농학박사학위논문

한국의 콩 불마름병균 *Xanthomonas citri* pv. *glycines*의 유전체 비교분석 및 레이스 동정

Comparative genome analysis and race identification of Korean isolates of *Xanthomonas citri* pv. *glycines* causing soybean bacterial pustule

2022년 2월

서울대학교 대학원

농생명공학부 식물미생물학전공

강 인 정

Comparative genome analysis and race identification of Korean isolates of *Xanthomonas citri* pv. *glycines* causing soybean bacterial pustule

A dissertation submitted in partial

fulfillment of the requirement for

the degree of

DOCTOR OF PHILOSOPHY

to the Faculty of

Department of Agricultural Biotechnology

at

SEOUL NATIONAL UNIVERSITY

by

In-Jeong Kang

February 2022

농학박사학위논문

한국의 콩 불마름병균 *Xanthomonas citri* pv. *glycines*의 유전체 비교분석 및 레이스 동정

지도교수 황 인 규

이 논문을 농학박사학위논문으로 제출함

2021년 12월

서울대학교 대학원
농생명공학부 식물미생물학전공

강 인 정

강인정의 박사 학위논문을 인준함

2022년 1월

위 원 장 _____ 김 국 형 (인)

부 위 원 장 _____ 황 인 규 (인)

위 원 _____ 이 상 기 (인)

위 원 _____ 김 진 우 (인)

위 원 _____ 문 제 선 (인)

A THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Comparative genome analysis and race identification of Korean isolates of *Xanthomonas citri* pv. *glycines* causing soybean bacterial pustule

UNDER THE DIRECTION OF DR. INGYU HWANG

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

BY

IN-JEONG KANG

PROGRAM IN PLANT MICROBIOLOGY
DEPARTMENT OF AGRICULTURAL BIOTECHNOLOGY

JANUARY 2022

APPROVED AS A QUALIFIED THESIS OF IN-JEONG KANG
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
BY THE COMMITTEE MEMBERS

CHAIRMAN _____Kook-Hyung Kim_

VICE CHAIRMAN _____Ingyu Hwang_

MEMBER _____Sangkee Rhee_

MEMBER _____Jinwoo Kim_

MEMBER _____Jae Sun Moon_

Comparative genome analysis and race identification of Korean isolates of *Xanthomonas citri* pv. *glycines* causing soybean bacterial pustule

In-Jeong Kang

ABSTRACT

Bacterial pustule caused by *Xanthomonas citri* pv. *glycines* (*Xcg*) is one of the most prevalent bacterial diseases of soybean [*Glycine max* (L.) Merr.] in Korea. In the first part of this study, the genes related to pathogenicity were analyzed based on sequence and structural similarity searches of genomic data of five Korean type strains and one strain from the United States, 8ra. The comparative analysis of candidate type III secretion system effector (T3SE) genes showed that the diversity and number of effectors in the five analyzed Korean strains were very similar, yet significantly differed from the 8ra strain. In the T3SEs, transcription activator-like effectors (TALEs) showed diverse repeat sizes and at least six forms of the repeat variable di-residue (RVD), and a phylogenetic tree based on the alignment of RVD sequences showed two distinct clusters with 17.5

repeats, suggesting that two distinct 17.5 RVD clusters have evolved to adapt *Xcg* to growth on distinct soybean cultivars. Additionally, the predicted effector binding elements of the TALEs fell into six groups and were strongly overlapping in sequence, suggesting evolving target specificity of the binding domains in soybean cultivars. Our findings reveal the variability and adaptability of T3SEs in the *Xcg* strains and enhance our understanding of *Xcg* pathogenicity in soybean.

In the second part of this study, the *Xcg* populations in Korea were evaluated based on *avrBs3* profiles. In particular, the study addressed if new *Xcg* groups have emerged since the six type strains were identified in 2008, and if such new groups represent new races of Korean *Xcg* isolates. The diversity of a nationwide collection of 106 *Xcg* isolates was characterized based on *avrBs3* banding patterns. The isolates fell into 11 groups, each represented by a type strain; only two of these were similar to isolates reported by the previous study. The diversity of *Xcg* strains increased and the dominant strains changed between 1999 and 2017, with three new type strains comprising 44% of the isolates examined in 2012 to 2017. In pathogenicity tests, Korean soybean cultivars, including the widely-grown cultivar, Daewon were susceptible to the 11 new type strains. The cultivar CNS, which carries the *rxp* resistance gene, was susceptible to most type strains, including two representing 83% of the Korean *Xcg* strains. In contrast, Williams 82, which

also carries *rxp*, showed resistance to at least five type strains. Collectively, these results suggest that Williams 82 has resistance loci in addition to *rxp*. The widespread distribution of *Xcg*, the high virulence of the current endemic strains, and the low resistance of most Korean soybean cultivars collectively favor widespread disease in Korea in years that are favorable to pustule development.

