opening insights into the structural and functional analysis of a variety of organisms in different areas of knowledge, for example in phytopathology (Klosterman et al., 2016). Likewise, comparative analyses of genomes have enabled researchers to investigate not only the genes involved in the plant-pathogen interaction, but also the genetic variability mechanisms of these pathogens (Sundin et al., 2016).

The bacterium *Ralstonia solanacearum* is one of the most devastating phytopathogens in economically important crops, able to infect around 250 hosts in 54 botanical families (Hayward, 1991; Mansfield et al., 2012). It is a Gram-negative, non-sporulating, soil-dwelling aerobic bacterium which invades the roots of the plants and lodges in the xylem, blocking flow of water to the leaves, causing the plant to wilt and, consequently, death (Denny, 2006).

Ralstonia solanacearum has variable strains due its genetic variability and a broad host range (Fegan & Prior, 2005). Therefore, *R. solanacearum* is known as *R. solanacearum* species complex (RSSC), which also includes the species *R. syzygii* and Blood Disease Bacteria (BDB). It is classified, respectively, in races, biovars, and phylotypes according to host range, physiological characterization and geographical origin, respectively (Buddenhagen et al., 1962; Fegan & Prior, 2005). Recently it was proposed that the RSSC comprises three distinct species: *R. pseudosolanacearum* (formerly phylotypes I and III), *R. solanacearum* (IIA and IIB) and *R. syzygii* (formerly phylotypes IV and BDB) (Safni et al. 2014, Prior et al. 2016).

In general, the taxonomic classification and pathogenesis of the RSSC is a reflection of the phenotypic and genotype variation of the species. The genome of the species is organized into two replicons: a chromosome containing on average 3.7 Mb and a megaplasmid containing

2.1 Mb. Both replicons can encode essential genes to the bacterium as well as accessory genes. Comparison of the sequences of the two replicons revealed similarities in the relative abundance of dinucleotides, codon usage and sequences of simple repetitions, which leads us to believe in a coevolution process between the replicons (Coenye & Vandamme, 2003; Guidot et al., 2007).

Comparative genomic analyses between representative strains of *R. solanacearum* have shown that characteristics such as size, number of genes and G+C content of the genome are conserved (Remenant et al., 2010; Li et al., 2016). In contrast, some rearrangements (e.g., inversion and translocation), deletions and insertions of exogenous DNA between species are noted, which supports the differentiation of the complex in species (Prior et al. 2016). Interestingly, Salanoubat et al. (2002) related the evolution of virulence in the GMI1000 genome, a strain of the species, to alternative codon usage regions (ACURs). These regions are flanked by mobile genetic elements (MGEs), such as prophages and insertion sequence (ISs), suggesting a possible mechanism for acquisition by horizontal gene transfer.

MGEs are segments of DNA which encode proteins that contribute to their own intracellular or intercellular mobility (Frost et al., 2005). The total MGEs present in the genome constitute the mobilome, which in the context of this study include: genomic islands (GIs), integrative and conjugative elements (ICEs), transposons and ISs. GIs are regions of the genome characterized by a G+C content different from the rest of genome; presence of direct repetitions at their extremities; generally associated with tRNA genes; have an integrase or transposase; use of a preferential codon (Lawrence & Ochman, 1998; Hacker et al., 2001). It is known that GIs contribute to the evolution of both pathogenic and non-pathogenic bacteria, and are referred to as saprophytic islands, symbiosis islands, resistance islands or pathogenicity islands (PAIs) (Hacker & Carniel, 1997; Juhas et al., 2009).

Similar to GIs, ICEs are elements found integrated into the genome which may undergo intercellular and intracellular mobility in the conjugative state or be transmitted to a

new cell (Burrus & Waldor, 2004; Carraro & Burrus, 2015). In addition to being transmitted by horizontal transfer, ICEs are passively propagated by vertical transfer during replication, segregation, and cell division. Structurally, ICEs contain a conserved central region, interspersed with genes not involved in transposition, which vary between ICEs, and often reflect the host's lifestyle (Seth-Smith & Croucher, 2009; Johnson & Grossman, 2015).

Another set of elements analyzed here are transposable elements, which are segments of DNA able to move in the genome (Frost et al., 2005). The role of these elements in the evolution of several species has been reported as they can increase plasticity and genomic diversity, modify gene structure and to be important sources of regulatory sequences (Shapiro, 2010). In bacteria, the main representatives of transposable elements are the ISs. These are considered the smallest and most numerous transposable elements, sized 700-2500 bp which essentially encode genes related to their motility, which includes a recombinase/transposase and, in some cases, a regulatory protein. ISs can be transposed by the mechanisms of conservative or replicative transposition when, respectively, excision of the element from the original locus and integration into a new locus (cut and paste) occurs, or when a new copy of the element is generated at the target site and that copy is integrated into a new locus (Snyder & Champness, 2007).

Overall, mobilome may become one of the main reservoirs for the acquisition/dissemination of pathogenicity genes, which include those coding for specialized secretion systems, phytotoxins, phytohormones, extracellular proteins, plant surface adhesion structures, lipopolysaccharides, exopolysaccharides (EPS) and micronutrient uptake systems (Arnold et al., 2003; Jackson et al., 2011). It is known that MGEs are important agents for the genetic variability of many pathogens. Therefore, investigating the dynamics of these elements in the genomes of *R. solanacearum*, will help us understand the genomic diversification that allows the phytopathogen to avoid or overcome host plant defense responses. Considering the

genetic plasticity of the RSSC, the mobilome may play an important evolutionary role in variability and pathogenicity. However, little is known about the RSSC MGEs and before this study there was no record on its mobilome. Thus, this study aimed to identify GIs, ICEs, ISs and transposons present in 121 genomes of the RSSC and related *Ralstonia* spp., and to understand the evolutionary impact of these elements on the pathogenicity and host adaptation of this important plant pathogen.

REFERENCES

ARNOLD, D. L.; PITMAN, A.; JACKSON, R. W. Pathogenicity and other genomic islands in plant pathogenic bacteria. **Molecular Plant Pathology**, v. 4, n. 5, p. 407–420, set. 2003.

BURRUS, V.; WALDOR, M. K. Shaping bacterial genomes with integrative and conjugative elements. **Research in Microbiology**, v. 155, n. 5, p. 376–386, 1 jun. 2004.

BUDDENHAGEN, I.; SEQUEIRA, L.; KELMAN, A. Designation of races in *Pseudomonas solanacearum*. **Phytopathology**, v. 52, p. 726, 14 set. 1962.

CARRARO, N.; BURRUS, V. The dualistic nature of integrative and conjugative elements. **Mobile Genetic Elements**, v. 5, n. 6, p. 98–102, 21 out. 2015.

COENYE, T.; VANDAMME, P. Simple sequence repeats and compositional bias in the bipartite *Ralstonia solanacearum* GMI1000 genome. **BMC Genomics**, v. 4, n. 10, p. 1–9, 2003.

DENNY, T. Plant pathogenic *Ralstonia* species. In: GNANAMANICKAM, S. S. (Ed.). Plant-Associated Bacteria. Dordrecht: **Springer Netherlands**, 2006. p. 573–644.

FEGAN, M.; PRIOR, P. How Complex is the *Ralstonia solanacearum* species complex. In: Allen C, Prior P, Hayward AC, eds. Bacterial Wilt Disease and the *Ralstonia solanacearum* Species Complex. St. Paul, MN: **American Phytopathological Society**; 2005:449–462.

FROST, L. S. et al. Mobile genetic elements: the agents of open source evolution. **Nature Reviews Microbiology**, v. 3, p. 722, 1 set. 2005.

GUIDOT, A. et al. Genomic structure and phylogeny of the plant pathogen *Ralstonia solanacearum* inferred from gene distribution analysis. **Journal of Bacteriology**, v. 189, n. 2, p. 377–387, 15 jan. 2007.

HAYWARD, A. C. Biology and epidemiology of bacterial wilt caused by *Pseudomonas solanacearum*. **Annual Review of Phytopathology**, v. 29, n. 1, p. 65–87, 1 set. 1991.

HACKER, J.; CARNIEL, E. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. **EMBO Rep**, v. 2, 2001.

HACKER, J. et al. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. **Molecular Microbiology**, v. 23, n. 6, p. 1089–1097, 1 mar. 1997.

JACKSON, R. W. et al. The influence of the accessory genome on bacterial pathogen evolution. **Mobile Genetic Elements**, v. 1, n. 1, p. 55–65, 1 maio 2011.

JOHNSON, C. M.; GROSSMAN, A. D. Integrative and conjugative elements (ICEs): what they do and how they work. **Annual Review of Genetics**, v. 49, n. 1, p. 577–601, 23 nov. 2015.

JUHAS, M. et al. Genomic islands: tools of bacterial horizontal gene transfer and evolution. **FEMS Microbiology Reviews**, v. 33, n. 2, p. 376–393, 30 mar. 2009.

KLOSTERMAN, S. J. et al. Disease Management in the Genomics Era—Summaries of Focus Issue Papers. **Phytopathology**, v. 106, n. 10, p. 1068–1070, 2016.

LAWRENCE, J. G.; OCHMAN, H. Molecular archaeology of the *Escherichia coli* genome. **Proceedings of the National Academy of Sciences U.S.A**, v. 95, n. 16, p. 9413–9417, 1998.

LI, P. et al. Genomic Analysis of Phylotype I Strain EP1 Reveals Substantial Divergence from Other Strains in the *Ralstonia solanacearum* Species Complex. **Frontiers in Microbiology**, v. 7, p. 1719, 26 out. 2016.

MANSFIELD, J. et al. Top 10 plant pathogenic bacteria in molecular plant pathology. **Molecular Plant Pathology**, v. 13, n. 6, p. 614–629, 1 ago. 2012.

METZKER, M. L. Sequencing technologies — the next generation. **Nature Reviews Genetics**, v. 11, p. 31, 8 dez. 2009.

PRIOR, P. et al. Genomic and proteomic evidence supporting the division of the plant pathogen *Ralstonia solanacearum* into three species. **BMC Genomics**, v. 17, p. 90, 1 fev. 2016.

REMENANT, B. et al. Genomes of three tomato pathogens within the *Ralstonia solanacearum* species complex reveal significant evolutionary divergence. **BMC Genomics**, v. 11, p. 379, 15 jun. 2010.

SAFNI, I. et al. 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. s. **International Journal of Systematic and Evolutionary Microbiology**, v. 64, n. Pt 9, p. 3087–3103, 2014.

SALANOUBAT, M. et al. Genome sequence of the plant pathogen *Ralstonia solanacearum*. **Nature**, v. 415, p. 497, 31 jan. 2002.

SETH-SMITH, H.; CROUCHER, N. J. Breaking the ICE. **Nature Reviews Microbiology**, v. 7, p. 328, 1 maio 2009.

SHAPIRO, J. A. Mobile DNA and evolution in the 21st century. **Mobile DNA**, v. 1, p. 4, 25 jan. 2010.

SIEFERT, J. L. Defining the Mobilome. In: GOGARTEN, M. B.; GOGARTEN, J. P.; OLENDZENSKI, L. C. (Eds.). **Horizontal Gene Transfer: Genomes in Flux**. Totowa, NJ: Humana Press, 2009. p. 13–27.

SIGUIER, P.; GOURBEYRE, E.; CHANDLER, M. Bacterial insertion sequences: their genomic impact and diversity. **FEMS Microbiology Reviews**, v. 38, n. 5, p. 865–891, 1 set. 2014.

SNYDER, L.; CHAMPNESS, W. **Molecular Genetics of Bacteria**. 3rd Edition ed. Washington: ASM Press, 2007.

SUNDIN, G. et al. Perspectives on the transition from bacterial phytopathogen genomics studies to applications enhancing disease management: from promise to practice. **Phytopathology**, v.106. n. 10, p. 1071-1082, jul. 2016.

CHAPTER 1

Integrative and Conjugative Elements (ICEs) and Genomic Islands reveal potential evolutionary impact to enhance fitness and pathogenicity in the *Ralstonia solanacearum* Species Complex

Gonçalves, O.S, Queiroz, M.V., Santana, M.F., Integrative and Conjugative Elements (ICEs) and Genomic Islands reveal potential evolutionary impact to enhance the fitness and pathogenicity in the *Ralstonia solanacearum* Species Complex**.** In preparation.

Integrative and Conjugative Elements (ICEs) and Genomic Islands reveal potential evolutionary impact to enhance fitness and pathogenicity in the *Ralstonia solanacearum* Species Complex

Osiel Silva Gonçalves[1]; Marisa Vieira de Queiroz[1], Mateus Ferreira Santana[1]*

[1]Departamento de Microbiologia, Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), Universidade Federal de Viçosa, Viçosa, MG, 36570-000; Brazil

*Corresponding author: Mateus Ferreira Santana

E-mail address: mateus.santana@ufv.br

Phone: +55 (31) 3612-2452

Abstract

Ralstonia solanacearum is one of the most devastating plant pathogenic bacteria found all over the world. This soil-borne pathogen is composed of a large-scale group of strains varying in genetic viability and pathogenic behavior known as *R. solanacearum* Species Complex (RSSC). In this context, mobile genetic elements (MGEs) may play an important role in shaping RSSC host range and adaptation to plants. Here, we performed a genome mining in 106 RSSC and 15 *Ralstonia* spp. genomes and comparative genomic analysis based on Integrative and Conjugative Elements (ICEs) and Genomic Islands (GIs). Overall, our results provided a collective dataset of 40 GIs and 12 ICEs in the RSSC. These elements represent an important fraction of the RSSC genomes (1-5%). Phylogenetic analysis showed that the elements are related to geographic distribution and/or species, but were found evidence of genetic flux between strains belonging to different countries from the Americas and horizontal gene transfer

between *R. syzygii* and *R. pseudosolanacearum*. In addition, the majority of GIs and ICEs identified are preferentially associated with *R. pseudosolanacearum* strains. ICEs and GIs carry accessory genes with potential impact on RSSC fitness and pathogenicity some coding for stress response and candidate virulence genes, such as several type III effector genes which may enable the bacteria to evolve rapidly and infect a wide range of hosts. To summarize, our results provide novel insight into the RSSC diversified adaptation, opening new paths to a better understanding of how these elements may affect the fitness and pathogenicity traits of this important plant pathogen.

Keywords: Genome evolution, microbial genomics, mobile DNA, pathogenicity islands, plant pathogen

Introduction

The soil-borne bacteria *R. solanacearum* is one of the most devastating phytopathogens worldwide, responsible for bacterial wilt disease in more than 250 plant species belonging to 54 different botanical families [1], [2]. Strains of *R. solanacearum* form a heterogenous group of species that are divided into a number of phylotypes corresponding to their geographic origin: Asia and Africa (phylotype I), America (II), Africa (III) and Indonesia (IV) [3] [4]. This complex is commonly known as the *R. solanacearum* species complex (RSSC), which has been assigned to *R. solanacearum*, *R. syzygii* and the blood disease bacterium (BDB) (Prior and Fegan 2005). Recently, a taxonomic and nomenclatural update has proposed that the RSSC should encompass three distinct species: *R. pseudosolanacearum* (formerly phylotypes I and III), *R. solanacearum* (IIA and IIB) and *R. syzygii* (formerly phylotype IV and BDB) (Safni et al. 2014, Prior et al. 2016).

R. solanacearum strains have a characteristic bipartite genome structure (chromosome and megaplasmid). In the genome of *R. pseudosolanacearum* GMI1000, a type strain, these replicons have a mosaic structure containing numerous mobile genetic elements (MGEs) signaling the potential for evolution through horizontal gene transfer (HGT) (Salanoubat et al. 2002). MGEs are efficient vectors for HGT that can contribute to bacterial genome evolution. In addition to MGEs encoding enzymes that mediate their inter or intracellular movement, these elements commonly encode accessory proteins that can confer important properties on the host cell (Frost et al. 2005).

Recently, as a result of the increase in bacterial genome sequences, the impact of MGEs in shaping the structure of bacterial diversification through the acquisition of novel genetic traits has been demonstrated (Croucher et al. 2014; Klemm and Dougan 2016). Furthermore, genomic islands (GIs) and integrative and conjugative elements (ICEs) are recognized as important MGEs that contribute to bacterial genome adaptation, plasticity, and evolution (Frost et al. 2005).

Genomic Islands are genomic regions horizontally acquired, recognized by CG-content and codon usage bias that usually differ from the rest of the genome, often inserted at tRNA genes which are flanked by 16–20-bp direct repeats (DR) (Juhas et al. 2009). GIs harbor components of mobile genetic elements which may have been implicated in mobilizing DNA (Gal-Mor and Finlay 2006). Furthermore, these regions often carry genes conferring a selective advantage to the host bacterium, which, depending on their gene functions, may be referred to as pathogenicity, symbiosis, metabolic, fitness or resistance islands (Hacker & Carniel 2001; Juhas et al. 2009). ICEs are chromosomal, self-transmissible mobile elements which can also be passively propagated during chromosomal replication, cell division or horizontal transfer via conjugation (Johnson and Grossman 2015). This genetic element has a modular structure such as a recombination module, a conjugation module, and a regulation module, which together

control and ensure the excision and transfer of the ICEs. The acquisition of ICEs and GIS may be attributed to a great source of genetic variation and ecological adaptation to certain plant pathogens (McCann et al. 2013; De Maayer et al. 2015). Therefore, in this study, to obtain insight into the genetic basis of the RSSC diversified adaptation, we conducted genome mining and comparative genomic analysis based on its repertoire of GIs and ICEs.

Materials and Methods

Data collection

106 RSSC and 15 *Ralstonia* spp. genome sequences were retrivied from the NCBI (https://www.ncbi.nlm.nih.gov/genome) (Supplementary material 1), in July 2018. We used the classification of replicons as chromosomes and megaplasmid as provided in the GenBank files, to perform our analyses.

Detection of GIs

GIs were predicted using the interface Island Viewer 4 (http://pathogenomics.sfu.ca/islandviewer) (Bertelli et al. 2017) executed with default parameters using the following GI prediction methods: IslandPick, IslandPath-DIMOB, SIGI-HMM and Islander. Accurate analysis of GIs, and only genomic regions detected by the three methods (IslandPick, IslandPath-DIMOB, SIGI-HMM) were considered. These three methods use: a comparative genomic prediction method to develop stringent data sets of GIs and non-GIs; abnormal sequence composition and presence of genes functionally related to mobile elements; and measurements of dinucleotide bias, respectively (Brinkman et al. 2003; Waack et al. 2006; Langille et al. 2008). The GIs were then manually filtered to avoid false positive and negative data sets according to the following criteria: (i) presence of mobility genes (*e.g.*, integrases and transposases); (ii) proximal structural RNA (tRNA and tmRNA), and; (iii) atypical guanine and

cytosine content. Predicted GIs were extracted and were systematically annotated by RAST server (http://rast.nmpdr.org/) (Aziz et al. 2008). Direct repeats and insertion sequences were manually identified in annotated sequences using the Geneious® 11.1.5 program (Biomatters Lts). Next, different genomic sequence signatures (*e.g.*, G+C content, GC-skew, codon usage) were also measured using Geneious® plugins. For the latter, CDS were categorized in clusters of orthologous groups (COGs) with COG v1.0 against NCBI preformatted CDD (conserved domains database) with an E-value cutoff of 1e-05. Identified GIs were named as follows: GI + Number of the GI + Strain.

Detection and delimitation of ICEs

We used two strategies to identify ICEs in the RSSC genomes. First, standard BLASTn (Altschul et al. 1990) searches against the ICEberg database version 1.0 (http://db-mml.sjtu.edu.cn/ICEberg/) (Bi et al. 2012), only matches with the E-value $\leq 10^{-5}$ and sequence coverage $\geq$ 85% were retained. Second, we found new ICEs using the presence of conserved conjugation and DNA processing genes (*e.g.*, trbE, traG, and trbL) in annotated genomes deposited in the NCBI database and compared them to closely related ICEs to identify conserved features indicative of conjugative elements. In addition, we also analyzed core genes flanking these loci to provide the upper bounds for the limits of the ICE. Thus, similar to Cury et al. (2017), we overlaid the presence of genes of the conjugation system, on GC content to delimit the ICEs. An element was considered as conjugative when it contained the following components of the conjugative system: a relaxosome, a coupling protein and a type IV secretion system (Guglielmini et al. 2014). Predicted ICEs were inspected for DRs that define the boundaries of the element. The complete nucleotide sequences of the ICEs, in GenBank format of corresponding records, were imported into Geneious® to help delimit genomic regions flanking the elements. We used the annotations of the GenBank files for our predicted ICEs and

categorized the CDS in clusters of orthologous groups (COGs) with COG v1.0 against NCBI preformatted CDD (conserved domains database) with an E-value cutoff of 1e-05. Identified ICEs were denominated as the following: ICE + Species + strain.

Virulence – associated genes in GIs and ICEs

We performed a BLASTp (Using parameters e-value $\leq 10^{-5}$ and amino acid identity > 30%) on the following database to analyze GIs and ICEs carrying virulence cargo genes: Virulence Factors Database (VFDB, http://www.mgc.ac.cn/VFs/) (Chen et al. 2005); Pathogen–Host Interactions database (PHI-base, www.phi-base.org) (Urban et al. 2017); The Pathogenicity Island Database (PAIDB, http://www.paidb.re.kr) (Yoon et al. 2015); Type III Secretion System Database (T3SEdb, http://effectors.bic.nus.edu.sg/T3SEdb) (Tay et al. 2010).

Element Comparisons

In order to infer phylogenetic relationships between ICEs and GIs identified in our study, we performed a whole sequence alignment on each type of element using ClustalW (Larkin et al. 2007). Multiple genome alignments were performed using the Mauve software (version 2.3.1) (Darling et al. 2004). The phylogenetic tree was constructed in MEGA X using Maximum Likelihood (1,000 bootstrap replicates) (Kumar et al. 2018). The predicted ICEs were analyzed for gene content and extracted from the genome to construct multiple sequence alignments using the Mauve software algorithm (Darling et al. 2004).

Results

The majority of GIs and ICEs are preferentially associated with *R. pseudosolanacearum* strains

In order to evaluate the evolutionary impact of GIs and ICEs on *Ralstonia* strains, we first characterized the composition and distribution of these two elements in the RSSC. From 121 genomes analyzed, GIs and ICEs were found only in the 37 complete genomes. We therefore manually created a curated dataset of 40 GIs and 12 ICEs from these 37 genomes. The size of GIs ranged from 27.5 Kb to 192 Kb and ICEs ranged from 41 Kb to 83Kb. The average size of GIs is slightly larger than that of ICEs (65.5 kb vs. 54.4 kb) (Fig. 1b, Table 1 and Table 2). Both GIs and ICEs were found to have GC lower than their host genomes (66% CG content), varying from 56.1 to 65.2, and 60.2 to 65.4, respectively (Fig. 1c). These genomic elements were identified on the chromosome of 58% (n=37) *Ralstonia* strains. None of the ICEs or GIs were identified in the megaplasmid sequences according to our criteria. Overall, GIs and ICEs are distributed throughout the complex, being the majority of GIs and ICEs are preferentially associated with R. *pseudosolanacearum* strains (fig. 1d). In addition, two GIs and three ICEs were found in tree non-plant pathogenic *Ralstonia* spp. (*R. pickettii*, *R. mannitolilytica*, and *R. insidiosa*). *R. solanacearum* encompasses seven different GIs, being GI1RS488 shared by two South American strains (RS 489 and UY031). For *R. pseudosolanacearum* eight GIs and six ICEs shared by mostly Asian strains (T98, SL3175, T60, SL3755, T42, SL3730, SL2729) and American strains (GMI1000, RS476 and CMRs218). *R. syzygii* encompasses five different GIs and one ICE. Interestingly, the *R. pseudosolanacearum* strain KACC10722 was found to share one GI and ICE among the *R. syzygii* strains. The number of predicted ORFs in ICEs ranges from 44 to 78, and ranged from 24 to 161 in GIs. Details of the 41 GIs and 12 ICEs are listed in Tables 1 and 2. In summary, these 40 GIs and 12 ICEs were identified as representing the largest collection of genomic regions identified for the complex to date. In subsequent systematic analyses, we looked at the evolutionary relationships between GIs and ICEs throughout the RSSC.

An ancestral acquisition of GIs and ICEs throughout RSSC

We constructed an alignment between GIs and ICEs to identify topologically and functionally conserved network sequences. The analysis revealed seven small clusters with sequence identity above 50% (Fig. S1). We used the scores of pairwise alignment to create an undirected network graph. As shown in Fig. 2, we found two main conserved network regions representing GIs and ICEs groups, which may reflect the functional and evolutionary relationships between the sequences. Through analysis of the network graph, we noticed a number of aspects that are noteworthy. The thickness of the edge is proportional to the score; therefore, we observed that certain elements are linked to each other. We also found GIs and ICEs sharing pairwise similarities between them. This finding represents associated genes that are related to functionally similar groups. In addition, we constructed phylogenetic trees to look closer topologically at these linked elements. The phylogeny shows several clusters containing highly related GIs (Fig. 3a) and ICEs (Fig. 3b), which were identified to compose the same species or cluster according to their geographic distribution. Members of the RSSC have formed a coherent group, according to geographical distribution and species, appearing as individual clusters. Notably, there is a clear gene flux between close strains and HGT between species. We found genetic flow between *R. pseudosolanacearum* strains from Brazil (RS476 and CMRs218) and French Guyana (GMI1000); and GI from *R. solanacearum* strains shared between Brazil (RS488 and RS489) and Uruguay (UY081). Surprisingly, we also noted HGT between *R. pseudosolanacearum* strain (KACC10722) and *R. syzygii* (SL2064, T51 and T95) from Korea. We also observed GIs between *R. pseudosolanacearum* and *R. solanacearum* Brazilian strains among *R. pseudosolanacearum* from China.

We combined the GIs into three groups, corresponding to the species clusters found in the phylogenetic tree to analyze their proportion on the *Ralstonia* chromosome. This analysis revealed that GI1 found in *R. pseudosolanacearum* strains occupied from 4 to 5% of its

chromosome, followed by GI2 found in the *R. solanacearum* and GI3 found in the *R. syzygii* (Fig. 4). The phylogeny demonstrates an ancestral acquisition of GIs and ICEs among RSSC strains, since these elements are diffused in the complex with a different geographic distribution.

Identification of several novel ICEs among RSSC

Our analysis revealed five novel ICEs identified in eight *Ralstonia* strains (Fig. 5). ICE, named here as ICEPsCRMRs218 and ICEPsRS476, were found on the chromosome of two *R. pseudosolanacearum* strains CRMRs218 and RS476, respectively. These ICEs are similar to the ICE observed in *R. pseudosolanacearum* GMI1000, which is related to the 55-kb transposon Tn4371 from *R. pickettii*, except for replaced biphenyl resistance (Salanoubat et al. 2002; Ryan et al. 2009). Similarly, we found three novel ICEs also in *R. pseudosolanacearum* strains (FQY_4, YC40-M, FJAT-91, and HA4I), FQY_4 e YC40-M sharing the same ICE. ICE from HA4-1 was observed in two *R. syzygii* strains (T98 and SL3175). In general, ICEs displayed a mosaic structure characterized by differing the gene cassettes (Fig. 5). The main putative cargo genes code for protein with hydrolase activity, carried by almost all ICEs, including limonene-1.2-epoxide hydrolase, amidohydrolase and several peptidases, which are involved in the degradation of plant cell walls. We also found genes predicted to encode proteins involved in stress response, in particular, oxidative stress, which may be involved in the detoxification of the active oxygen species produced by infected plants. An interesting Cluster of Orthologous Groups (COG) related to aromatic compound metabolism was found in four Asian strains FQY_4, YC40-M, SL3175 and T98. These genes may be important to the bacteria metabolizing this compound and, thus, its adaptability to the environment in the soil (Fuchs et al. 2011). Details for accessory genes found in ICEs are listed in Table 3.

Despite a mosaic structure, we also observed collinear syntenic blocks, representing the core modular structure, which displayed a high degree of similarity between these ICEs (Fig. 6). These modular genes code for mobilization proteins (*e.g.,* VirD4), mating pair formation (Mpf) (e.g., TrbB, TrbC, TrbD, TrbE, TrbF, TrbG, TrbI), secretion system proteins (e.g. VirD2), regulatory proteins (e.g., RepA), maintenance protein (e.g., TA system), integrase/ excisionase and others (e.g., ATPases, phage-related proteins) (Fig. 5, Supplementary material 4). A point to note, these five ICEs are only related to *Ralstonia* strains that share a low degree of identity (<50%) with other species (data not shown), which might possibly indicate an exclusive association.

The functional repertoires of genes of GIs and ICEs reveal the potential impact on fitness and pathogenicity

The presence of GIs and ICEs in the bacteria host can confer selective advantages because these genomic regions often encode additional functions beyond what is essential for bacterial growth (Carraro et al. 2017; Delavat et al. 2017). We then investigated whether fitness and pathogenicity genes can be linked to these two elements. We searched for the repertoires of genes in our dataset. In total, we located 2,330 putative ORFs in island regions and 832 putative ORFs in ICEs. Thus, we were able to calculate the relative ratio of the different functions between the two types of elements (Fig. 7a). Next, we categorized these genes into two major functional groups.

GIs and ICEs encoding ecological fitness genes. As shown in Fig. 7a, the most prevalent ORFs correspond to poorly characterized genes, which, for the most part represents hypothetical protein. The high percentage of hypothetical protein genes has been commonly related to GIs and ICEs (Hsiao et al. 2005; Che et al. 2014; Cury et al. 2017). A second COG

we identified is phages, prophages, transposable elements, and plasmids, which are often involved in the integration, recombination, or excision of both elements. Other major clusters found were based on protein metabolism, which was assigned to a number of set subgroups, including transferase activity, transaminase activity, catalytic activity, and ATP binding. In addition, we found in the GIs genes coding for protein involved in cofactors, vitamins, prosthetic groups, pigments. A set of ORFs involved in the Iron acquisition and metabolism, sulfur metabolism, cell wall and capsule, and cell division and cell cycle. Details of the cargo genes are listed in Supplementary material 4 and Supplementary material 5.

GIs and ICEs encoding variety of Virulence Factors. Figure 7A also shows a higher frequency of virulence, disease, and defense category, which represent 9.4% and 1 % of the ORFS found in GIs and ICEs, respectively. In subsequent analysis, we grouped the GIs virulence ORFs into eight classes (VF1 to VF8) corresponding, respectively, to Adhesion/Surface proteins, Hydrolytic enzymes/Host cell wall degradation, Plant hormones & signaling molecules, Potential Type III secretion-dependent effectors, Resistance to oxidative stress, Secretion system, Toxins, and Type III secretion system/secreted effectors, according to Salanoubat et al. (2002). In Fig. 7b, we demonstrated the distribution of each virulence class in the 39 GIs identified. The VF2 (Hydrolytic enzymes/Host cell wall degradation) and VF6 (Secretion system) were found to be the most prevalent classes throughout the GIs, 51% (n=110) and 23% (n=23), respectively. Another class that we highlight is VF8 (Type III secretion system and secreted effectors) found in 8.4% (n=18) of GIs. This class, essential for pathogenicity, includes Type III effector protein code for Ript and Skwp 4, popP2, and popP1 (Fig. 8b, Supplementary material 5). We found that the most frequent virulent factor in ICEs corresponds to the Hydrolytic enzymes/Host cell wall degradation class (22/31).

Discussion

R. solanacearum is a useful model for exploring the evolution patterns in closely connected strains due to the formation of emerging ecotypes (highly host-adapted strains) adapting to local environmental conditions (Genin and Denny 2012; Liu et al. 2017). Herein, we provide genomic insights into this diversified adaptation based on the repertoire of GIs and ICEs in a large set of strains within the RSSC. We conducted genome mining and comparative genomic analysis in 121 sequenced *Ralstonia* strains. Our analyses identified a total of 52 integrative mobile regions in 37 complete genomes throughout the species complex. The number of these two elements can be even higher due to the fact that most *Ralstonia* genomes are in draft assemblies (split into multiple contigs), and it became a daunting task to identify GIs and ICEs (Ricker et al. 2012). In addition, GIs and ICEs are often known to be absent in several closely related genomes (Langille et al. 2010). We compared the composition and distribution of the GIs and ICEs and evaluated their evolutionary impact. These elements are characterized by their large size (>10 Kb) and a different G+C content compared with the rest of the genome (Langille et al. 2010; Bertelli et al. 2018). We found that GIs and ICEs constitute a large fraction of the *Ralstonia* genomes, varying in size and GC content. Especially in the case of the ICEs, C+G content tends to become more similar to the compositional signature of their host, suggesting that ICEs are undergoing a domestication process (Lawrence and Ochman 1997; Marri and Golding 2008). Certain GIs and ICEs are hundreds of kilobases long. However, we noticed no significant effect on the chromosome structure nor on genome organization.

None of the ICEs or GIs was found in the megaplasmid sequences following our criteria (see Methods: Detection of GIs). This may be explained by the number of tRNA genes found on the chromosome and megaplasmid, once the most predominant ICEs and GIs were preferentially associated with tRNA and tmRNA genes. We found an average of 55 tRNA in the chromosome and three tRNA genes in the megaplasmid (data not shown). Therefore, GIs

and ICEs associated with tRNA and tmRNA appear to be good indicators for site-specific integration and HGT in RSSC (Hacker et al. 1997).

The majority of GIs and ICEs identified are present in *R. pseudosolanacearum* followed by the GIs from *R. solanacearum* strains. These observations could be due to the fact that both *R. pseudosolanacearum*, as well as *R. solanacearum*, cover a significant number of genomes analyzed in this study. We also revealed that GIs from these two species occupied a large fraction on their chromosomes. In addition, these two species are recognized to have been arranged in clonal complexes worldwide, which may also contribute to this preference (Castillo and Greenberg 2007). Consequently, Safni et al. (2014) suggest further work to better understand and propose the designation of subspecies for one or both species, especially for *R. pseudosolanacearum*, where a clear geographical division exists.

We have also demonstrated that these identified elements constitute phylogenetically related MGEs that have adapted to the same species and/or geographic distribution. This finding also showed evidence that isolates from the RSSC have recently fallen under genetic flux in different countries in the Americas (Brazil, Uruguay and Peru). To our knowledge, this is the first report of genetic flux between *Ralstonia* strains from different countries. Thus, our finding underscores the necessity of adopting strict quarantine measures by countries. In addition, the HGT between *R. pseudosolanacearum* and *R. syzygii* from Korea is also clear. Gene flow between species is an important mechanism for genetic diversity between populations. Furthermore, GIs and ICEs enable this dynamic lateral gene flow, which allows for the bacteria to rapidly acquire new traits (Wozniak and Waldor 2010).

Interestingly, our analysis suggests that GIs and ICEs may remain stably incorporated in *Ralstonia* genomes, and that could play an important role in the evolution of the bacterial genome. It has been suggested an ancestral acquisition of other elements in distinct RSSC, *e.g.*, CRISPR-Cas system (Xavier et al. 2019). This finding is supported by Castillo and Greenberg

(2007) who showed remarkable accumulation of fixed polymorphisms in the *R. solanacearum* phylotypes and also indicated that *R. solanacearum* is an ancient pathogen.

In addition to the ICEs already reported, *e.g.*, ICE from the *R. pseudosolanacearum* strain GMI1000 (Salanoubat et al. 2002), Tn4371 from *R. pickettii* and *R. insidiosa* (Ryan et al. 2009), we identified five novel ICEs in the *Ralstonia* species. In addition to an unreported ICE observed in two *R. pseudosolanacearum* closely related to GMI1000, two novel ICEs were found in the species; and one novel ICE in *R. syzygii* strains. Despite a mosaic structure shown by these novel ICEs, we have demonstrated a collinearity identity between them. In general, this identity represents the core modular structure related functions, including conjugation, recombination, and regulation (Johnson and Grossman 2015; Cury et al. 2017).

GIs and ICEs were analyzed and further searches carried out for cargo genes. A repertoire of genes was found in GIs and ICEs with potential impact in RSSC fitness and pathogenicity. We found genes of putative function related to bacteria fitness. Among these, we highlight the cluster of genes coding for protein metabolism, cofactors, vitamins and pigment, genes involved in iron acquisition and sulfur metabolism. These subsets of functional clusters may help the RSSC to thrive in the plant environment, contributing to its diversified adaptation (Levy et al. 2018).

Several putative genes coding for virulence factors were found. Next we represented in eight classes corresponding to adhesion/surface proteins, hydrolytic enzymes/host cell wall degradation, plant hormones & signaling molecules, potential type iii secretion-dependent effectors, resistance to oxidative stress, secretion system, toxins and type iii secretion system and secreted effectors. These eight classes were represented in the majority of GIs identified, which shows their widespread dissemination throughout these genomic elements. We also found 8% of GIs encoded type III secretion system and secreted effectors, which are recognized as essential for RSSC pathogenicity (Genin and Denny 2012). Type III secretion system is

responsible for the synthesis and also injection of a variety of effector proteins into the plant cells during the infection stage, which can cause disease in a susceptible host or a hypersensitive response in the resistant hosts (Aldon et al. 2000; Valls et al. 2006). In *Ralstonia* this system is tightly controlled by the central regulator HrpG, which transfers the activation signal through a cascade of transcriptional regulators that activate and repress the expression of many effectors genes (Brito et al. 2002; Coll and Valls 2013).

In addition to type III secretion system and secreted effectors, *R. solanacearum* pathogenicity is also determined by a number of other virulence factors, such as plant cell wall–degrading enzymes. These genes are secreted across the outer membrane by the type II secretion system (T2SS), recognized as essential for the colonization phase (Liu et al. 2005). Herein, we identified a collection of GIs and ICEs carrying important hydrolytic enzymes that are secreted by this system. During the infection phase, *R. solanacearum* face a variety of reactive oxygen species (ROS) made by plants, the expression of genes that confer resistance to oxidative stress allow the bacteria to tolerate this oxidative environment. Not surprisingly, mutants lacking genes that confer resistance to oxidative stress, *e.g*., peroxidase and regulator of oxidative stress genes, significantly reduce virulence (Colburn-Clifford et al. 2010). Also, evidence of positively regulation of phytohormones production, e.g., ethylene and auxin, during interaction with plants might contribute to the virulence of *R. solanacearum*, but its role is still poorly understood (Valls et al. 2006). It should be noted that our results have shown a great variety of *R. solanacearum* virulence factors such as genes cargo in GIs and ICEs, which might provide a selective phenotype for this important pathogen. In summary, these results demonstrate a series of putative ORFs that may be involved in the RSSC's diversified adaptation to environmental conditions and genes for their pathogenicity.

Funding information

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001; Fundação Arthur Bernardes (Furnabe; Minas Gerais, Brasil) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; Brasília, Brasil).

Compliance with ethical standards

This study does not contain any experiment with human participants or animals performed by any of the authors.

Conflict of Interest

The authors have no conflicts of interest to disclose.

Publisher's note

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

References

Aldon D, Brito B, Boucher C, Genin S (2000) A bacterial sensor of plant cell contact controls the transcriptional induction of *Ralstonia solanacearum* pathogenicity genes. EMBO J 19(10):2304–2314. https://doi.org/10.1093/emboj/19.10.2304

Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS (2016) PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44: W16–W21. https://doi.org/10.1093/nar/gkw387

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards, RA, Zagnitko O (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75

Bertelli C, Laird MR, Williams KP, Simon F, Lau BY, Hoad G, Brinkman FSL (2017) IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res 45(W1): W30–W35. http://dx.doi.org/10.1093/nar/gkx343

Bertelli C, Tilley KE, Brinkman FSL (2018) Microbial genomic island discovery, visualization and analysis. Brief Bioinform 42:1-14. https://doi.org/10.1093/ bib/bby042

Bi D, Xu Z, Harrison EM, Tai C, Wei Y, He X, Ou HY (2012) ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. Nucleic Acids Res 40: D621–D626. https://doi.org/10.1093/nar/gkr846

Brinkman FSL, Wan I, Jones SJ, Hsiao W (2003) IslandPath: aiding detection of genomic islands in prokaryotes. Bioinformatics 19(3):418–420. https://doi.org/10.1093/bioinformatics/btg004

Brito B, Aldon D, Barberis P, Boucher C, Genin S (2002) A signal transfer system through three compartments transduces the plant cell contact-dependent signal controlling

Ralstonia solanacearum hrp genes. Mol Plant-Microbe Interact 15(2):109–119. https://doi.org/10.1094/MPMI.2002.15.2.109

Burrus V, Waldor MK (2004) Shaping bacterial genomes with integrative and conjugative elements. Res Microbiol 155(5):376–386. https://doi.org/10.1016/J.RESMIC.2004.01.012

Carraro N, Rivard N, Burrus V, Ceccarelli D (2017) Mobilizable genomic islands, different strategies for the dissemination of multidrug resistance and other adaptive traits. Mob Genet Elements 7(2):1–6. https://doi.org/10.1080/ 2159256X.2017.1304193

Castillo JA, Greenberg JT (2007) Evolutionary dynamics of *Ralstonia solanacearum.* Appl Environ Microbiol 73(4):1225–1238. https://doi.org/10.1128/AEM.01253-06

Che D, Hasan M, Chen B (2014) Identifying pathogenicity islands in bacterial pathogenomics using computational approaches. Pathogens 3(1):36–56. https://doi.org/10.3390/pathogens3010036

Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q (2005) VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 33: D325–D328. https://doi.org/10.1093/nar/gki008

Colburn-Clifford JM, Scherf JM, Allen C (2010) *Ralstonia solanacearum* Dps contributes to oxidative stress tolerance and to colonization of and virulence on tomato plants. Appl Environ Microbiol 76(22):7392-7399. https://doi.org/ 10.1128/AEM.01742-10

Coll NS, Valls M (2013) Current knowledge on the *Ralstonia solanacearum* type III secretion system. Microb Biotechnol 6(6):614–620. https://doi.org/10.1111/1751-7915.12056

Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP (2014) Diversification of bacterial genome content through distinct mechanisms over different timescales. Nat Commu 5:5471. https://doi.org/10.1038/ncomms6471

Cury J, Touchon M, Rocha EPC (2017) Integrative and conjugative elements and their hosts: composition, distribution and organization. Nucleic Acids Res 45(15):8943–8956. https://doi.org/10.1093/nar/gkx607

Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14(7):1394–1403. https://doi.org/10.1101/gr.2289704

De Maayer P, Chan WY, Martin DAJ, Blom J, Venter SN, Duffy B, Coutinho TA (2015) Integrative conjugative elements of the ICEPan family play a potential role in *Pantoea ananatis* ecological diversification and antibiosis. Front Microbiol 6:576. https://www.frontiersin.org/article/10.3389/fmicb.2015.00576

Delavat F, Carraro N, Pradervand N, Van der Meer JR, Miyazaki R (2017) The hidden life of integrative and conjugative elements. FEMS Microbiol Rev 41(4):512–537. https://doi.org/10.1093/femsre/fux008

Dy RL, Przybilski R, Semeijn K, Salmond GPC, Fineran PC (2014) A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. Nucleic Acids Res 42(7):4590–4605. https://doi.org/10.1093/ nar/gkt1419

Fegan M, Prior P (2005) How Complex is the "*Ralstonia Solanacearum* Species Complex. In Bact Wilt Dis Ralstonia Solanacearum Species Complex.

Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3:722. http://dx.doi.org/10.1038/nrmicro1235

Fuchs G, Boll M, Heider J (2011) Microbial degradation of aromatic compounds — from one strategy to four. Nat Rev Microbiol 9:803. https://doi.org/10.1038/nrmicro2652

Gal-Mor O, Finlay BB (2006) Pathogenicity islands: a molecular toolbox for bacterial virulence. Cell Microbiol 8(11):1707–1719. https://doi.org/10.1111/j.1462-5822.2006.00794.x

Genin S, Denny TP (2012) Pathogenomics of the *Ralstonia solanacearum* Species Complex. Annu Rev Phytopathol 50(1):67–89. https://doi.org/10.1146/annurev-phyto-081211-173000

Guglielmini J, Néron B, Abby SS, Garcillán-Barcia MP, de la Cruz F, Rocha EPC (2014) Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. Nucleic Acids Res 42(4):5715–5727. https://doi.org/10.1093/nar/gku194

Hacker J, Blum-Oehler G, Mühldorfer I, Tschäpe H (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol Microbiol 23(6):1089–1097. https://doi.org/10.1046/j.1365-2958.1997.3101672.x

Hacker J, Carniel E (2001) Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. EMBO Rep 2 https://doi.org/10.1093/embo-reports/kve097

Hayward AC (1991) Biology and epidemiology of bacterial wilt caused by *Pseudomonas solanacearum*. Annu Rev Phylopathol 2:65–87. https://www.annualreviews.org/doi/pdf/10.1146/annurev.py.29.090191.000433

Hsiao WWL, Ung K, Aeschliman D, Bryan J, Finlay BB, Brinkman FSL (2005) Evidence of a large novel gene pool associated with prokaryotic genomic islands. PLoS Genet 1(5): e62. https://doi.org/10.1371/journal.pgen.0010062

Johnson CM, Grossman AD (2015) Integrative and Conjugative Elements (ICEs): What they do and how they work. Annu Rev Genet 49(1):577–601. https://doi.org/10.1146/annurev-genet-112414-055018

Juhas M, Van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW (2009) Genomic islands: tools of bacterial horizontal gene transfer and evolution. Fems Microbiol Rev 33(2):376–393. https://doi.org/10.1111/j.1574-6976.2008.00136.x

Klemm E, Dougan G (2016) Advances in understanding bacterial pathogenesis gained from whole-genome sequencing and phylogenetics. Cell Host Microbe 19(5):599–610. https://doi.org/10.1016/J.CHOM.2016.04.015

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution (35). https://doi.org/10.1093/molbev/msy096

Langille MGI, Hsiao WWL, Brinkman FSL (2008) Evaluation of genomic island predictors using a comparative genomics approach. BMC Bioinformatics 9(1):329. https://doi.org/10.1186/1471-2105-9-329

Langille MGI, Hsiao WWL, Brinkman FSL (2010) Detecting genomic islands using bioinformatics approaches. Nat Rev Microbiol 8:373. https://doi.org/10.1038/nrmicro2350

Larkin MA, Wilm A, Higgins DG, Valentin F, Blackshields G, McWilliam, H, Gibson TJ (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Lawrence J, Ochman H (1997) Amelioration of bacterial genomes: Rates of change and exchange. J Mol Evol (44). https://doi.org/10.1007/PL00006158

Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Dangl JL (2018) Genomic features of bacterial adaptation to plants. Nat Genet 50(1):138–150. https://doi.org/10.1038/s41588-017-0012-9

Liu B, Pop M (2009) ARDB—Antibiotic Resistance Genes Database. Nucleic Acids Res 37: D443–D447. https://doi.org/10.1093/nar/gkn656

Liu H, Zhang S, Schell MA, Denny TP (2005) Pyramiding unmarked deletions in *ralstonia solanacearum* shows that secreted proteins in addition to plant cell-wall-degrading

enzymes contribute to virulence. Mol Plant-Microbe Interact 18(12):1296–1305. https://doi.org/10.1094/MPMI-18-1296

Liu Y, Tang Y, Qin X, Yang L, Jiang G, Li S, Ding W (2017) Genome sequencing of *ralstonia solanacearum* cqps-1, a phylotype I strain collected from a highland area with continuous cropping of tobacco. Front Microbiol 8:974. https://doi.org/10.3389/fmicb.2017.00974

Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, Foster GD (2012) Top 10 plant pathogenic bacteria in molecular plant pathology. Mol Plant Pathol 13(16):614–629. https://doi.org/10.1111/j.1364-3703.2012.00804.x

Marri PR, Golding GB (2008) Gene amelioration demonstrated: the journey of nascent genes in bacteria. Genome 51(2):164–168. https://doi.org/10.1139/G07-105

McCann HC, Rikkerink EHA, Bertels F, Fiers M, Lu A, Rees-George J, Templeton MD (2013) Genomic analysis of the kiwifruit pathogen *pseudomonas syringae* pv. *actinidiae* provides insight into the origins of an emergent plant disease. PLOS Pathog 9(7): e1003503. https://doi.org/10.1371/journal.ppat.1003503

MF, Alm EJ, Hanage WP (2013) Horizontal gene transfer and the evolution of bacterial and archaeal population structure. Trends Genet 29(3):170–175. https://doi.org/10.1016/j.tig.2012.12.006

Prior P, Ailloud F, Dalsing BL, Remenant B, Sanchez B, Allen C (2016) Genomic and proteomic evidence supporting the division of the plant pathogen *Ralstonia solanacearum* into three species. BMC Genomic 17:90. https://doi.org/10.1186/ s12864-016-2413-z

Ravenhall M, Skunca N, Lassalle F, Dessimoz C (2015) Inferring horizontal gene transfer. PLoS Comput Biol 11(5): e1004095–e1004095. https://doi.org/10.1371/journal.pcbi.1004095

Ricker N, Qian H, Fulthorpe RR (2012) The limitations of draft assemblies for understanding prokaryotic adaptation and evolution. Genomics 100(3):167–175. https://doi.org/10.1016/J.YGENO.2012.06.009

Ryan MP, Pembroke JT, Adley CC (2009) Novel Tn4371-ICE like element in *Ralstonia pickettii* and genome mining for comparative elements. BMC Microbiol 9:242. https://doi.org/10.1186/1471-2180-9-242

Safni I, Cleenwerck I, Paul De Vos, 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. Int J Syst Evol Microbiol 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, Boucher CA (2002) Genome sequence of the plant pathogen *Ralstonia solanacearum*. Nature 415:497. http://dx.doi.org/10.1038/415497a

Shao Y, Harrison EM, Bi D, Tai C, He X, Ou HY, Deng Z (2011) TADB: a web-based resource for Type 2 toxin-antitoxin loci in bacteria and archaea. Nucleic Acids Res 39: D606–D611. https://doi.org/10.1093/nar/gkq908

Sim SH, Yu Y, Lin CH, Karuturi RKM, Wuthiekanun V, Tuanyok A, Tan P (2008) The core and accessory genomes of *Burkholderia pseudomallei*: Implications for human melioidosis. PLoS Pathog 4(10): e1000178. https://doi.org/10.1371/journal.ppat.1000178

Stevens P, Van Elsas JD (2010) A putative genomic island, PGI-1, in *Ralstonia solanacearum* biovar 2 revealed by subtractive hybridization. Antonie Van Leeuwenhoek 98(3):359–377. https://doi.org/10.1007/s10482-010-9450-4

Sundin GW (2007) Genomic insights into the contribution of phytopathogenic bacterial plasmids to the evolutionary history of their hosts. Annu Rev Phytopathol 45(1): 129–151. https://doi.org/10.1146/annurev.phyto.45.062806.094317

Tay DMM, Govindarajan KR, Khan AM, Ong TYR, Samad HM, Soh WW, Tan TW (2010) T3SEdb: data warehousing of virulence effectors secreted by the bacterial Type III Secretion System. BMC Bioinformatics 7(7): S4–S4. https://doi. org/10.1186/1471-2105-11-S7-S4

Urban M, Cuzick A, Rutherford K, Irvine A, Pedro H, Pant R, Hammond-Kosack KE (2017) PHI-base: a new interface and further additions for the multi-species pathogen–host interactions database. Nucleic Acids Res 45: D604–D610. https://doi.org/10.1093/nar/gkw1089

Valls M, Genin S, Boucher C (2006) Integrated regulation of the type III secretion system and other virulence determinants in *Ralstonia solanacearum*. PLoS Pathog 2(8): e82–e82. https://doi.org/10.1371/journal.ppat.0020082

Van Melderen L (2010) Toxin–antitoxin systems: why so many, what for? Curr Opin Microbiol, 13(6):781–785. https://doi.org/10.1016/j.mib.2010.10.006

Van Melderen L, Saavedra De Bast M (2009) Bacterial Toxin–Antitoxin systems: more than selfish entities? PLOS Genet 5(3): e1000437. https://doi.org/10.1371/journal.pgen.1000437

Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, Merkl R (2006) Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC Bioinformatics 7:142. https://doi.org/10.1186/1471-2105-7-142

Wicker E, Lefeuvre P, de Cambiaire JC, Lemaire C, Poussier S, Prior P (2012) Contrasting recombination patterns and demographic histories of the plant pathogen *Ralstonia*

solanacearum inferred from MLSA. ISME J 6(5):961–974. https://doi.org/10.1038/ismej.2011.160

Wozniak RAF, Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol 8:552. http://dx.doi.org/10.1038/nrmicro2382

Xavier ADS, Rousseau GM, Almeida JCF, Rezende RR, Melo AG, Moineau S, Alfenas-Zerbini P (2018) Characterization of CRISPR-Cas systems in the *Ralstonia solanacearum* species complex. Mol Plant Pathol 20(2):223–239. https://doi. org/10.1111/mpp.12750

Yoon SH, Park YK, Kim JF (2015) PAIDB v2.0: exploration and analysis of pathogenicity and resistance islands. Nucleic Acids Res 43: D624–D630. https://doi.org/10.1093/nar/gku985

Table 1. Characteristics and coordinates of Genomic Islands (GIs) identified in the *Ralstonia solanacearum* Species Complex and *Ralstonia* spp. genomes

Strain	Replicon	Islands	Lenght	coordinates	GC% content	Number of predicted ORFs	Insertion site
R. pseudosolanacearum							
GMI1000	Chromosome	GI1	94.3 Kb	3417490-3511839	63.3	96	tRNA-Ala
	Chromosome	GI2	80.7 Kb	872116-943347	64.2	79	tmRNA
RS476	Chromosome	GI1	94,3 Kb	3417490-3511839	63.3	96	tRNA-Ala
	Chromosome	GI2	71.2 Kb	872131-943355	64.2	79	tmRNA
CRMRs218	Chromosome	GI1	94,3 Kb	3417490-3511839	63.3	96	tRNA-Ala
	Chromosome	GI2	80.7 Kb	872131-943355	64.2	79	tmRNA
EP1	Chromosome	GI1	60.2 Kb	1234176-1294537	56.1	61	tRNA-Ser
FQY_4	Chromosome	GI2	48.5 Kb	3055877-3104433	59.1	48	tRNA-Thr
T60	Chromosome	GI1	68.9 Kb	2888558-2957522	61.8	64	tRNA-Ser
T42	Chromosome	GI1	57.6 Kb	1190504-1248147	57.6	56	tRNA-Ser
SL3822	Chromosome	GI1	71.1 Kb	1205679-1276801	59.7	56	tRNA-Ser
SL3755	Chromosome	GI1	192.5 Kb	2726603-2919152	62.4	161	tmRNA
SL3730	Chromosome	GI1	85.4 Kb	1190367-1275794	61.6	81	tRNA-Ser
SL3175	Chromosome	GI1	46.6 Kb	2672173-2718819	60.6	48	-
	Chromosome	GI2	69.3 Kb	1760989-1830354	65.2	46	-
SL3103	Chromosome	GI1	56.2 Kb	205572-261836	62.6	57	tRNA-Ala
	Chromosome	GI2	78.6 Kb	2534222-2612828	62.9	75	tmRNA
SL2729	Chromosome	GI1	94.2 Kb	1190420-1284664	62	83	tRNA-Ser
R. solanacearum							
OE1-1	Chromosome	GI1	45.3 Kb	2553544-2598873	58.4	37	tRNA-Ser
	Chromosome	GI2	35.3 Kb	3214582-3249933	58.4	35	tRNA-Thr
	Chromosome	GI3	102.6 Kb	919175-1022124	62.7	96	-
IBSBF1900	Chromosome	GI1	46.5 Kb	2622912-2669496	57.1	38	tRNA-Met
YC40-M	Chromosome	GI1	38.9 Kb	1726826-1765845	64.7	33	tRNA-Arg
	Chromosome	GI2	48.5 Kb	597355-645927	59.1	48	tRNA-Thr
Po82	Chromosome	GI1	64.7 Kb	278322-343070	63.1	55	-

UW163	Chromosome	GI1	31.8 Kb	218685-250457	62.2	32	-
IBSBF1503	Chromosome	GI1	27.5 Kb	2617958-2645495	56.8	24	tRNA-Met
RS 488	Chromosome	GI1	57.4 Kb	1201498-1258905	56.8	58	tRNA-Ser
UY031	Chromosome	GI1	57.4 Kb	1201506-1258913	56.8	58	tRNA-Ser
RS 489	Chromosome	GI1	56.0 Kb	1202826-1258883	56.8	56	tRNA-Ser
KACC 10722	Chromosome	GI1	68.6 Kb	1100412-1169028	58.7	64	tRNA-Ser
Ralstonia syzygii							
T98	Chromosome	GI1	46.6 Kb	2672156-2718802	60.6	48	-
T82	Chromosome	GI1	43.2 Kb	2363558-2406806	60.2	39	tRNA-Lys
T95	Chromosome	GI1	68.6 Kb	1107972-1176588	58.7	64	tRNA-Ser
T51	Chromosome	GI1	68.6 Kb	1053922-1122546	58.7	64	tRNA-Ser
SL2312	Chromosome	GI1	41.1 Kb	2363523-2404879	60	36	tRNA-Lys
SL2064	Chromosome	GI1	71.0 Kb	1107930-1179022	59.0	62	tRNA-Ser
PSI07 (subsp. *indonesiensis*)	Chromosome	GI1	60.3 Kb	260245-320613	60.3	65	-
R. pickettii							
12J	Chromosome	GI1	50.1 Kb	2802607-2852753	64.0	56	tRNA-Leu
12D	Chromosome	GI1	45.7 Kb	2424327-2470021	64.1	46	-

Table 2. Characteristics and coordinates of Integrative and conjugative element identified (ICEs) in the *Ralstonia solanacearum* Species Complex and *Ralstonia* spp. genomes

Name	Size (kb)	%GC	coordinates	Replicon	Number of predicted ORFs	Integrase
R. pseudosolanacearum						
ICERpGMI1000	45Kb	63.2	2780151-2825764	chromosome	50	Insertion between Ser-integrase and tRNA gene
ICERpRS476	45Kb	63.2	2780170-2825773	chromosome	50	Ser
ICERpCRMRs218	45Kb	63.2	2780419-2826032	chromosome	50	Ser
ICERpFJAT-91	51.2 Kb	60.9	1283877-1335170	chromosome	54	Tyr
ICERpFQY_4	60.3 Kb	63	1113417-1173796	chromosome	61	Tyr
ICERpYC40-M	41.4 Kb	62.3	2546800-2588238	chromosome	44	Tyr
ICERpHA4I	46.8 Kb	62.2	1709954-1756828	chromosome	48	Tyr
R. syzyzii						
ICERsT98	60.2 Kb	60.2	1925451-1985747	chromosome	60	
ICERsSL3175	60.2 Kb	60.2	1925467-1985763	chromosome	60	
R. pickettii						
ICERpi12J	54.2 Kb	64.6	2715217-2769400	chromosome	61	Tyr
R. mannitolilytica						
ICERmSN82F48	83 Kb	63.5	1635815-1718887	chromosome	78	
R. insidiosa						
ICERiFC1138	52.3 Kb	65.4	1322050-1374772	chromosome	64	

Table 2. Putative functions of cargo ORFs within conjugative element identified (ICEs) identified in the *Ralstonia solanacearum* Species Complex

ICEs	Related ICE	Cargo gene	COGs	Retention system
ICEPsGMI1000	ICEPsCRMRs218	limonene-1,2-epoxide hydrolase	Hydrolase activity	None identified
	ICEPsRS476	limonene-1,2-epoxide hydrolase	Hydrolase activity	
		putative carboxylesterase, type b; protein	Hydrolase activity	
		antibiotic biosynthesis monooxygenase YgiN	Energy production and conversion	
		putative amidase related to nicotinamidase; protein	Coenzyme transport and metabolism	
		glutathione S-transferase family protein	Posttranslational modification, protein turnover, chaperones	
ICEPsFJAT-91		peptidase	Hydrolase activity	Toxin-antitoxin system (RelE/ParE)
		alpha/beta hydrolase	Coenzyme transport and metabolism	
		enoyl-CoA hydratase	Chemical binding site	
ICEPsFQY-4	ICEPsYC40-M	Organic hydroperoxide resistance transcriptionalregulator	Resistance/Transcription	Toxin-antitoxin system (RelE/ParE)
		thiamine biosynthesis protein ThiF	Stress Response	
		thiamine biosynthesis protein ThiF	Stress Response	
		competence protein ComEC	Membrane transport	
		Glycosidase	Sugar metabolism	
		mandelate racemase/muconate lactonizing enzyme family protein	Aromatic compound metabolism	
		oxidoreductase, short-chain dehydrogenase/reductase family	Oxidation-reduction process	
		Gfa-like protein	Protein metabolism	
		putative lipoprotein	Membrane transport	
ICERsSL3175	ICERsT98	amidohydrolase	Hydrolase activity	Toxin-antitoxin system (AbiEii/AbiGii)

	type 1 glutamine amidotransferase domain-containing protein	Stress Response	
	hydroxyacid dehydrogenase	Oxidation-reduction process	
	superoxide dismutase	Stress Response	
	carboxymuconolactone decarboxylase	Aromatic compound metabolism	
	cupin	Nutrient reservoir activity	
	peptidase	Hydrolase activity	
ICERpHA4I	TIR domain-containing protein	Signal transduction	Antitoxin
	peptidase	Hidrolytic enzyme	
	glutathione S-transferase	Transferase activity	
	Hydrolase TatD	Hydrolase activity	

Figure Legend

Figure 1. Comparison between GIs (blue) and ICEs (orange). (A) Dot plot representing the distribution of the sizes between the elements. Averages sizes: GIs (65.5 kb) and ICEs (54.4 kb). (B) Dot plot showing the distribution of the GC content between GIs and ICEs. (C) Plot of the size of the element in comparison of their hosts' genome size. Shaded regions indicate the 95% confidence interval. According to the Pearson correlation coefficient, GIs plots has no linear relationship and ICEs has a negative linear relationship.

Figure 2. Representation of pairwise alignment network between GIs and ICEs. The nodes represent the GIs and ICEs, and the edges link the pairwise alignment score (the thickness of the edge is proportional to the score). The color, blue thickness represents the pairwise similarity between GIs; orange thickness represents the pairwise similarity between ICEs and purple thickness, the pairwise similarity between both elements.