This study is the first to characterize TALEs and their effector binding elements at the whole genome level for Korean *Xcg* type strains and identify new races of Korean *Xcg* isolates and their pathogenicity against soybean cultivars. This study provides guidance in developing strategies to deploy soybean varieties resistant to bacterial pustule caused by *Xcg* in Korea.

KEY WORDS: *Xanthomonas citri* pv. *glycines,* bacterial pustule of soybean, comparative genome, new race, pathogenicity

Student number: 2016-38299

CONTENTS

LIST OF TABLES

Page

CHAPTER I

CHAPTER II

LIST OF FIGURES

Page

CHAPTER I

CHAPTER II

INTRODUCTION

Bacterial diseases occurring during the growing season of soybeans in Korea are reported to be bacterial pustule, bacterial blight, bacterial brown spot, and wildfire. Whereas bacterial blight is the most common bacterial disease of soybeans in USA, bacterial pustule and wildfire are the main diseases prevalent in Korea. Among these, bacterial pustule caused by *Xanthomonas citri* pv. *glycines* (*Xcg*) (Constantin et al., 2016) is the most problematic, widespread bacterial disease of soybean [*Glycine max* (L.) Merr.] in Korea. The causal agent was previously classified as *Xanthomonas axonopodis* pv. *glycines* (Vauterin et al. 1995) and, prior to that, as *Xanthomonas campestris* pv. *glycines* (Dye, 1978). *Xcg* cells are Gram-negative, aerobic, and motile with a single polar flagellum; they are rod-shaped, with a size range of 0.5~0.9×1.4~2.3 μm in width and length, respectively. Colonies on tryptic soy agar are small, circular, and smooth, and have entire margins. The bacterium is slow-growing in culture and is known to produce auxins, bacteriocins and exopolysaccharides. It also liquefies gelatin, produces acid but not gas from sucrose, and rapidly hydrolyzes starch. The optimal temperature for growth is 30~33℃.

The etiology of the bacterial pustule disease is well understood. *Xcg* enters the plant through natural openings and wounds and multiples

intercellularly. *Xcg* invades and multiplies within the apoplast, causing localized leaf spots or leaf streaks (Boch et al., 2010). Bacterial pustule symptoms include small yellowish spots that later develop a reddish-brown color and then slightly raised pustules in the center of the lesions (Groth and Braun, 1998); these lesions usually occur on the lower leaf surfaces. The lesions vary from specks to large, irregularly shape, mottled-brown areas, which develop when the lesions coalesce. The leaf spots can form without developing pustules. Infections are more common on younger leaves, which may reflect that they are more susceptible than older leaves.

Bacterial pustule can significantly impact soybean yields. Bacterial pustule has been reported in the most soybean-growing area of the world in which warm temperatures and sufficient moisture prevails during the growing season. The disease may cause premature defoliation, which may decrease yield by reducing seed size and number. The effect of bacterial pustule on the growth and development of soybean largely depends on infection levels, with high infection levels decreasing yields by 20 to 60% following natural or artificial infection (Hokawat et al., 1991; Shukla, 1994; Hong et al., 2011). In Korea, bacterial pustule occurs nationwide and can be observed any time from July to September. A survey of bacterial pustule in the 1990s estimated that 86% of Korean soybean cultivation fields appeared to have bacterial pustule (Lee, 1999). In 2005 and 2006, a similar level of disease incidence, 89.7%,

was observed when surveyed in the southern regions of Korea (Hong et al., 2010).

Xanthomonas pathogens such as *Xcg* employ various pathogenicity factors during disease establishment. Successful infection and bacterial multiplication in the host tissue often depend on secreted virulence factors, including adhesins, polysaccharides, and degradative enzymes. One of the key pathogenicity factors is the type III secretion system (T3SS), which injects effector proteins into the host cell cytosol to manipulate plant cellular processes; for example, these effectors may suppress the host's basal defenses, thus benefitting the pathogen (Büttner and Bonas, 2010). *Xanthomonas spp.* employ a T3SS to inject an especially large number of effector proteins into the cytosol of host cells. The T3SS in *Xcg,* called a Hrp T3SS, is encoded by hypersensitive reaction and pathogenicity (*hrp*) and *hrp*-conserved (*hrc*) genes, which often occur in pathogenicity islands in the genome (Alfano et al., 2000). Kim et al. (2003) identified HR-eliciting proteins and characterized the virulence genes contained in the Hrp T3SS in *Xcg*. *Xanthomonas* strains typically express 20-40 type III effectors including transcription activator-like effectors (TALEs). TALEs are DNA-binding proteins that reprogram the expression of specific genes in the host, boosting the expression of susceptibility genes or activating the expression of resistance genes, depending on the host genotype (Bogdanove et al., 2010). The transcription