Figure 3. Maximum likelihood phylogeny based on the whole element sequences. (A) Phylogenetic tree of the GIs showing sequence clusters (highlighted colored rectangles) according to species and/or geographic location. (B) the phylogenetic tree of the ICEs also highlights the same result as GIs. The tree was generated by the MEGA-X software program using the Maximum Likelihood and the algorithm of Jukes and Cantor with 1,000 bootstrap re-samplings. Bootstrap values (≥50%) are shown beside each node. Codes in blue represent the names of countries according to The International Organization for Standardization (ISO) 3166: BRA, Brazil; CHN, China; GUF, French Guyana; JPN, Japan; KOR, Korea; PER, Peru; URY, Uruguay. Star indicates genetic flux; Arrow head HTG.

Figure 4. Percentage proportion of three groups of GIs, and ICEs for each *Ralstonia* spp. chromosome. Ribbon colors represent the following categories sky blue: GIs found in *R. pseudosolanacearum* strains; Green: GIs found in *R. solanacearum* strains; Blue sapphire: GIs found in *R. syzygii* strains; pink: ICEs.

Figure 5. Schematic representation of the several novel or unrelated ICEs structures identified in the RSSC genomes. Genes are represented by arrows with different colors according to their functions. Modules clusters are color coded and their functional designations are labeled on the figure. Abbreviation for: MPF (Mating Pair Formation); Mob (mobility).

Figure 6. Mauve alignment of the integrative and conjugative elements from the *Ralstonia solanacearum* Species Complex. Colored blocks represent co-linear blocks and the histogram inside each box shows the average level of conservation in that region. Areas that are completely white were not aligned and contain sequence elements specific to a particular ICE.

Figure 7. Functional comparisons of the ORFs carried by GIs (blue) and ICEs (orange) from *Ralstonia* genomes. (A) Bar graph representing the relative ratio of COGs in each element. (B) Percent proportion of seven classes of virulence factor in the GIs. VF1 (Adhesion/Surface proteins), VF2 (Hydrolytic enzymes/Host cell wall degradation), VF3 (Plant hormones & signaling molecules), VF4 (Potential Type III secretion-dependent effectors), VF5 (Resistance to oxidative stress), VF6 (Secretion system), VF7 (Toxins) and VF8 (Type III secretion system and secreted effectors).

Supplementary information's legends

Supplementary Figure S1. Pairwise Identity Matrix of the distribution of Genomic Island and Integrative and Conjugative elements among the RSSC genomes.

Figure 1

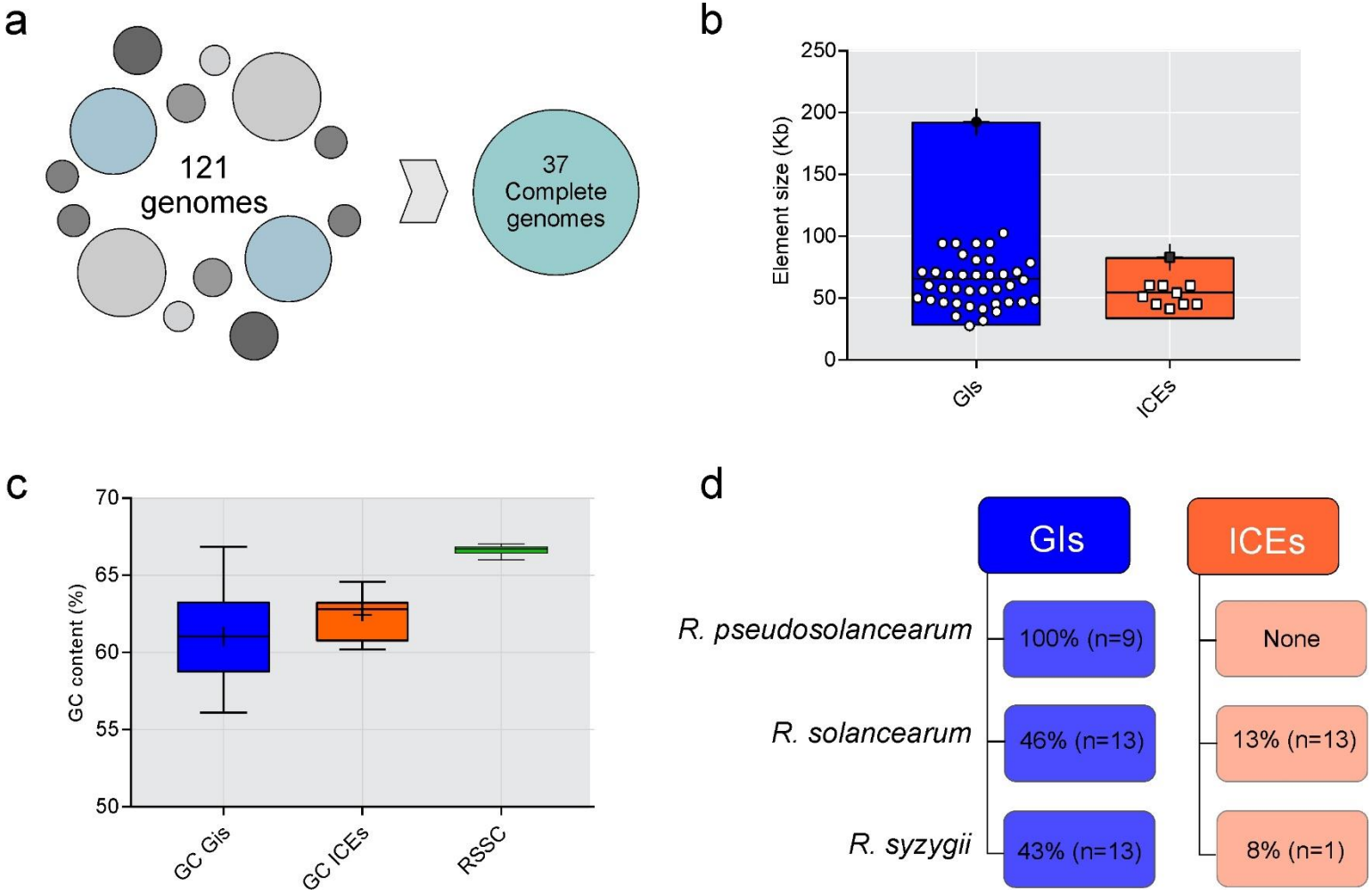

a
121
genomes
37
Complete
genomes
b
Element size (Kb)
250
200
150
100
50
0
GIs
ICEs
c
GC content (%)
70
65
60
55
50
GC GIs
GC ICEs
RSSC
d
GIs
ICEs
R. pseudosolancearum
100% (n=9)
None
R. solancearum
46% (n=13)
13% (n=13)
R. syzygii
43% (n=13)
8% (n=1)

Figure 2.

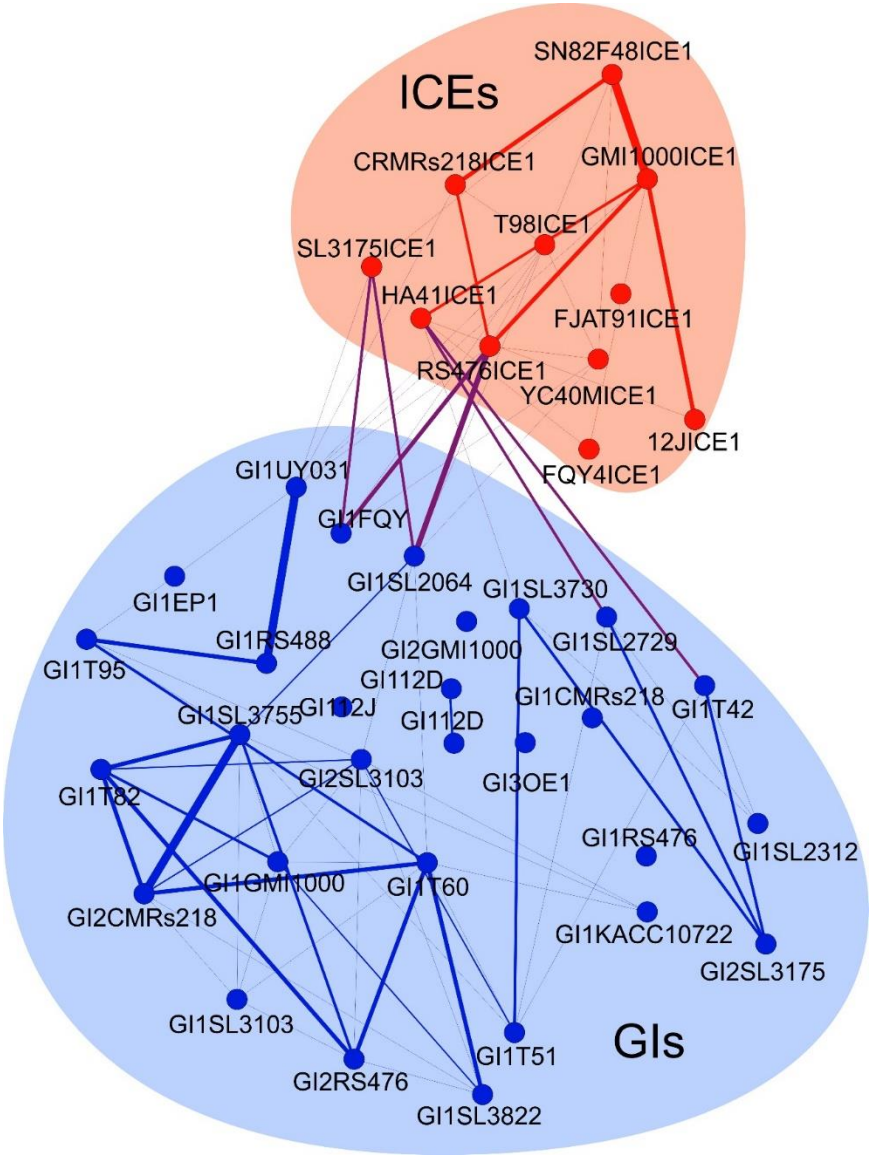

ICEs
SN82F48ICE1
CRMRs218ICE1
GMI1000ICE1
T98ICE1
SL3175ICE1
HA41ICE1
FJAT91ICE1
RS476ICE1
YC40MICE1
12JICE1
FQY4ICE1
GI1UY031
GI1FQY
GI1SL2064
GI1SL3730
GI1EP1
GI1RS488
GI2GMI1000
GI1SL2729
GI1T95
GI112D
GI1CMRs218
GI1T42
GI1SL3755
GI112J
GI112D
GI2SL3103
GI3OE1
GI1T82
GI1RS476
GI1SL2312
GI1GMI1000
GI1T60
GI2CMRs218
GI1KACC10722
GI2SL3175
GI1SL3103
GI1T51
GIs
GI2RS476
GI1SL3822

Figure 3.

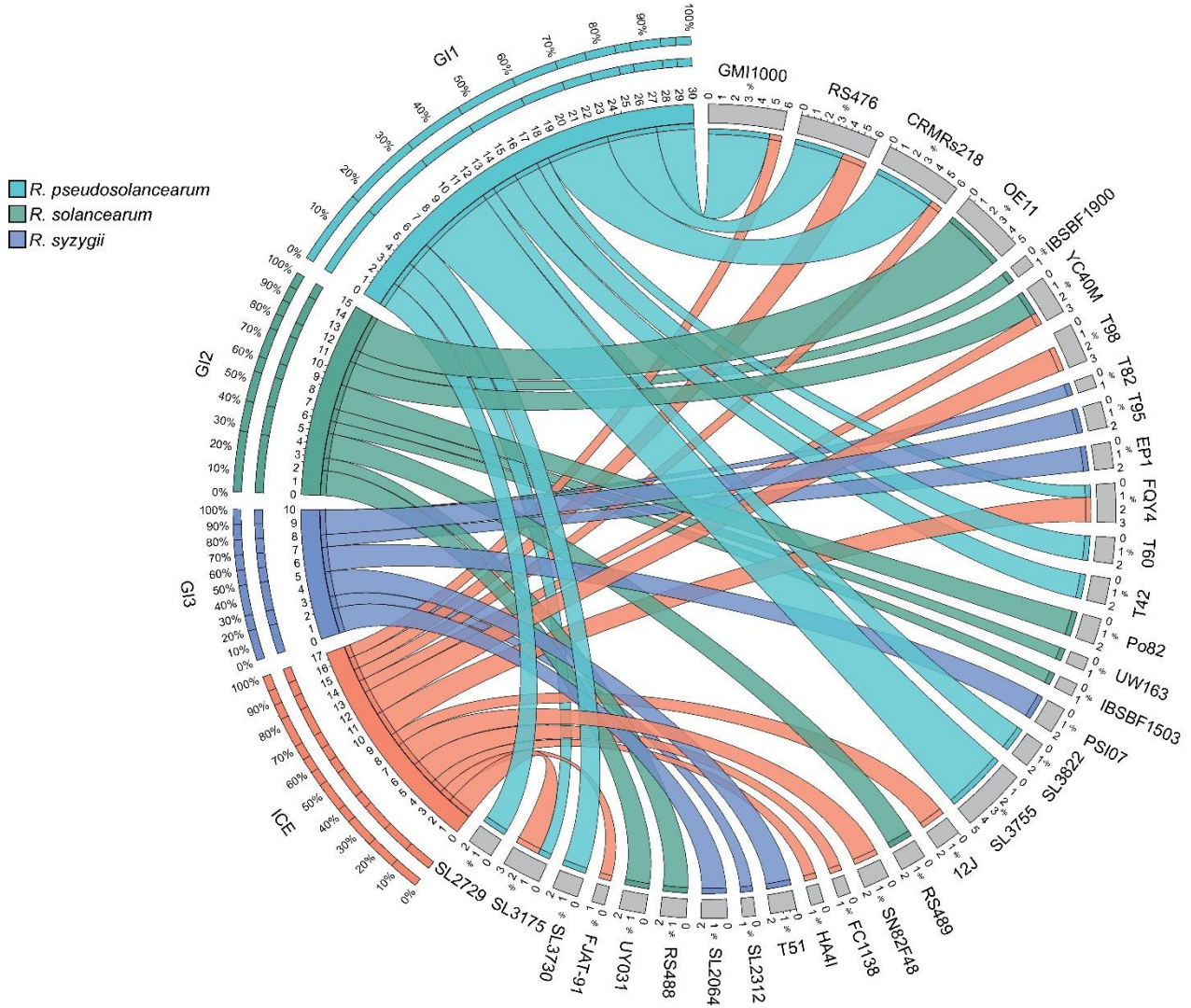

R. pseudosolancearum
R. solancearum
R. syzygii
GI1
GI2
GI3
ICE
GMI1000
RS476
CRMRs218
OE11
IBSBF1900
YC40M
T98
T82
T95
EP1
FQY4
T60
T42
Po82
UW163
IBSBF1503
PSI07
SL3822
SL3755
12J
RS489
SN82F48
FC1138
HA4I
T51
SL2312
SL2064
RS488
UY031
FJAT-91
SL3730
SL3175
SL2729

Figure 4

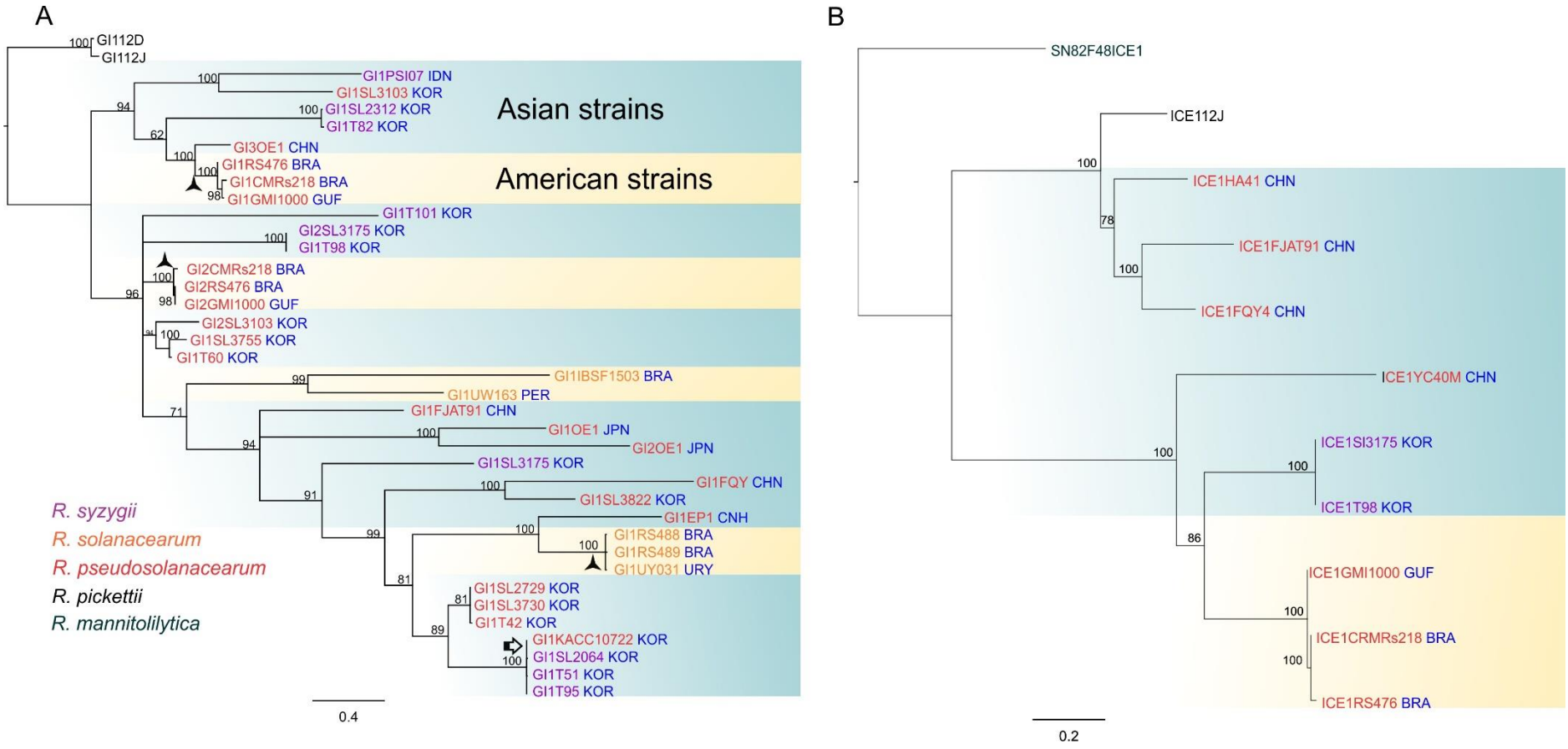

A
GI112D
GI112J
GI1PSI07 IDN
GI1SL3103 KOR
GI1SL2312 KOR
GI1T82 KOR
GI3OE1 CHN
GI1RS476 BRA
GI1CMRs218 BRA
GI1GMI1000 GUF
Asian strains
American strains
GI1T101 KOR
GI2SL3175 KOR
GI1T98 KOR
GI2CMRs218 BRA
GI2RS476 BRA
GI2GMI1000 GUF
GI2SL3103 KOR
GI1SL3755 KOR
GI1T60 KOR
GI1IBSF1503 BRA
GI1UW163 PER
GI1FJAT91 CHN
GI1OE1 JPN
GI2OE1 JPN
GI1SL3175 KOR
GI1FQY CHN
GI1SL3822 KOR
GI1EP1 CNH
GI1RS488 BRA
GI1RS489 BRA
GI1UY031 URY
GI1SL2729 KOR
GI1SL3730 KOR
GI1T42 KOR
GI1KACC10722 KOR
GI1SL2064 KOR
GI1T51 KOR
GI1T95 KOR
R. syzygii
R. solanacearum
R. pseudosolanacearum
R. pickettii
R. mannitolilytica
0.4
B
SN82F48ICE1
ICE112J
ICE1HA41 CHN
ICE1FJAT91 CHN
ICE1FQY4 CHN
ICE1YC40M CHN
ICE1SI3175 KOR
ICE1T98 KOR
ICE1GMI1000 GUF
ICE1CRMRs218 BRA
ICE1RS476 BRA
0.2

Figure 5.

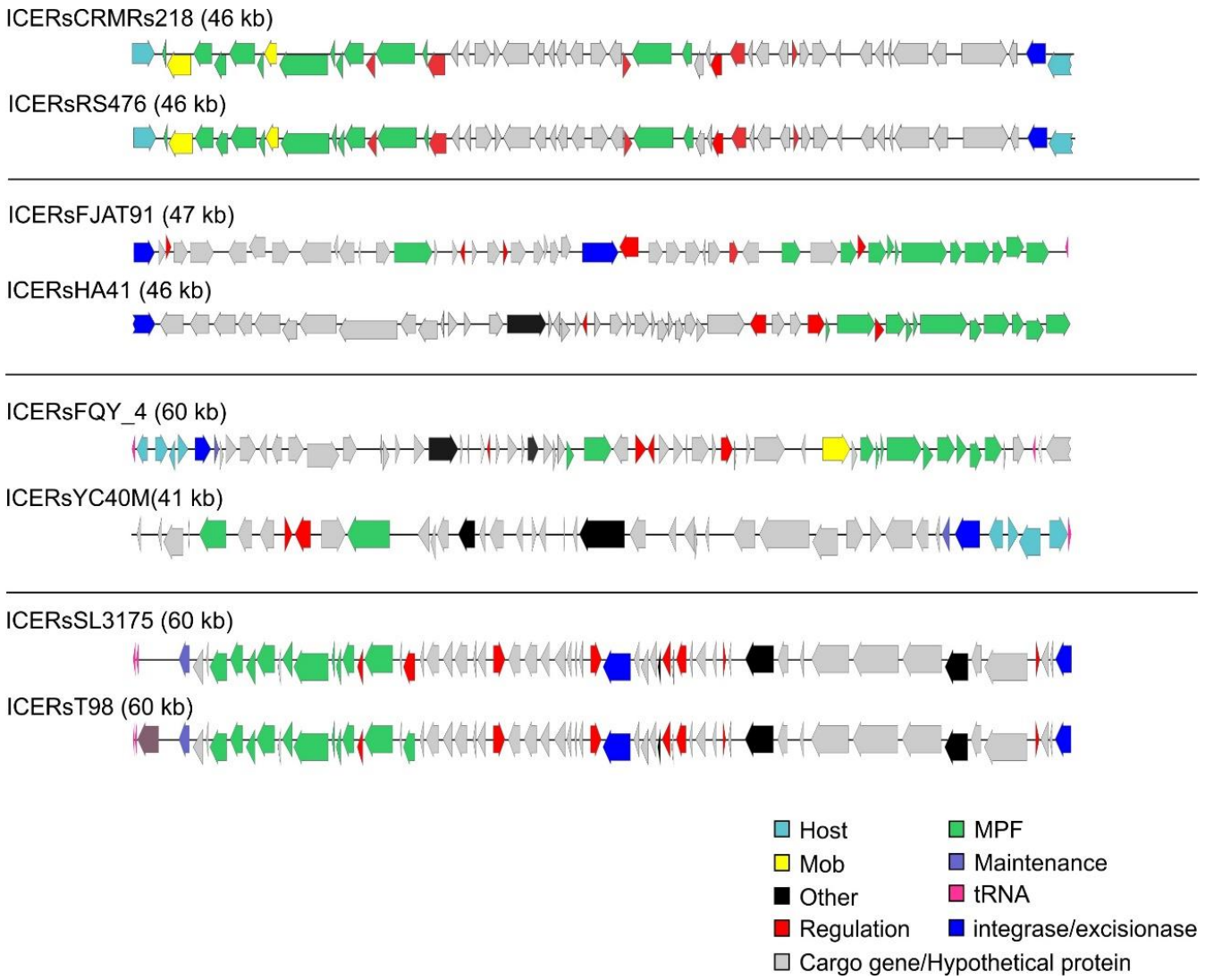

ICERsCRMRs218 (46 kb)
ICERsRS476 (46 kb)
ICERsFJAT91 (47 kb)
ICERsHA41 (46 kb)
ICERsFQY_4 (60 kb)
ICERsYC40M(41 kb)
ICERsSL3175 (60 kb)
ICERsT98 (60 kb)
Host
MPF
Mob
Maintenance
Other
tRNA
Regulation
integrase/excisionase
Cargo gene/Hypothetical protein

Figure 6.

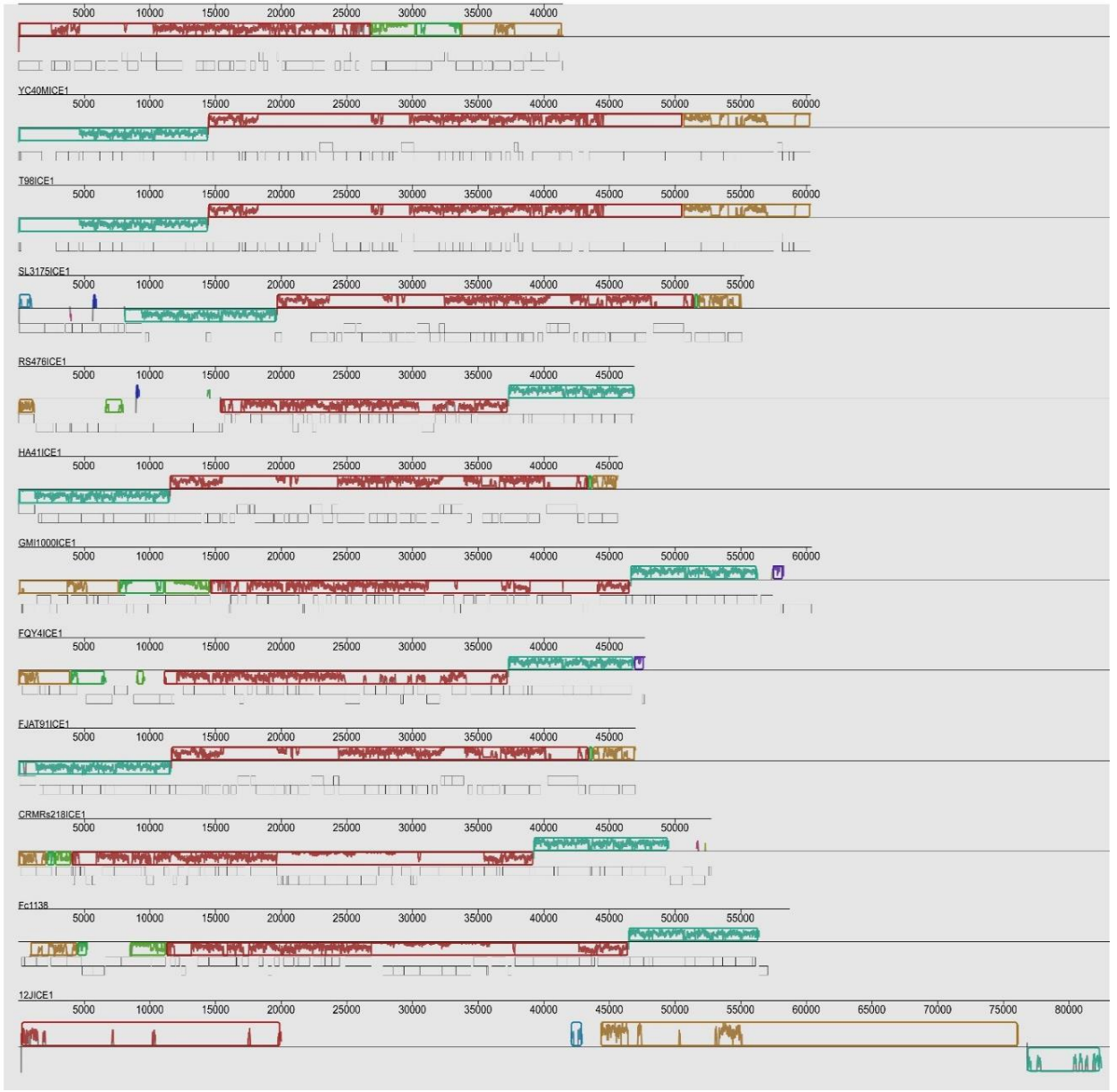
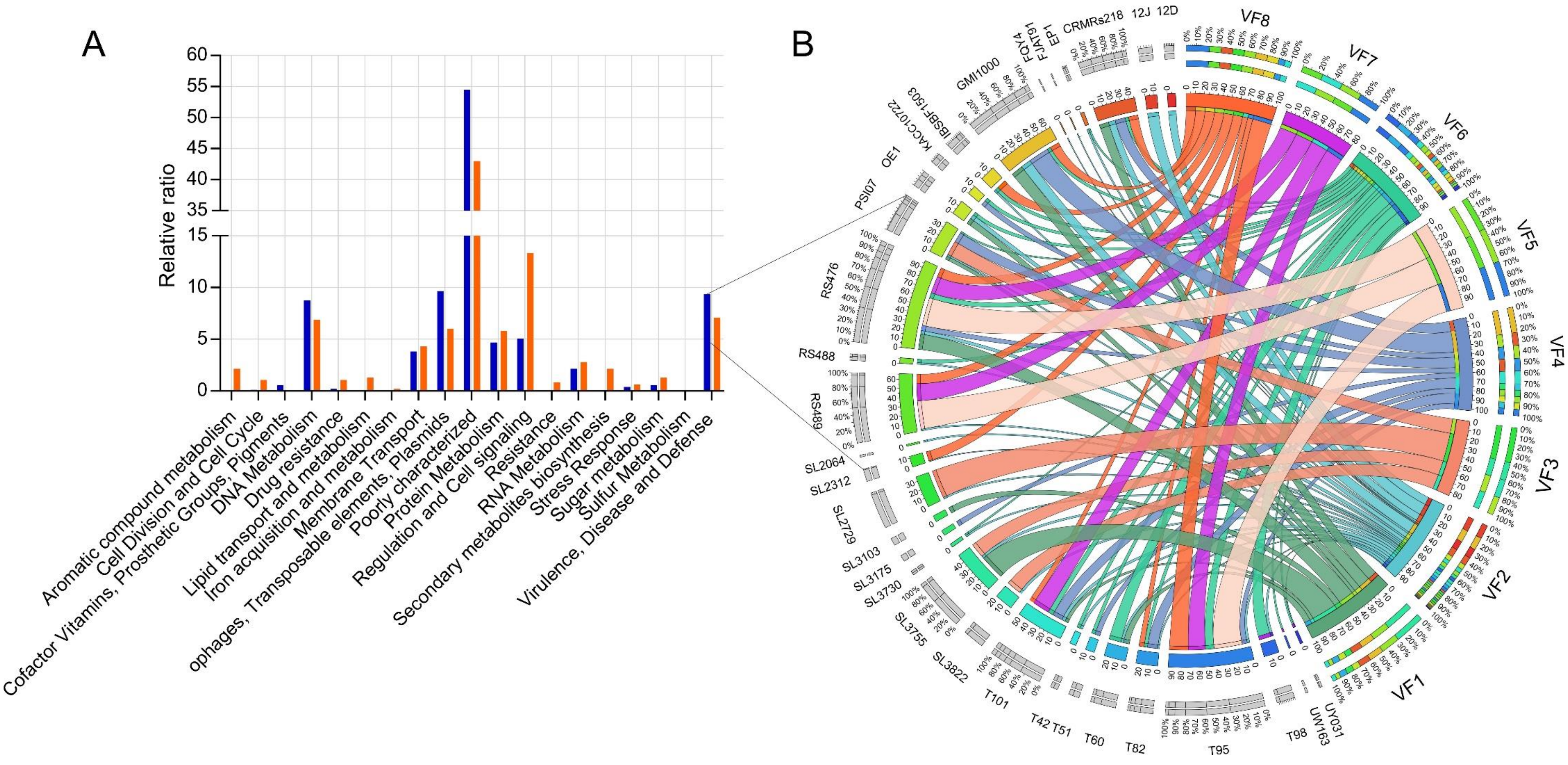
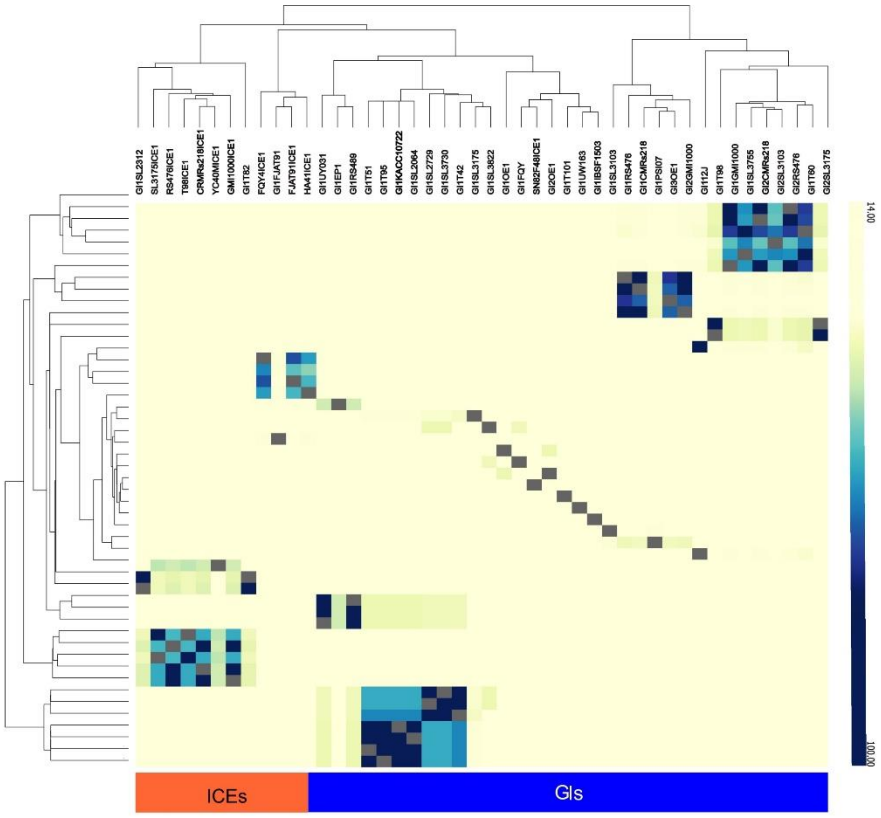

YC40MICE1
T98ICE1
SL3175ICE1
RS476ICE1
HA41ICE1
GMI1000ICE1
FQY4ICE1
FJAT91ICE1
CRMRs218ICE1
Fc1138
12JICE1

Figure 7.


A
B
Relative ratio
Aromatic compound metabolism
Cell Division and Cell Cycle
Cofactor Vitamins, Prosthetic Groups, Pigments
DNA Metabolism
Drug resistance
Lipid transport and metabolism
Iron acquisition and metabolism
Membrane Transport
ophages, Transposable elements, Plasmids
Poorly characterized
Protein Metabolism
Regulation and Cell signaling
Resistance
RNA Metabolism
Secondary metabolites biosynthesis
Stress Response
Sugar metabolism
Sulfur Metabolism
Virulence, Disease and Defense
VF1
VF2
VF3
VF4
VF5
VF6
VF7
VF8
GMI1000
FQY4
FJAT91
EP1
CRMRs218
12J
12D
IBSBF1503
KACC10722
OE1
PSI07
RS476
RS488
RS489
SL2064
SL2312
SL2729
SL3103
SL3175
SL3730
SL3755
SL3822
T101
T42
T51
T60
T82
T95
T98
UW163
UY031

Supplementary Figure S1.



CHAPTER 2

Transposable elements of the plant pathogen *Ralstonia solanacearum* provides impact on genome plasticity of the species complex

Gonçalves, O.S, Campos, K.F., Souza, T.S., Rodrigues, L.G.C., Queiroz, M.V., Santana, M.F., Transposable elements of the plant pathogen *Ralstonia solanacearum* provides impact on genome plasticity of the species complex**.** ***Genes*** will be submitted.

Transposable elements in the plant pathogen *Ralstonia solanacearum* provides impact on genome plasticity of the species complex

Osiel Silva Gonçalves[1]; Kiara França Campos[1], Thamires Santos Souza1, Luiz Guilherme do Carmo Rodrigues[1], Marisa Vieira de Queiroz[1], Mateus Ferreira Santana[1]*

[1]Departamento de Microbiologia, Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), Universidade Federal de Viçosa, Viçosa, MG, 36570-000; Brazil

*Corresponding author: Mateus Ferreira Santana

E-mail address: mateus.santana@ufv.br

Phone: +55 (31) 3612-2452

Running Title: TE in *Ralstonia solanacearum*

Abstract

The extensive genetic diversity of *Ralstonia solanacearum*, a serious soil-borne plant pathogen, led to the concept that *R. solanacearum* encompasses a species complex (RSSC). In addition, it has been suggesting that Insertion Sequences (ISs) play an important role in the genome evolution of this pathogen. Here, we aimed to identify transposable elements, ISs and Transposon, in 106 RSSC genomes and 15 *Ralstonia* spp.

2,957 IS elements were found in the genome of 59 representative RSSC strains and closely related *Ralstonia* spp. A unique set of 13 IS families were found highly conserved across the strains. In particular to the IS family, IS3 and IS5 were found as the most abundant. Strains closely related tend to have similar patterns of ISs, suggesting their ancestral acquisition. Our

results also showed four novel Tn3 transposons sequences carrying passenger genes related to antibiotic resistance and avirulence protein. Internal rearrangements events associated with IS were demonstrated in *R. pseudosolanacearum* strains. We also mapped IS elements interrupting avirulence genes, which provides evidence that IS playing an important role RSSC pathogenicity evolution. In addition, the activity of ISs was demonstrated by DNA hybridization in environmental population. Our results provide a collective data TEs in RSSC genomes, opening a new path for understanding their evolutionary impact over the genome evolution, and pathogenicity of this important plant pathogen.

Keywords: Genome evolution, Insertion sequence, mobile DNA, Transposon

Introduction

Plant–pathogen interaction is intimate, complex and ancient, resulted of never-ending war (Schneider and Collmer 2010; Peyraud et al. 2017). Understanding how plant pathogenic bacteria are evolving to overcome plant resistance is crucial for designing strategy of disease control. However, many evolutionary aspects of this interaction remain under-studied. In recent association with hosts, some bacterial genomes are subject of remarkable variation such insertions, duplications, inversions, and translocations until stabilizing a long-term association with hosts (Mira et al. 2002; Bobay and Ochman 2017). In part, this process can be archived by the accumulation of repetitive DNA, including transposable elements (TEs), prophages and paralogous genes, and that many of these sequences are recognize as non-functional sequences, which can play an important evolutionary role as concerns to specialized host-adaptation (Gil and Latorre 2012).

Pay special attention in TEs, several recent pathogens possess relatively high numbers of these mobile elements, which may be responsible for bottlenecking relationship over

pathogen and hosts (Bobay and Ochman 2017). In bacteria, TEs are self-replicable intracellular mobile genetic elements assign to transposons (Tn) and insertion sequences (ISs). Typically, TEs have a single or multiple open reading frames (ORF) that encodes a transposase protein, which it is required for insertion and through so-called "cut and paste" mechanism. They have inverted terminal inverted repeats (TIRs) and are flanked by short direct repeats (DRs). These elements are distinguished basically because Tn carries passenger genes not involved in catalyzing or regulating TE movement (Kleckner 1981). By contrast, ISs elements are typically the smallest TEs (<2 Kb), which dramatically shape genetic sequence content by causing mutation, insertion, deletion, inversion of DNA and alteration of gene expression (Siguier et al. 2014).

It is believing that this process represents a great source of genomic diversification, allowing rapid evolution of a pathogen or stimulate the emergence of new pathogenic races that now might outbreak the disease to overcome host defenses (Arnold and Jackson 2011). For the bacteria *Ralstonia solanacearum*, a serious soil-borne plant pathogen to agricultural production due to its an extensive host range and aggressiveness, it has been suggesting that ISs may be an important role in genome evolution (Salanoubat et al. 2002). However, no complete analysis of TEs has been reported in the *R. solanacearum.*

R. solanacearum genome is organized into two circular replicons, chromosome and megaplasmid. Both replicons encoded housekeeping and accessory genes, they have same genomic features (dinucleotide relative abundances, codon usage, and similar distribution and composition of simple sequence repeats), which suggests the co-evolution over a long-time span (Salanoubat et al. 2002, Castillo and Greenberg 2007). Genome comparison of representative strains of *R. solanacearum* showed that genomic features, such as size, G+C content, number of genes, were conserved across the strains. However, it has been demonstrated many genomic rearrangements (e.g., inversion and translocation), deletion and insertion of

DNA among the strains (Remenant et al. 2010; Li et al. 2016). Due to genome differentiation, it was proposed that *R. solanacearum* Species Complex (RSSC), which include *R. syzygii* and blood disease bacteria (BDB), encompasses three distinct species: *R. pseudosolanacearum* (formerly phylotypes I and III), *R. solanacearum* (IIA and IIB) and *R. syzygii* (formerly phylotypes IV and BDB) (Safni et al. 2014, Prior et al. 2016). In order to investigate the evolutionary impact of TEs over genome evolution of RSSC, we analyzed these elements together in the genomes of 106 RSSC strains and 15 *Ralstonia* spp. collected from diverse plant hosts and geographic origin available from public sequence database.

Material and Methods

Genome data

In this study, we performed the analysis in the 121 genomes from *Ralstonia* genera. A total number of genomes are as follows: 106 RSSC strains spanning 3 different species and 15 *Ralstonia* species belonging to *R. pickettii* and *R. mannitolilytica*, two closest species to *R. solanacearum* (Table S1). We expanded the genomes number in order to strengthen our analysis. The genomes were downloaded from the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/ genome) in December 2018.