activator-like effector (TALE) family is a distinct family of type III protein effectors, which includes members with cognate susceptibility and/or resistance genes. The diversity of type III effectors reflects the diversity in tissue and host specificity of members of the genus (White et al., 2009). The DNA binding domain of TALEs consists of highly repetitive consecutive repeats of a 33-35 highly conserved amino-acid sequence. The number of repeats present in TALEs varies with an average of 17 repeats and up to 30 repeats (Boch et al., 2009). The amino acids in positions 12 and 13 are called repeats variable di-residue (RVD), and the sequence of RVDs determines the TALE target specificity, with each RVD binding to a specific target sequence in the host. Other candidate type III effectors beyond TALEs also affect virulence and disease symptomatology (White et al., 2009). For instance, both XopD and XopN of *X. campestris* pv. *vesicatoria* contribute to virulence and disease symptoms in tomato (Kim et al., 2008).

The genomic resources available for *Xcg* are improving. Draft genome sequences provide valuable information on gene content and enable genomic comparisons among species or even among strains. Recently, draft genome sequences became available for three *Xcg* strains, 8ra, 12-2, and EB08, and these were generated using PacBio long reads data (Carpenter et al., 2019). Such draft genome sequences support genome-wide studies of pathogenicity effectors and pathogen variability among *Xcg* strains in Korea, and may allow

us to understand the extent of variation among distinct races of *Xcg* in Korea. *Xcg* strains fall into distinct races and exhibit differences in their aggressiveness (Athinuwat et al., 2009; Hwang and Lim, 1998; Park et al., 2008; Kaewnum et al., 2005), which are likely associated with distinct profiles of effector genes. Park et al. (2008) reported that *Xcg* strains fall into at least six distinct groups based on their profiles of infection and aggressiveness on a collection of soybean cultivars, and the strains representing these groups varied in the number of *avrBs3* homologs in the genome based on hybridization to an *avrBs3* gene family probe. However, the extent of sequence variation among these *avrBs3* family genes in *Xcg* is not known.

Various approaches can be taken to control bacterial pustule. For example, planting pathogen-free seeds may reduce the occurrence of bacterial pustule because *Xcg* is seed-borne. Cultural practices such as rotating with a nonhost crop (e.g. maize) and avoiding field activities when foliage is wet may also reduce the occurrence of pustule by reducing the inoculum on host leaves. Delaying the planting of soybean cultivars may reduce the incidence of bacterial pustule because delayed planting might lead to avoiding the rainy season when bacterial pustule is intensively occurring in Korea at the flowering stage of soybean (Hong et al., 2012). Bactericides such as copper can be used to control bacterial pustule by directly killing the pathogen. The

most effective management approach, however, would be to use soybean varieties that are known to be resistant against bacterial pustule. Only one gene, *rxp* (resistance to *Xanthomonas phaseoli*), is currently known to confer resistance in soybean, and this resistance is only partial (Bernard, 1973). The *rxp* gene was originally found as a recessive resistance gene in the cultivar CNS (Hartwig and Lehman, 1951) and has been introduced into the cultivars Williams and Williams 82 that were developed by the United States Department of Agriculture (Hartwig and Lehman, 1951; Narvel et al., 2001). The *rxp* gene has not yet been cloned, and the availability of bacterial pustule-resistant commercial varieties is still limited. To date, the incorporation of *rxp* into Korean soybean cultivars has not been reported, and the *Xcg* avirulence gene that triggers *rxp*-mediated resistance has not been identified.

Bacterial pustule is a continuing problem in soybean despite that commercial soybean varieties with at least partial resistance to bacterial pustule have been developed for other areas of the world. Therefore, the objectives of this study were to track the incidence and severity of bacterial pustule by surveying the current distribution of bacterial pustule throughout Korea, and to evaluate the virulence of *Xcg* in Korea against cultivars that were previously used for characterizing virulence to assess potential changes in *Xcg* populations in Korea. This study is the first to evaluate the diversity of Korean *Xcg* strains since 2008. This knowledge is essential for developing

resistant soybean varieties and understanding the diversity of the native Korean *Xcg* strains to help ensure the development of durable host resistance. This study provides an important basis for further studies of *Xcg*-plant interactions in Korea and provides an excellent tool to study *Xcg* genetic determinants related to pathogenicity.