Detection of TE sequences

ISs were predicted by BLASTn alignment against the ISfinder database, using default parameters (E-value $\leq 10^{-5}$) (Siguier et al. 2006). The information of direct repeats (DR), insertion sites and tandem inverted repeat (TIR) were manually identified and annotated using Geneious® 11.1.5 (Biomatters Lts). Tn sequence was identified, by screening our local database of ISs to search Tn transposase family. In the following, predicted transposase were inspected for DR and TIR sequences that define the boundaries of the transposon. The complete

nucleotide sequence in Genbank format of corresponding records was imported into Geneious® to help delimit genomic regions flanking the element.

Virulence and Antimicrobial resistance–associated genes in Tn elements

Virulence and antimicrobial resistance genes were identified by performed a BLASTp (Using parameters e-value ≤ 10^{-4} and amino acid identity > 30%) on Pathogen–Host Interactions database (PHI-base, www.phi-base.org) (Urban et al. 2017) and by a standard BLASTn searches against the The Comprehensive Antibiotic Resistance Database (CARD, http://arpcard.mcmaster.ca) (Jia et al. 2017), respectively.

Integration profile analysis

We selected seven *R. solanacearum* strains that were isolated from the soil sample, as detailed in the Table S2. The isolates were cultured in CPG medium containing casein (1g/L), peptone (10g/L) and glucose (5g/L) at 28°C with shaking at 150 rpm. Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega Co, United States), according to manufacture recommendation, checked for quality using a NanoDrop 2000 (Thermo Scientific, Waltham, MA, United States) and gel electrophoresis. Probes for IS1021 and IS010 were prepared using PCR DIG ´Probe Synthesis Kit (Hoffmann–La Roche, AG, Switzerland). For genomic Southern hybridization, 10 µg DNA was digested with EcoRI and incubated overnight at 37°C. The gel was treated with 0.25 N HCl for 30 min, followed by denaturation buffer (0.5 M NaOH. 1.5 M NaCl), and the digested chromosomes were transferred to a nylon membrane (Hybond N+; Amersham, Amersham, UK) washed in 0.4 M NaOH for about 72 hours. The following procedure after the transfer was performed as an early study. For stripping the hybridized probe, the used membrane was washed twice, for 15 min each, with 0.2 M NaOH, 0.1% SDS at 37°C, then the membranes were soaked in 2×SSC for 5 min and dried (Sambrook

and Russsel 2001). To confirm absence of hybridizations, IS1021 and IS010 genes were amplified (initial denaturation step at 95°C for 5 min, followed by 35 cycles at 95°C for 30 s, 53°C for 45 s and 72°C for 45 s, followed by a final step at 72°C for 5 min). 16S rRNA genes were also amplified as positive control (initial denaturation step at 94°C for 4 min, followed by 30 cycles at 94°C for 40 s, 58°C for 45 s and 72°C for 1,30 min, followed by a final step at 72°C for 7 min) using the universal primers 27F and 1401R.

Chromosomal rearrangements comparison

The 16S rRNA gene sequences were obtained from the NCBI database and a distance matrix o was constructed using ClustalW (Larkin et al. 2007). Next, all the sequences were aligned and a phylogeny tree was constructed in MEGA X using Maximum Likelihood (1,000 bootstrap replicates) (Kumar et al. 2018). The generated output file (.tree) was visualized and annotated on the Interactive Tree of Life (iTOL) interface v4 (https://iTOL.embl.de/) (Letunic and Bork 2019). Genome sequences of the strains KACC10722, T110, and SEPPX05 were obtained from NCBI in .gbk format. Our main selection criteria were the richness of IS copies on the chromosome. Multiple genome alignments were performed by the Mauve software (version 2.3.1) (Darling et al. 2004) following the parameter: (computational settings: alignment with progressive Mauve (aligner: Muscle 3.6), default seed weight (15), full alignment (minimum island size: 50, maximum backbone gap size: 50, minimum backbone size: 50), use of seed families: yes, iterative refinement: yes, determination of LCBs: yes).

Results

Great diversity of IS elements and Transposons in the *Ralstonia* spp. genomes

Our analysis shows high numbers of IS copies in the chromosome and megaplasmid of *Ralstonia* strains. From 121 genomes, we identified IS sequences only in complete genomes

(n=59). The overall numbers of IS elements were 1,544 copies in the chromosome and 1,413 copies in the megaplasmid, representing a total of 2,957 IS copies (Table S3). In the attempt correlated the all IS elements in the frame of genome sequences accumulation, we performed a correlation analysis. Results imply that there is no linear correlation between the IS size and genome size ($R^2 = 0.08$) (Fig. 1a). In the following, we computed the IS family distribution in the replicons. Our results show a unique set of 13 IS families highly conserved across the chromosome and megaplasmid of the RSSC, being the IS5 the most abundant family in number of copies (n=1,944), follow by IS3 family with 496 copies (Table 1). IS length ranged from 864 to 2956, found respectively in the IS1421 and Tn3 family. Replicon containing exclusive IS family was found, such the family IS1595 and IS1182 only found integrated in the megaplasmid, and IS200-IS605 family in the chromosome (Fig. 1b). Altogether, IS elements cover in average 1% of *Ralstonia* genome, varying from 0.1% in the *R. syzygii* to 3.9% in the *R. pseudosolanacearum*.

Within 13 families, 64 distinct IS elements were found. Phylogenetic analysis reveals several coherent clusters, which were identified to compose the IS3, IS5, Tn3, IS21, IS256 and IS1182 family (Fig. 2). We compared the IS families against the corresponding set of *Ralstonia* complete genomes. Genome comparisons revealed pattern of IS families among the RSSC strains (Fig. 3). Genomes that are closely related tend to have similar a pattern of IS element. All RSSC genomes share a set of IS belong the IS3 and IS5 family, suggesting that these elements are conserved or ancestral acquired through the complex. *R. pseudosolanacearum* strains share numerous and diversity of IS elements (n= 2,284), follow by *R. solanacearum* (n=422) and *R. syzygii* strains (n=251). Several species-specific IS elements were noticed, such IS200-IS605 shared among *R. pseudosolanacearum* (strains GMI1000, RS476 and CRMRs218); IS1595 only found in the strains KACC10709, SL3103, SL2330, SL3300 and T25; and IS1182 specific for African strain CMR15. The average copies of ISs is roughly larger

on the RSSC genomes than non-plant pathogenic *Ralstonia* spp. (51 vs 16). The 2,957 insertions found in 59 genomes of *Ralstonia* spp. were characterized in detail (Table S3).

Four transposons belong to Tn3 family were identified in two *R. pseudosolanacearum* strains RSSCM and HA4-1, named here as Tn3RSSCM_1, Tn3RSSCM_2, Tn3pHA41 and Tn3HA41_2 (Fig. 4). Tn3RSSCM_1 and 5Tn3RSSCM_2 were respectively found in the chromosome and megaplasmid of strain RSSCM; and Tn3pHA41 and Tn3HA41_2 found in the megaplasmid and plasmid of the strain HA4-1, respectively. Tn3RSSCM_1 and 5Tn3RSSCM_2 range in size from 5.1 Kb to 5.8 Kb long, respectively. While Tn3pHA41 and Tn3HA41_2 range from 6.6 kb to 8.5 kb long. Together these transposons share from 78% to 89% of sequences identity and are exclusively identified in these two Chinese strains. Commonly, all transposon coding Tn3 transposase family, recombinase gene, which ensures the transposition process. They are flanked by a typical Tn3 family IRL sequence with 51 bp long (GGGGCCGTCTCAGAAAACGGAAAAAATCGTACGCTAAGCCCGGGTTGATGC) and IRR sequence with 42 bp long (GGGGTCGTCTCAGA AAACGGAAAAAATCGTACGCTAAGCTCG), also an 8bp long DR (CAAGATGG). Tn3RSSCM_1 chromosome copy also includes two hypothetical protein and nucleotidyltransferase as passenger genes. Tn3RSSCM_2 megaplasmid copy contains additional recombinase and passenger genes coding for peptidase C55. This is similar to *Tn3HA41_2*, excepted for the IS21 family transposase. *Tn3pHA41* plasmid copy contains a pair of IS5 family transposase, hypothetical protein gene and also additional passenger genes coding for avirulence effector proteins AviRxv. For our knowledge, this is the first available study reporting transposon element in the RSSC genome. Our results opening new paths for classical studies on transposition and their potential impact in pathogenicity of *R. pseudosolanacear*um.

IS mediates genomic rearrangements

IS elements can shape genetic rearrangements by causing many insertions, deletion, inversion (Vandecraen et al. 2017). To assess the impact IS elements, we performed a synteny alignment analysis of *Ralstonia* chromosome. Our main selection criteria were the richness of IS copies on the chromosome. Three *R. pseudosolacearum* strains (SEPPX05, KACC10722, and T110) were selected. KACC10722 have 2 copies, while T110 has none IS copies on the chromosome sequence. Notably, these two strains share collinear syntenic blocks. By contrast, *R. pseudosolacearum* strain SEPPX05 have 120 IS copies, therefore our analysis revealed numerous internal rearrangements events with a subset being associated with IS (Fig. 5). This suggests that genomes possessing a higher number of IS copies are under larger impact of these elements on its organization.

Transposable elements linked to virulence coding regions

It is thought that pathogens under bottlenecking relationship over hosts have a strong effect of IS elements on its genome (Stapley et al. 2015). We pay special attention in the insertion of these elements in intergenic regions of RSSC virulence factors. To assess the impact of IS elements, they were classified into three classes: IS insertions within virulence ORF; with one or two truncated virulence ORF nearby (less than 100 nucleotides distant); nearby ORF coding virulence gene. Most of the elements were found truncating, inserted nearby and between intergenic regions of virulence factor mainly found in chromosome sequence (Fig. 5a). These insertions cover hemagglutinin-related gene, genes related to the type II secretion system, and a number of type III effector protein (T3EP). Details of the insertion are found in the Table S4. Our analysis reveals that 51% (n=18) of T3EP genes may be affected by insertions of ISs. Fig. 5b illustrates three examples, representing the classes, mapped across the RSSC genomes. In *R. pseudosolanacearum* strain YC40-M an IS5 insertion site is present within T3EP gene, and an IS116/IS110/IS902 element disrupted T3EP gene. An intergenic insertion site is present

upstream of the T3EP gene and downstream hypothetical protein in *R. pseudosolanacearum* strain T117. Disrupted T3EP gene by ISs, representing the most common insertion site. In subsequent analysis, we performed a BLASTx analysis of T3EP against PHI-database to characterize the genes. More than half of the T3EP genes were identified as avirulence genes (*avr*), indicating the inactivation of *avr* by IS elements (Table S5). Interestingly, we found the IRL and DR sequences of the IS5/IS1021 elements overlapping the start of the *avr* gene in the related strains RS488, RS489 and UY081 (Fig. S1). Apparently, the sequence is complete, and that might suggest activity of IS over *avr* expression.

Concordance with IS elements activity in *R. solanacearum* isolates

Having demonstrated many IS copies *in silico* analysis, we searched for evidence that such elements manifest *in vitro* in the *R. solanacearum* population. We selected seven *R. solanacearum* strains that were isolated from the soil sample in Minas Gerais and Brasília, Brazil (Figure S2a, Table S2), and performed Southern blot experiment using IS5 family transposase IS1021 and IS010 elements as probes (Fig. S2b). Prevalence and abundance of these elements in Brazilian *R. solanacearum* strains were our main selection criteria. Most of the isolates show the hybridization pattern for IS1021 and IS010 elements (Fig. S2c). This suggests that the elements are detected in Brazilian *R. solanacearum* isolates. Inside this population, we found a great polymorphism in the number of copies and also isolates with no DNA band patterns detected. For those, we confirmed by doing PCR (data not shown).

Discussion

We report a systematic TEs identification in the 120 genomes from *Ralstonia solanacearum* Species Complex and closely related *Ralstonia* spp. (*R. pickettii*, *R. insidiosa*, *R. mannitolilytica*). Nevertheless, our analysis only found TEs elements in 59 complete genome

sequences. ISfinder missed out the identification in draft genomes. In total 2,957 ISs copies were found in the chromosome and megaplasmid sequences. Although we have demonstrated a high number of ISs, our result did not show suitable evidence of its accumulation in *Ralstonia* genome. We show that *Ralstonia* spp. share a unique set of IS families, mainly IS3 and IS5 family, which might indicate their ancestral acquisition throughout the species. IS3 family have been found in 270 bacterial species, over 554 members. This family is characterized by two consecutive transposases containing 1,200 and 1,550 bp and inverted terminal repeats in the range of 20 to 40 bp (Kleckner 1981, Siguier et al. 2015). Similar, IS5 also encode two consecutive transposases, but this group is relatively heterogeneous (Siguier et al. 2015). Together, these findings may indicate the preference of consecutive transposases configurations in the genome of *Ralstonia* spp. Especially for the RSSC, closely related strains tend to have similar patterns of IS element. It has been demonstrated that IS elements have the ability to quickly multiply in genomes, resulting in that closely related strains have a number of IS elements similar (Adam et al. 2016). We noticed that *R. pseudosolanacearum* strains share a numerous and diversity of IS families, follow by *R. solanacearum* and *R. syzygii*. This result also reflects over the coverage of these elements in the genome, being the *R. pseudosolanacearum* strain SEPPX05 genome composed of 3.9% IS sequences.

Besides IS elements, we also reported the presence of four novel transposons sequences to belongs to Tn3 family. These transposons share more than 70% of identity and were found in the Chinese *pseudosolanacearum* strains RSSCM and HA4-1 isolated from *Cucurbita maxima* and *Arachis hypogaea*, respectively (She et al. 2017). Interestingly, theses transposons were only found in these strains. The transposon Tn3RSSCM_1 encode an aminoglycoside nucleotidyltransferase as passenger genes. This enzyme confers resistance to a wide range of aminoglycoside such as kanamycin A, which act transferring the nucleoside monophosphate group from a nucleotide to the 4'-hydroxyl group of kanamycin A (Fling and Richards, 1985;

Krause et al. 2016). Although the wilt disease caused by *R. solanacearum* is not managed by antibiotics, our result shows acquisition of antibiotic resistance in this important plant pathogen. Such knowledge is also critical because the genes are transferred by the mobile element, which potentially might evolve other bacteria in the environmental condition to the acquisition of resistance via Horizontal Gene Transfer (HGT). Genes related to *avr* were also mapped as passenger genes in the sequences of Tn3RSSCM_2, Tn3HA41_2 and Tn3pHa41. *avr* genes is a general term to indicate an effector gene that encodes any determinant of the specificity of the interaction with the host (Leach and White 1996). Therefore, *avr* being HGT is recognized as a major epidemiological factor in new disease outbreaks (Gabriel, 1999), suggesting the role of these transposons on the pathogenicity of RSSC. Together, our findings demonstrated a collective data, which primarily shows that these elements might have a remarkable potential to impact *Ralstonia* genome.

As we demonstrated, the genome of *R. pseudosolanacearum* strain SEPPX05, which possess a high number of IS copies, is the under larger impact of these elements on its organization, as compared with two other *R. pseudosolanacearum* strains (KACC10722 and T110) with the low number of copies. Besides our study, it has been demonstrated some deletions, insertions, and inversions, of SEPPX05 as compared to most representative RSSC strains (Li et al. 2018). In addition, KACC10722 and T110 have in common the potato as host and all three strains have very high economic damage in the crop from China (Jiang et al. 2017). To our knowledge, SEPPX05 can only cause sesame bacterial wilt disease. Therefore, we might suggest the important role of IS-mediated genome rearrangements from SEPPX05 to new host colonization.

Having shown the effect of IS elements in genome plasticity, we then look closer to the impact of IS in modulating RSSC virulence genes. Most IS elements were found between intergenic of hemagglutinin-related protein, a class of adhesins produced by diverse pathogenic

bacteria, which is responsible for the bacteria attachment during plant-pathogen interaction (Genin and Denny, 2009). IS transposition can active the expression of a gene when its insertion creates an alternative target host gene promoter or as result of read-through transcription (Vandecraen et al. 2017). We mapped a few examples of inserted within virulence ORFs, such as IS overlapping type II secretion system F family and hemolysin-like, both secreted proteins exported by ABC systems (Preston et al. 2005). Thus, possible interference of ISs in the transcription of these genes may be suggested. Jeong and Timmis (2000) reported transposition mediated ISRso4 (IS5family) in *R. solanacearum*, the inactivation of the global regulatory gene *phcA*, which in turn modulated the expression of extracellular polysaccharides. In this study, a great number of *avr* genes interrupted by the insertion of an IS were found. *Avr* genes are one major key during the plant – pathogen interaction as described by the gene-for-gene theory (Flor, 1971). The theory relies on the relationship between pathogen and host plant cultivars, this interaction occurs between an *avr* gene in the pathogen and an *R* (Resistance) gene in the plant. When a pathogen possessing an *avr* attacks the plant which carries the corresponding *R* gene, resistance is induced in the plant and the disease does not occur. Therefore, the inactivation of *avr* genes in the bacteria can lead to virulence on a resistant host plant (Leach and White 1996; Grennan 2006). Similar to this result, IS sequences interrupting *avr* genes have been related to in *Pseudomonas syringae* (Kim et al. 1998; Deslandes and Rivas 2012) and too as a mechanism for the emergence of races in *Fusarium oxysporum* (Inami et al. 2012). In conclusion, these results underlying evidence that mobile genetic elements may be one of the driving forces RSSC pathogenicity evolution.

To assess the activity of IS elements predicted *in silico*, two IS elements were analyzed in seven strains of *R. solanacearum*. The observed band polymorphism leads the hypothesis that these elements are involved in diversification (Lee et al. 2016). Our analysis shows suitable evidence between *in silico* predicted IS elements and *in vitro*, confirming the widespread

distribution of these elements among environmental isolates. Also, might indicate a recent activity of IS elements among the *R. solanacearum* population from Minas Gerais and Brasília. In conclusion, the research described here opens up new avenues for understanding the evolutionary impact of TEs over the genome evolution, and pathogenicity of the RSSC.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001; Fundação Arthur Bernardes (Furnabe; Minas Gerais, Brasil) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; Brasília, Brasil).

Conflict of Interest

The authors have no conflicts of interest to disclose.

References

Adams MD, Bishop B, Wright, MS (2016) Quantitative assessment of insertion sequence impact on bacterial genome architecture. Microb Genomics 2: e000062–e000062. https://doi.org/10.1099/mgen.0.000062

Arnold DL, Jackson RW (2011) Bacterial genomes: evolution of pathogenicity. Curr Opin Plant Biol 14: 385–391. https://doi.org/https://doi.org/10.1016/j.pbi.2011. 03.001

Bobay LM, Ochman H (2017) The Evolution of Bacterial Genome Architecture. Front Genet 8:72. https://www.frontiersin.org/article/10.3389/fgene.2017.00072

Castillo JA, Greenberg JT (2007) Evolutionary dynamics of *Ralstonia solanacearum*. Appl Environ Microbiol 73:1225–1238. https://doi.org/10.1128/AEM.01253-06

Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704

Deslandes L, Rivas S (2012) Catch me if you can: bacterial effectors and plant targets. Trends Plant Sci 17:644–655. https://doi.org/https://doi.org/10.1016/j.tplants.2012. 06.011

Fling ME, Kopf J, Richards C (1985) Nucleotide sequence of the transposon Tn7 gene encoding an aminoglycoside-modifying enzyme, 3"(9)-O-nucleotidyltransferase. Nucleic Acids Res, 13:7095–7106. https://doi.org/10.1093/nar/13.19.7095

Flor HH (1971) Current Status of the Gene-For-Gene Concept. Annu Rev Phytopathol 9:275–296. https://doi.org/10.1146/annurev.py.09.090171.001423

Gabriel DW (1999) Why do pathogens carry avirulence genes? Physiol Mol Plant Pathol 55:205–214. https://doi.org/10.1006/PMPP.1999.0230

Genin S, Denny TP (2012) Pathogenomics of the *Ralstonia solanacearum* Species Complex. Annu Rev Phytopathol 50:67–89. https://doi.org/10.1146/annurev-phyto-081211-173000

Gil R, Latorre A (2012) Factors behind junk DNA in bacteria. Genes (Basel), 3:634–650. https://doi.org/10.3390/genes3040634

Grennan AK (2006) Plant response to bacterial pathogens Overlap between innate and gene-for-gene defense response. Plant Physiol 142:809–811. https://doi.org/10.1104/pp.106.900207

Inami K, Yoshioka-Akiyama C, Morita Y, Yamasaki M, Teraoka T, Arie T (2012) A Genetic Mechanism for Emergence of Races in *Fusarium oxysporum f. sp. lycopersici*: Inactivation of Avirulence Gene AVR1 by Transposon Insertion. PLoS One 7: e44101. https://doi.org/10.1371/journal.pone.0044101

Jeong EL, Timmis JN (2000) Novel Insertion Sequence Elements Associated with Genetic Heterogeneity and Phenotype Conversion in *Ralstonia solanacearum*. J Bacteriol 182:6541. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC94808/

Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang, KK, McArthur AG (2016) CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 45: D566–D573. https://doi.org/10.1093/nar/gkw1004

Jiang G, Wei Z, Xu J, Chen H, Zhang Y, She X, Liao B (2017) Bacterial Wilt in China: History, Current Status, and Future Perspectives. Front Plant Sci 8: 1549. https://doi.org/10.3389/fpls.2017.01549

Kim JF, Charkowski AO, Alfano JR, Collmer A, Beer SV (1998) Sequences Related to Transposable Elements and Bacteriophages Flank Avirulence Genes of *Pseudomonas*

syringae. Mol Plant-Microbe Interact 11:1247–1252. https://doi.org/10.1094/MPMI.1998.11.12.1247

Kleckner N (1981) Transposable Elements in Prokaryotes. Structure 15:341–404.

Krause KM, Serio AW, Kane TR, Connolly LE (2016) Aminoglycosides: An Overview. Cold Spring Harb Perspect Med 6(6): a027029. https://doi.org/10.1101/csh perspect.a027029

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular biology and evolution 35. https://doi.org/10.1093/molbev/msy096

Larkin MA, Wilm A, Higgins DG, Valentin F, Blackshields G, McWilliam H, Gibson TJ (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Leach JE, White FF (1996) Bacterial avirulence genes. Annu Rev Phytopathol 34:153–179. https://doi.org/10.1146/annurev.phyto.34.1.153

Lee H, Doak TG, Popodi E, Foster PL, Tang H (2016) Insertion sequence-caused large-scale rearrangements in the genome of *Escherichia coli*. Nucleic Acids Res 44:7109–7119. https://doi.org/10.1093/nar/gkw647

Letunic I, Bork P (2019) Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 1. https://doi.org/10.1093/nar/gkz239

Li P, Wang D, Yan J, Zhou J, Deng Y, Jiang Z, Zhang L (2016) Genomic Analysis of Phylotype I Strain EP1 Reveals Substantial Divergence from Other Strains in the *Ralstonia*

solanacearum Species Complex. Front Microbiol 7:1719. https://doi.org/10.3389/fmicb.2016.01719

Li X, Huang X, Chen G, Zou L, Wei L, Hua J (2018) Complete genome sequence of the sesame pathogen *Ralstonia solanacearum* strain SEPPX05. Genes Genomics 40:657–668. https://doi.org/10.1007/s13258-018-0667-3

Mira A, Klasson L, Andersson SGE (2002) Microbial genome evolution: sources of variability. Curr Opin Microbiol 5:506–512. https://doi.org/https://doi.org/10.1016/ S1369-5274(02)00358-2

Peyraud R, Dubiella U, Barbacci A, Genin S, Raffaele S, Roby D (2017) Advances on plant-pathogen interactions from molecular toward systems biology perspectives. Plant J 90:720–737. https://doi.org/10.1111/tpj.13429

Preston GM, Studholme DJ, Caldelari I (2005) Profiling the secretomes of plant pathogenic Proteobacteria. FEMS Microbiol Rev 29:331–360. https://doi.org/ 10.1016/j.fmrre.2005.01.001

Prior P, Ailloud F, Dalsing BL, Remenant B, Sanchez B, Allen C (2016) Genomic and proteomic evidence supporting the division of the plant pathogen *Ralstonia solanacearum* into three species. BMC Genomics 17:90. https://doi.org/10.1186/ s12864-016-2413-z

Remenant B, Coupat-Goutaland B, Guidot A, Cellier G, Wicker E, Allen C, Prior P (2010) Genomes of three tomato pathogens within the *Ralstonia solanacearum* species complex

reveal significant evolutionary divergence. BMC Genomics 11: 379. https://doi.org/10.1186/1471-2164-11-379

Safni I, Cleenwerck I, Paul De Vos, 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*. Int J Syst Evol Microbiol 64:3087–3103. https://doi.org/10.1099/ijs.0.066712-0

Salanoubat M, Genin S, Artiguenave F, Gouzy J, Mangenot S, Arlat M, Boucher C A (2002) Genome sequence of the plant pathogen *Ralstonia solanacearum*. Nature 415:497. http://dx.doi.org/10.1038/415497a

Schneider DJ, Collmer A (2010) Studying Plant-Pathogen Interactions in the Genomics Era: Beyond Molecular Koch's Postulates to Systems Biology. Annu Rev Phytopathol 48:457–479. https://doi.org/10.1146/annurev-phyto-073009-114411

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

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M (2006) ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34: D32–D36. https://doi.org/10.1093/nar/gkj014

Siguier P, Gourbeyre E, Chandler, M. (2014). Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev, 38(5), 865–891 http://dx.doi.org/10.1111/1574-6976.12067

Siguier P, Gourbeyre E, Varani A, Ton-Hoang B, Chandler M (2015) Everyman's Guide to Bacterial Insertion Sequences. Microbiol Spectr 3(2):550–590. http://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.DNA3-0030-2014

Stapley J, Santure AW, Dennis SR. (2015) Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. Mol Ecol 24(9): 2241–2252. https://doi.org/10.1111/mec.13089

Urban M, Cuzick A, Rutherford K, Irvine A, Pedro H, Pant R, Hammond-Kosack KE (2017) PHI-base: a new interface and further additions for the multi-species pathogen–host interactions database. Nucleic Acids Res 45: D604–D610. https://doi.org/10.1093/nar/gkw1089

Vandecraen J, Chandler M, Aertsen A, Van Houdt R (2017) The impact of insertion sequences on bacterial genome plasticity and adaptability. Crit Rev Microbiol 43(6): 709–730. https://doi.org/10.1080/1040841X.2017.1303661

Table 1. Characteristics of IS elements found in the *Ralstonia* Species Complex and *Ralstonia* spp. genomes

IS Family	IS Length (bp)	Number of copies		DR	IR	Most relevant IS
		Chromosome	Megaplasmid			
IS21	1890	84	54	-	5′TGGGTCAGGTTTACTTCGGCGGGTGGGTCAGTTTTACATCGGCGCTAACA 3′ 3′TGTCAACGCCGACCGAATTCTGACCCGCCTTTCGGTGTTTCGCCGAAGTA 5′	ISRs019
IS256	1330	68	32	CGACGGAAAG	5′CCTGTCCGCATTTTTGTGGGCGAGGCGGTTAAAAAG 3′ 3′CTTTTTAACCGCCTCGCCCACAAAAATGCGGACAGG 5′	ISBcen18
IS3	1300	347	149	AGCTG	5′TGGAGCGGCCCCCTGAATCTCCGGACACAGTCACCCACTTAAACTACGGG 3′ 3′ATCATCAACTCAAGTGTGAGGCGGTGTCCGGAGGTACGGGGGCAAATCCA 5′	ISRs010
IS4	1460	12	10	AGGCTAAGAGC	5′ACTGCCCGCGTACTCAAAAAGGCCCGGGCGGAAAGACCTTGAAAGGGGTG 3′ 3′CACCCCTTTCAAGGTCTTGTCGCACAGAAGTGCCGTATTCTTCCGCTCGC 5′	ISRso13
IS5	1180	909	1035	TTAG	5′GCGAACCTTCCTTCGACTGCCCGCGCTTGCGGGAGTTTTGAACACTATCC 3′ 3′GGATGGTGTTCAGAACTCATTCCGTCGAACTGAGTCATCGAAGCGAAGCG 5′	IS1405
IS630	1170	20	5	CTAG	5′TGTCCTGTTCCGTTGTTAACTGGTCTAAGTTTACATAGGATGTTGGCATC 3′ 3′ATCCTCTCCAAGCTCAGCAGACTTTGCGAACGAATTAGCGGAACGGGACA 5′	ISRs05
IS701	1480	19	45	CTAT	5′TCTTACTGTGTCATAAAAAGTAGCGCATTATATGTACTTCCTAATTCAAG 3′ 3′TGCCCATGCTGTGCGTCGGGCTGATCAATCACATTTGTGACACAGTAAGA 5′	ISRso17
ISL3	1320	20	14	TAAAAAAT/T	5′GGTTCTTCGAAGAATTCCGTGAGCCTGAACTGGCACATTCCGTATGCTGA 3′ 3′ACTTCTTCCTCAAGATCCGCGCCGCGTTCCCCGGAATTCCTCGATGAACC 5′	ISRs015
IS200/IS605	1605	4	0	TGAC	5′TACATCGAACAACAGCAGACCCCACACTGAAAAACCACAGCAAGGACGGCTACGCCGT 3′ 3′TTGACATCCTCCCCGACCTGTGAGCCGCCCGAGTCGCTTCGGTGGGTTCCTGATGCTGACG5′	ISAba30
Tn3	2956	17	32	-	-	ISPa38
IS1595	1100	0	18	ATATCCAT	-	ISRama1
IS110	1200	17	10	-	-	ISBma3
IS1182	1500	43	3	ACGCCACCCGAAGGTG GCGTT	-	ISBma2

DR: Direct Repeat / IR: Inverted Repeat

Figure legends

Figure 1. IS elements in the frame of *Ralstonia* spp. replicon. (A) Plot of the size of the IS elements in comparison of their hosts 'genome size. Shaded regions indicate the 95% confidence interval. According with the Pearson correlation coefficient, ISs plots have no linear relationship. (B) Dot plot representing the distribution of IS families in the chromosome (blue) and megaplasmid (red).

Figure 2. Phylogeny of the 64 distinct IS elements from the *Ralstonia* spp. reveling clusters (highlighted colored) according to IS family. The phylogenetic tree was generated with the Maximum Likelihood method using MEGA X software (1000 bootstrap replications) and the algorithm GTR+G (General Time-Reversible + Gamma distribution) as substitution model. Tree was visualized and annotated using iTOL.

Figure 3. Representation IS family's distribution in a RSSC phylogenetic context based in the 16S rRNA gene. The phylogenetic tree was generated with the Maximum Likelihood method using MEGA X software (1000 replications) and the algorithm TN93+G+I (Tamura-Nei + Gamma distribution + Invariable) as substitution model. Tree was visualized and annotated using iTOL.

Figure 4. Schematic representation of four Tn3 transposon located in *R. pseudosolanacearum* strains RSSCM and HA4-1. Genes are indicated by colored boxes, with the direction of transcription shown by the arrowheads. Transposition-related genes, passenger genes and terminal inverted repeats are shown in legend description.

Figure 5. Insertions of ISs in the intergenic regions of virulence genes. (A) Dot plot showing the insertion of this elements in three classes across the chromosome and megaplasmid sequence. (B) Schematic representation of three examples, representing the classes, mapped in two strains of RSSC.

Figure 6. Mauve alignment of the three *R. pseudosolanacearum* genome revealing numerous internal rearrangements for the strains SEPPX05. Colored blocks represent co-linear blocks. Multiple genome alignments were performed by the Mauve software.

Supplementary information's legends

Supplementary Figure S1. Schematic representation of *avr* gene overlapped by the terminal inverted repeat of the IS5/IS1182 element and the sequence alignment shown a putative completeness of *avr* protein (WP_054277986.1).

Supplementary Figure S2. Calculation of d_N/d_S ratio by site in transposase near to *avr* genes using JCoDA. The red line represents the score of d_N/d_S by site. Maximum Likelihood phylogenetic tree (100 replicates) was generated by aligned transposase protein sequences.

Supplementary Figure S3. Evidence of IS elements activity in *R. solanacearum* isolates. (A) Seven *R. solanacearum* were isolated from soil samples in the state of Minas Gerais and Distrito Federal. (B) IS1021 and IS010 were used as probes. Arrows indicated the primers in the direction forward (F) and reverse (R). (C) Southern hybridization membrane for the probe IS1021 and IS010.

Fig. 1

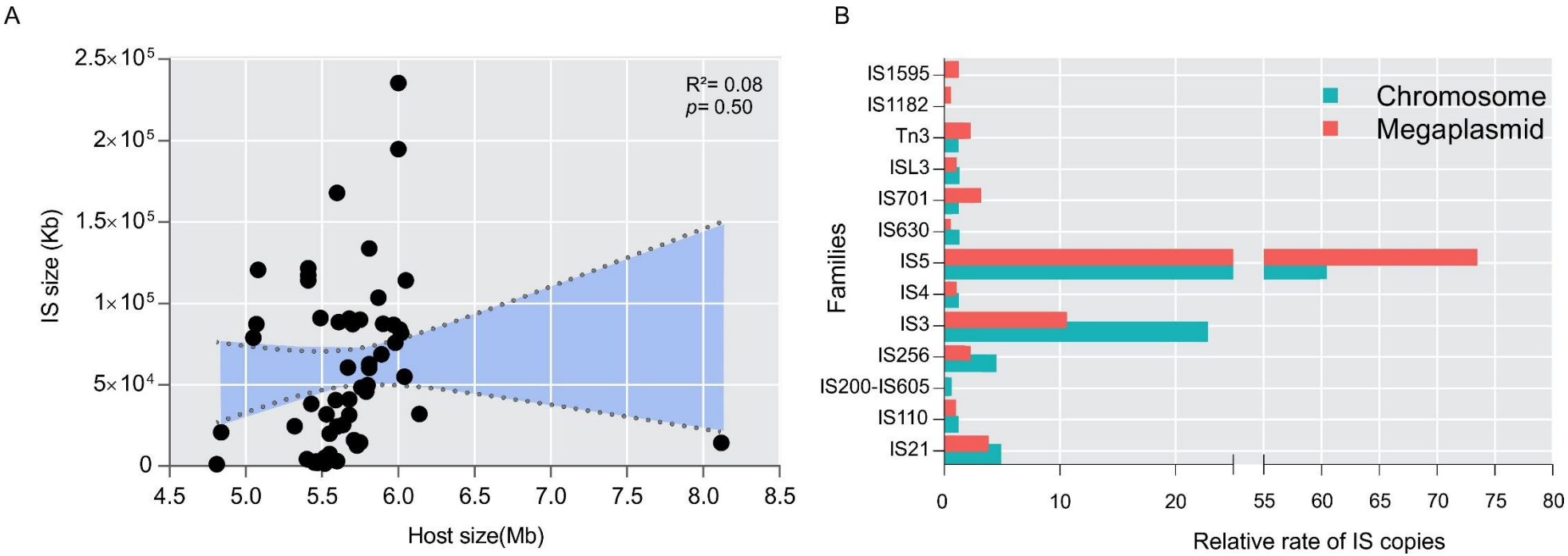

A
2.5×10^5
2×10^5
1.5×10^5
1×10^5
5×10^4
0
IS size (Kb)
R²= 0.08
p= 0.50
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
Host size(Mb)
B
IS1595
IS1182
Tn3
ISL3
IS701
IS630
IS5
IS4
IS3
IS256
IS200-IS605
IS110
IS21
Families
Chromosome
Megaplasmid
0
10
20
55
60
65
70
75
80
Relative rate of IS copies

Fig. 2

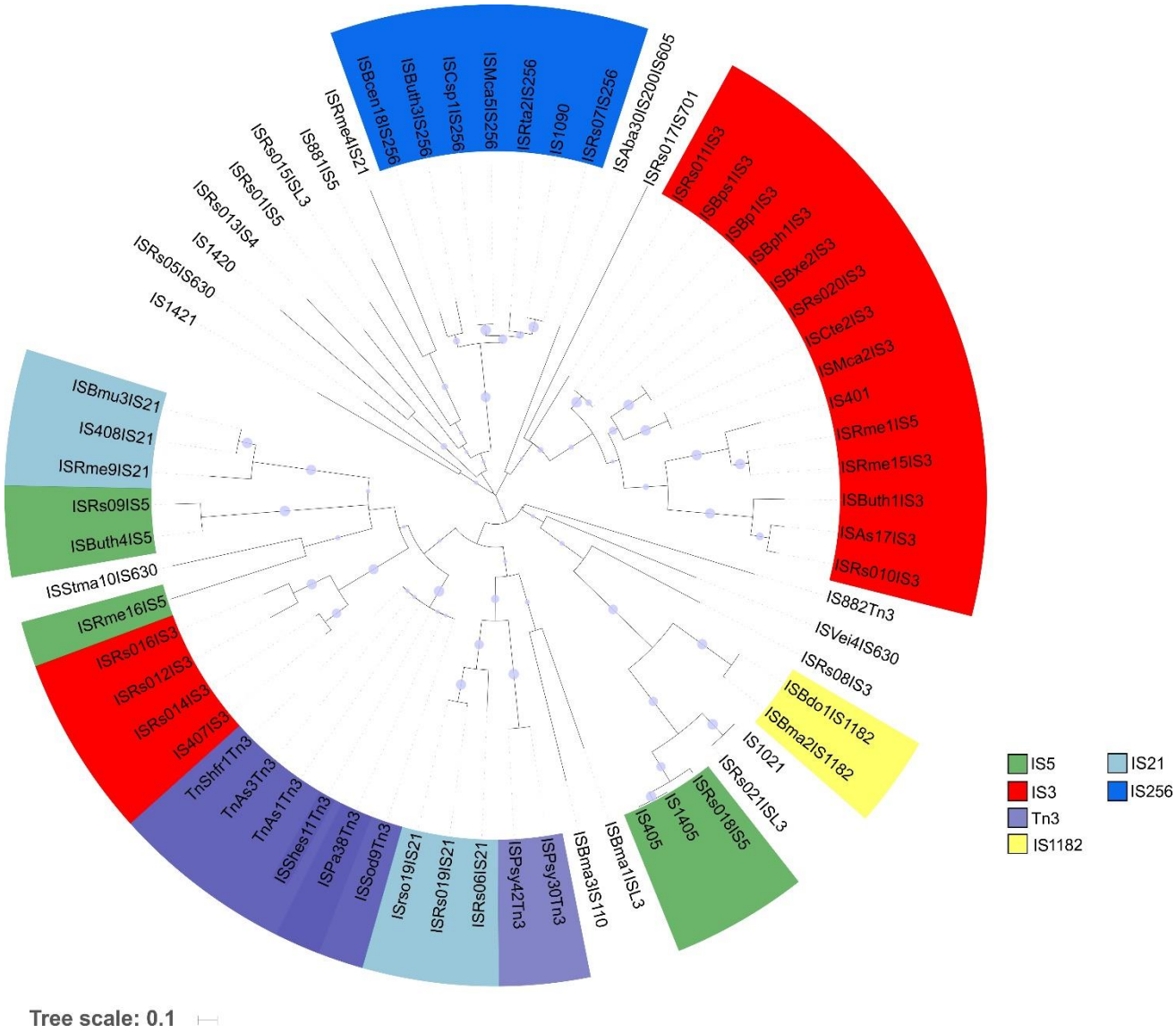

ISBcen18IS256
ISButh3IS256
ISCsp11IS256
ISMca5IS256
ISRta2IS256
IS1090
ISRs07IS256
ISAba30IS200IS605
ISRs017IS701
ISRs011IS3
ISBps1IS3
ISBp1IS3
ISBph1IS3
ISBxe2IS3
ISRs020IS3
ISCte2IS3
ISMca2IS3
IS401
ISRme1IS5
ISRme15IS3
ISButh1IS3
ISAs17IS3
ISRs010IS3
IS882Tn3
ISVei4IS630
ISRs08IS3
ISBdo1IS1182
ISBma2IS1182
IS1021
ISRs021ISL3
ISRs018IS5
IS1405
IS405
ISBma1ISL3
ISBma3IS110
ISPsy30Tn3
ISPsy42Tn3
ISRso6ISL21
ISRso10IS21
ISRso19IS21
ISSodgTn3
ISPa38Tn3
ISShes11Tn3
TnAs1Tn3
TnAs3Tn3
TnShfr1Tn3
IS407IS3
ISRs014IS3
ISRs012IS3
ISRs016IS3
ISRme16IS5
ISStma10IS630
ISButh4IS5
ISRs09IS5
ISRme9IS21
IS408IS21
ISBmu3IS21
IS1421
ISRs05IS630
IS1420
ISRs013IS4
ISRs011IS5
ISRs015ISL3
IS981IS5
ISRme4IS21
IS5
IS3
Tn3
IS1182
IS21
IS256
Tree scale: 0.1