LITERATURE CITED

Alfano, J. R., Charkowski, A. O., Deng, W. L., Badel, J. L., Petnicki-Ocwieja, T., Van Dijk, K. and Collmer, A. 2000. The *Pseudomonas syringae* Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants. *PNAS*. 97: 4856-4861.

Athinuwat, D., Prathuangwong, S., Cursino, L., Burr, T. 2009. *Xanthomonas axonopodis* pv. *glycines* Soybean Cultivar Virulence Specificity Is Determined by *avrBs3* Homolog a*vrXg1*. *Phytopathol*. 99: 996-1004.

Bernard, R. L. and Weiss, M. U. 1973. Qualitative genetics. In: *Soybeans: Improvement, production, and uses*, ed. by B. E. Caldwell, pp. 117-164. American Society of Agronomy, Madison, WI, USA.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S. and Bonas, U. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*. 326: 1509-1512.

Boch, J. and Bonas, U. 2010. *Xanthomonas* AvrBs3 family-type III effectors: Discovery and function. *Annu. Rev. Phytopathol*. 48: 419-436.

Bogdanove, A. J., Schornack, S. and Lahaye, T. 2010. TAL effectors: finding plant genes for disease and defense. *Curr. Opin. Plant Biol.* 13: 394-401.

Büttner, D. and Bonas, U. 2010. Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol. Rev.* 34: 107-133.

Carpenter, S. C., Kladsuwan, L., Han, S. W., Prathuangwong, S. and Bogdanove, A. J. 2019. Complete genome sequences of *Xanthomonas axonopodis* pv. *glycines* isolates from the United States and Thailand reveal conserved transcription activator-like effectors. *Genome Biol. Evol*. 11: 1380-1384.

Constantin, E. C., Cleenwerck, I., Maes, M., Baeyen, S., Van Malderghem, C., De Vos, P. and Cottyn, B. 2016. Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. *Plant Pathol*. 65: 792-806.

Dye, D.W. 1978. Genus *Xanthomonas* Dowson 1939. In: Young, J.M., Dye, D.W., Bradbury, J.F., Panagopoulos, C.G., and Robbs, C.F. A proposed nomenclature and classification for plant pathogenic bacteria. *New Zealand J. Agricultural Res*. 21: 153-177.

Groth, D. E. and Braun, E. J. 1989. Survival, seed transmission and epiphytic development of *Xanthomonas campestris* pv. *glycines* in the north-central

United States. *Plant Dis.* 73: 326-330.

Hartwig, E. E. and Lehman, S. G. 1951. Inheritance of resistance to the bacterial pustule disease in Soybean. *J. Agron.* 43.

Hong, S. J., Kim, Y. K., Jee, H. J., Lee, B. C., Yoon, Y. N. and Park, S. T. 2010. Selection of bactericides for controlling soybean bacterial pustule. *Res. Plant Dis.* 16: 266-273. (In Korean)

Hong, S. J., Kim, Y. K., Jee, H. J., Shim, C. K., Kim, M. J., Park, J. H. and Lee, B. C. 2011. Influence of disease severity of bacterial pustule caused by *Xanthomonas axonopodis* pv. *glycines* on soybean yield. *Res. Plant Dis.* 17: 317-325. (In Korean)

Hong, J. K., Sung, C. H., Kim, D. K., Yun, H. T., Jung, W. and Kim, K. D. 2012. Differential effect of delayed planting on soybean cultivars varying in susceptibility to bacterial pustule and wildfire in Korea. *Crop Prot*. 42: 244-249.

Hokawat, S. and Rudolph, K. 1991. Variation in pathogenicity and virulence of strains of *Xanthomonas campestris* pv. *glycines*, the incitant of bacterial pustule of soybean. *J. Phytopathol*. 131: 73-83.

Hwang, I., Lim, S.M. 1998. Pathogenic variability in isolates of *Xanthomonas campestris* pv. *glycines*. *Korean J. Plant Pathol.* 14:19-22.

Kaewnum, S., Prathuangwong, S., Burr, T.J. 2005. Aggressiveness of *Xanthomonas axonopodis* pv. *glycines* isolates to soybean and hypersensitivity responses by other plants. *Plant Pathol.* 54: 409-415.