Fig. 3

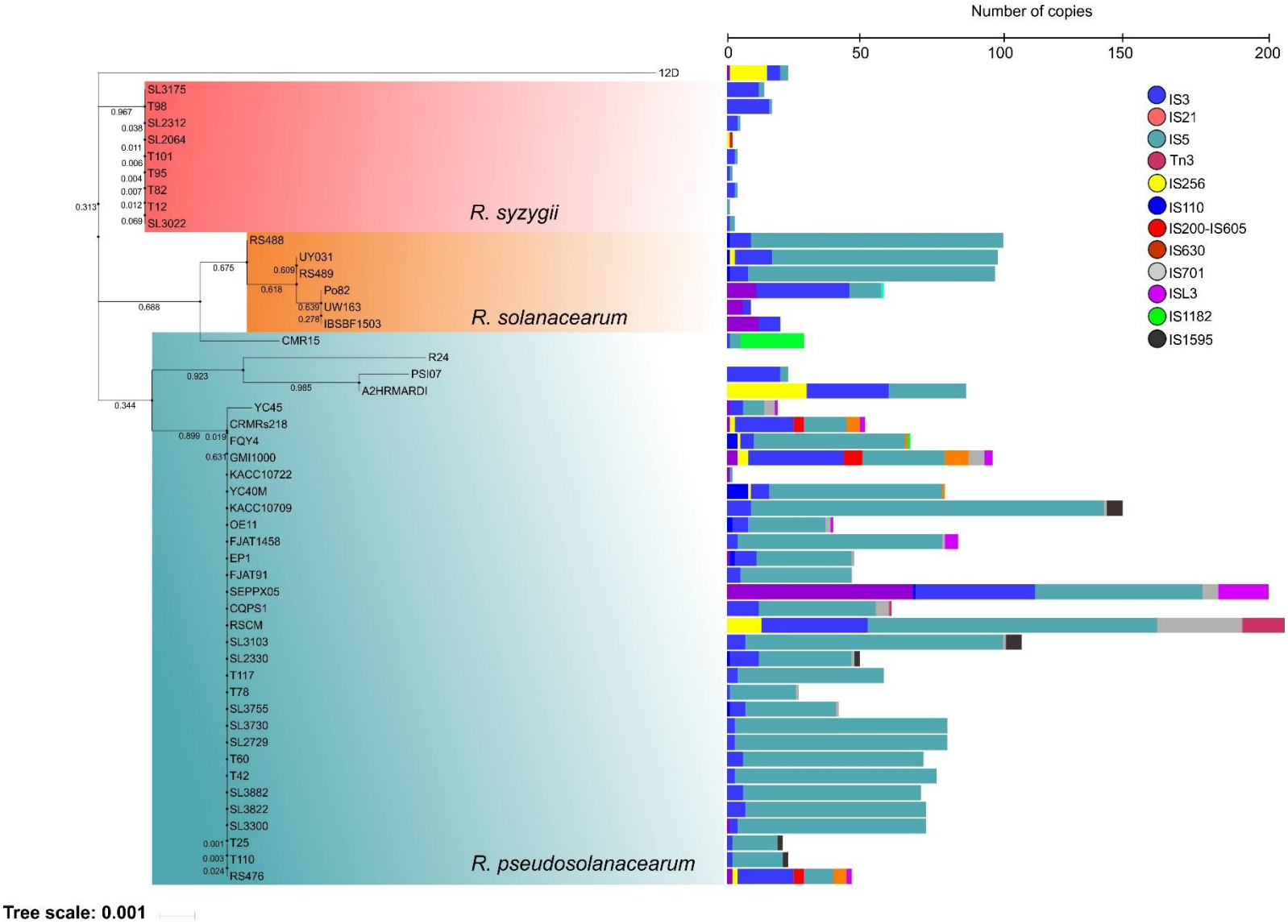

Number of copies
0
50
100
150
200
12D
SL3175
T98
SL2312
SL2064
T101
T95
T82
T12
SL3022
R. syzygii
RS488
UY031
RS489
Po82
UW163
IBSBF1503
R. solanacearum
CMR15
R24
PSI07
A2HRMARDI
YC45
CRMRs218
FQY4
GMI1000
KACC10722
YC40M
KACC10709
OE11
FJAT1458
EP1
FJAT91
SEPPX05
CQPS1
RSCM
SL3103
SL2330
T117
T78
SL3755
SL3730
SL2729
T60
T42
SL3882
SL3822
SL3300
T25
T110
RS476
R. pseudosolanacearum
IS3
IS21
IS5
Tn3
IS256
IS110
IS200-IS605
IS630
IS701
ISL3
IS1182
IS1595
Tree scale: 0.001

Fig. 4

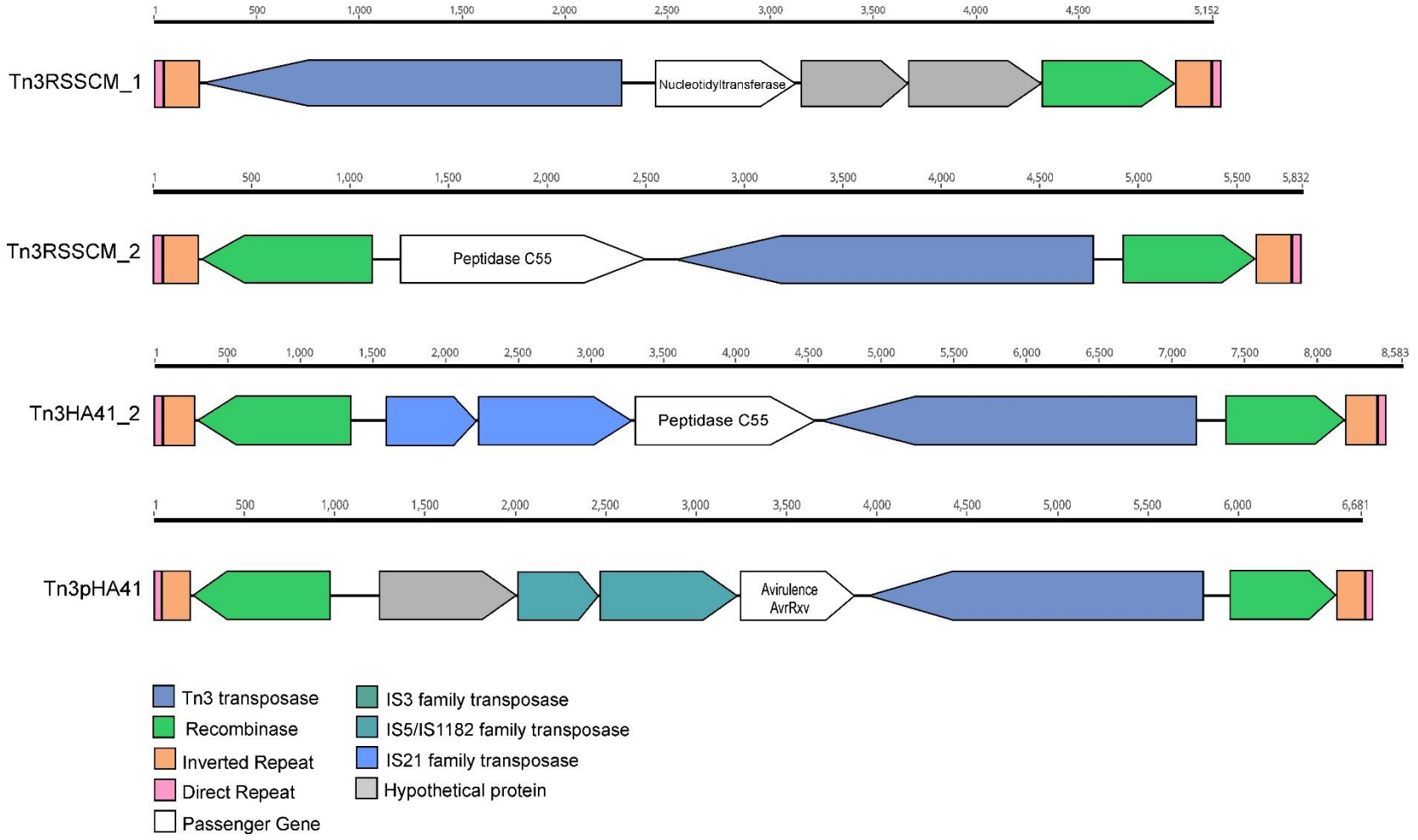

1
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,152
Tn3RSSCM_1
Nucleotidyltransferase
1
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
5,500
5,832
Tn3RSSCM_2
Peptidase C55
1
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
5,500
6,000
6,500
7,000
7,500
8,000
8,583
Tn3HA41_2
Peptidase C55
1
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
5,500
6,000
6,681
Tn3pHA41
Avirulence
AvrRxv
Tn3 transposase
Recombinase
Inverted Repeat
Direct Repeat
Passenger Gene
IS3 family transposase
IS5/IS1182 family transposase
IS21 family transposase
Hypothetical protein

Fig. 5

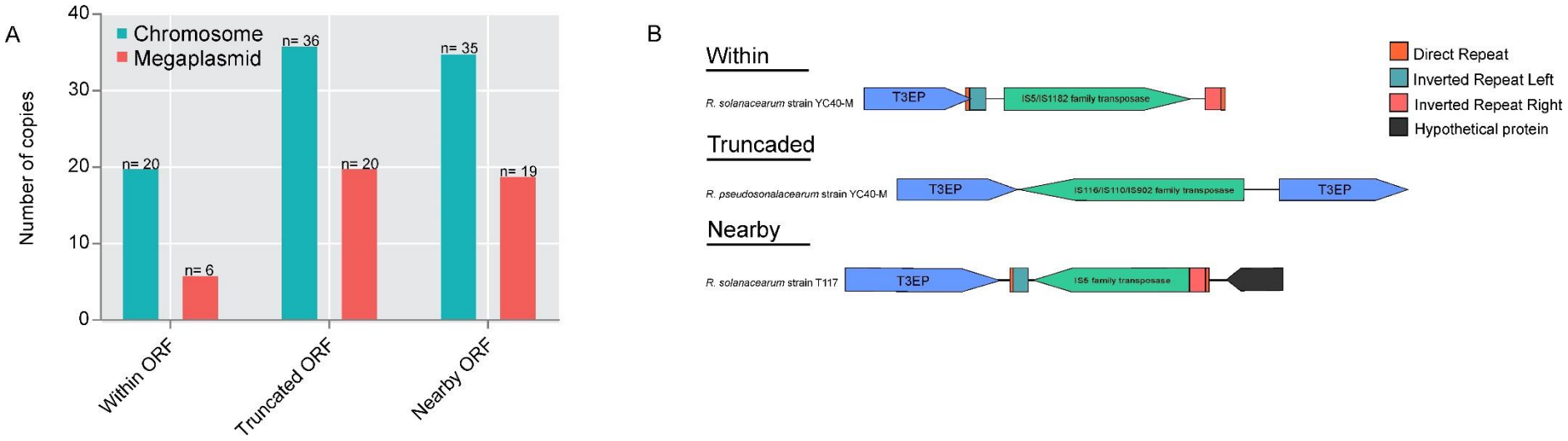
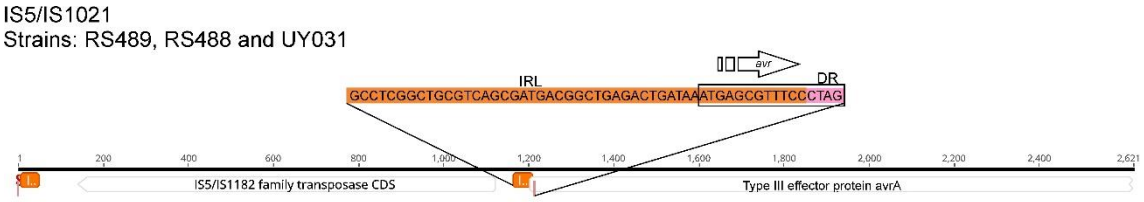
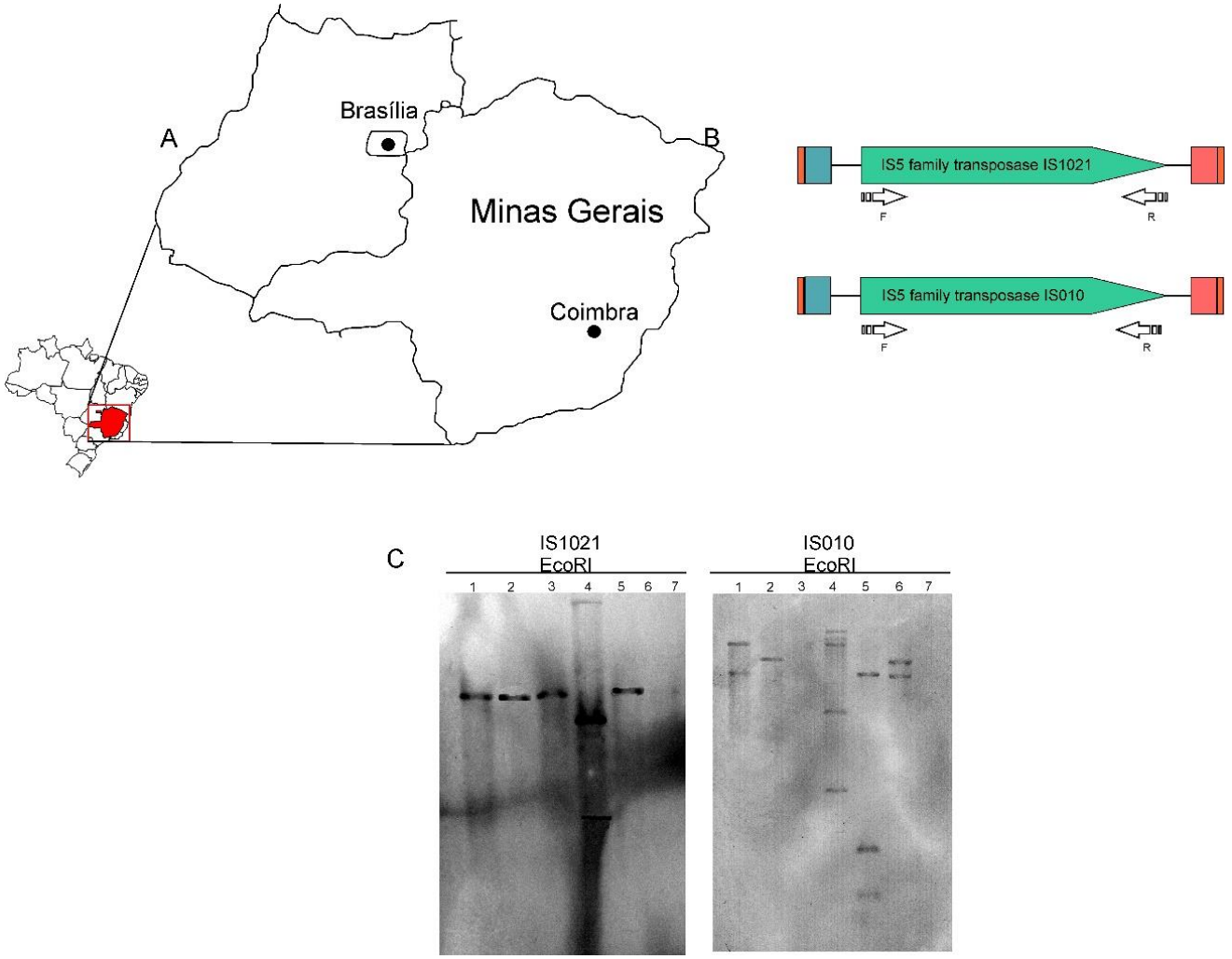

A
Chromosome
Megaplasmid
Number of copies
40
30
20
10
0
n= 20
n= 6
n= 36
n= 20
n= 35
n= 19
Within ORF
Truncated ORF
Nearby ORF
B
Within
R. solanacearum strain YC40-M
T3EP
IS5/IS1182 family transposase
Truncaded
R. pseudosonalacearum strain YC40-M
T3EP
IS116/IS110/IS902 family transposase
T3EP
Nearby
R. solanacearum strain T117
T3EP
IS5 family transposase
Direct Repeat
Inverted Repeat Left
Inverted Repeat Right
Hypothetical protein

Fig. S1




```
WP_054277986.1      MSVSLDSPLLPQAPGHTASAPKPLPVGRRSPIQSELRSKLSQAPHSSQPPLLKPPVKLTP 60
IS                  MSVSLDSPPESQAPGHTASAPKPLPVGRRSPIQSELRSKLSQAPHSSQPPLLKPPVKLTP 60
                    ******** * **************************************************

WP_054277986.1      AFQEKLKSPQETVPKTHEAGGSAKAAPNASDVKPVVVLDWDDCLRYEKGMNYQLVHNALV 120
IS                  ALQEKLKSPQETVPKTQEAGGSAKAAPNASDVKPVVVLDWDDCLRYEKGMNYQLVHNALV 120
                    *:***************:********************************************

WP_054277986.1      IAARMHAQSLPELAEAVDRLHARMQRGEQPGDGDPLIMKSQEDCENYLAANSGIYKRGLV 180
IS                  IAARMHAQSLPKLGEAVDRLHARMQRGEQPGDGDPLIMKSLEDCSNYLAATPGIYKRGLV 180
                    ***********:*.************************ ***.*****. ********

WP_054277986.1      QDFVRTMLPDIDSSMAEAITDAAYTQCALEYNRLMAAPDQKESWRQDLPFPDVRIALMPG 240
IS                  QDFVRTMLPDIESSMAEAITEAAYTQCVLEYSRLMAAPEQKESWRQDLPFPDVRIALMPG 240
                    ***********:********:******.***.******.********************

WP_054277986.1      ARELLDTSRAEGSPVLLISNRAHSDLQKEVRYLGMQQDFDVVSGAPTITDKKSKTTPSPM 300
IS                  ARELLDTSRADGSPVLLISNRAHSDLQKEVHYLGMQQDFDVVSGCPTITRNKSKTTPSPM 300
                    **********:*******************:*************:**** :*********

WP_054277986.1      PQTLEQQLIAALKGD---DDGALRTALEAASPYAHPNTTSVKQTDHKPRDTRLLNGLERL 357
IS                  PQTLEQQLIAALQGDDDDDDGALRAALEAASPYAHPNTTSVTQTDHKPRDTRLQDGLQRL 360
                    ***********:**    ******:***************.********** :**:**

WP_054277986.1      SVPSDAPIVLYGDQESDISQAATLAAAGRHVEGVLIDPGRDDVGQQIGIQGIPTRVIGSL 417
IS                  SVPSDAPIVLYGDQPSDISQAATLAAAGRHVEGVLIDPGRNDVGQQVDIQGIPTRVIGSL 420
                    ************* ************************:*****:.***********

WP_054277986.1      TDADAPWKTAAASRASLPSALAPIPQALRPGSAGPVQDGHLFSVTRSVPFILSRPGQAAP 477
IS                  TDADAPWKTAAASRASLPSALAPIPQALRPGSAGPVQGGHLFSVTRSVPFILSRPGQAAP 480
                    ************************************.**********************

WP_054277986.1      SLPVAAGGGVVYSRDKQALGGSDVWMVAWMDGNEHLVKTGLPRSLLGQIVSCWATGQEAD 537
IS                  SLPVAAGGGVVYSRDKQALGGSDVGMVAWMDGNEHLVKTGVSRGLLGQILSCWATGQEAD 540
                    ********************** ***************: *.*****:**********

WP_054277986.1      GDKRVGYLIGIWARLSGNDENKDVLETQLKTALRTFIQSENGSALFEKSKSEALDLADVS 597
IS                  GDKRAGYLIGNWARLSGNDENKDVLETQLKTALRTFIQSENGSTLFEKSKSEALDLADVN 600
                    ****.***** ********************************:***************.

WP_054277986.1      AIHRALTEACPELKNPLGMPALFELVNGAASQALANALQRTYLPEDQVPDATLLVSEDNV 657
IS                  AIHRALTAACPELKNPLGMPAIFDVVNGAASQALANALQRTYLPEGQVPDATLLVSEDNV 660
                    ******* ************:*::********************.**************

WP_054277986.1      LLASRLLPDAVPMDDFLTRSFLPQGVSLQDAKLAAARIKANPEGDSAELRSARELIARLC 717
IS                  LLASRLLPNAVPMDAFLTRSFLPQGVSLQDAKLAAARIKANPEGDSAELRSARELIARLC 720
                    ********:***** ********************************************

WP_054277986.1      DPQQLEAGRAGLTQALAAQGMDGFFASVLARLTIGENHKDLGPDNMLLVRGADGRNKAVN 777
IS                  APQQLEAGRAGLTQALAAQGMDGFFASVLARLTIGENHKDLGPDNMLLVRGADGRNKAVN 780
                     ********************************************************
```

Fig. S2


A
Brasília
B
Minas Gerais
Coimbra
IS5 family transposase IS1021
F
R
IS5 family transposase IS010
F
R
C
IS1021
EcoRI
1 2 3 4 5 6 7
IS010
EcoRI
1 2 3 4 5 6 7

GENERAL CONCLUSIONS

- Overall, our results provide a comprehensive identification of GIs and ICEs in the *R. solanacearum* Species Complex, these elements constituting an important fraction of the *Ralstonia* genome and which are phylogenetically related, which could indicate their ancestral acquisition. Together the GIs and ICEs have a repertoire of genes with potential impact in the *Ralstonia* fitness and pathogenicity, which may enable the bacteria to evolve rapidly and infect a wide range of hosts.

- A great diversity of TEs in the representative RSSC strains and closely related *Ralstonia* spp. were identified. A unique set of IS families were found highly conserved across the strains, suggesting their ancestral acquisition. Novel Tn3 transposons sequences carrying genes related to antibiotic resistance and avirulence protein as passenger genes. Numerous internal rearrangements events with a subset being associated with IS were demonstrated, indicating the impact of these elements on RSSC genome evolution. We also mapped a numerous of IS elements interrupting avirulence genes, which provides evidence that IS may be one of the driving forces RSSC pathogenicity evolution.