Kim, J. G., Park, B. K., Yoo, C. H., Jeon, E., Oh, J. and Hwang, I. 2003. Characterization of the *Xanthomonas axonopodis* pv. *glycines* Hrp pathogenicity island. *J. Bacteriol. Res.* 185: 3155-3166.

Kim, J. G., Taylor, K. W., Hotson, A., Keegan, M., Schmelz, E. A. and Mudgett, M. B. 2008. XopD SUMO protease affects host transcription, promotes pathogen growth, and delays symptom development in *Xanthomonas*-infected tomato leaves. *The Plant Cell*. 20: 1915-1929.

Lee, S. D. 1999. Occurrence and characterization of major plant bacterial diseases in Korea. Ph.D. thesis. Seoul National University, Seoul, Korea.

Narvel, J.M., Jakkula, L.R., Phillips, D.V., Wang, T., Lee, S.H. and Boerma, H.R. 2001. Molecular mapping of *rxp* conditioning reaction to bacterial pustule in soybean. *J. Hered.* 92: 267-270.

Park, H.J., Han, S.W., Oh, C., Lee, S., Ra, D., Lee, S.H., Heu, S. 2008. Avirulence gene diversity of *Xanthomonas axonopodis* pv. *glycines* isolated in Korea. *J. Microbiol. Biotechnol.* 18: 1500-1509.

Shukla, A. K. 1994. Pilot estimation studies of soybean (*Glycine max*) yield losses by various levels of bacterial pustule (*Xanthomonas campestris* pv. *glycines*) infection. *Int. J. Pest Manage*. 40: 249-251.

Vauterin, L., Hoste, B., Kersters, K. and Swings, J. 1995. Reclassification of *Xanthomonas*. *Int. J. Syst. Evol.* 45: 472-489.

White, F. F., Potnis, N., Jones, J. B. and Koebnik, R. 2009. The type III effectors of *Xanthomonas*. *Mol. Plant Pathol*. 10: 749-766.

CHAPTER I

Pan-Genome Analysis of Effectors in Korean Strains of the Soybean Pathogen *Xanthomonas citri* pv. *glycines*

Contents of this chapter have been published in *Microorganism* (Kang et al. 2021. Pan-genome analysis of effectors in Korean strains of the soybean pathogen *Xanthomonas citri* pv. *glycines*. *Microorganisms*. 9: 2065, p1-16).

ABSTRACT

Xanthomonas citri pv. *glycines* is a major pathogen of soybean in Korea. Here, we analyzed pathogenicity genes based on a comparative genome analysis of five Korean strains and one strain from the United States, 8ra. Whereas all six strains had nearly identical profiles of carbohydrate active enzymes, they varied in diversity and number of candidate type III secretion system effector (T3SE) genes. The five Korean strains were similar in their effectors, but differed from the 8ra strain. Across the six strains, transcription activator-like effectors (TALEs) showed diverse repeat sizes and at least six forms of the repeat variable di-residue (RVD) sequences, with differences not correlated with the origin of the strains. However, a phylogenetic tree based on the alignment of RVD sequences showed two distinct clusters with 17.5 repeats, suggesting that two distinct 17.5 RVD clusters have evolved, potentially to adapt *Xcg* to growth on distinct soy-bean cultivars. The predicted effector binding elements of the TALEs fell into six groups and were strongly overlapping in sequence, suggesting evolving target specificity of the binding domains in soybean cultivars. Our findings reveal the variability and adaptability of T3SEs in the *Xcg* strains and enhance our understanding of *Xcg* pathogenicity in soybean.

INTRODUCTION

Xanthomonas citri pv. *glycines* (*Xcg*), previously classified as *X. axonopodis* pv. *glycine* (Constantin et al., 2016), causes bacterial pustules in soybean and is one of the most prevalent soybean (*Glycine max*) pathogens in the Republic of Korea (Hong et al., 2010). *Xcg* invades and multiplies extracellularly within the apoplast, causing localized leaf spots or leaf streaks (Boch et al., 2010). *Xcg* employs a suite of virulence factors to colonize plant tissue, including cell wall-degrading enzymes, extracellular polysaccharide and protein secretion systems (Büttner et al., 2010) and a type III protein secretion system.

In the Republic of Korea, a report estimated that 89.7% of soybean fields were infected by *Xcg* in 2010 (Hong et al., 2010). The recessive disease resistance locus *rxp* (resistance to *Xanthomonas phaseoli*) confers partial resistance to *Xcg* (Narvel et al., 2001) and has been introduced into many commercial soybean cultivars. However, the *Xcg* avirulence gene that triggers *rxp*-mediated resistance is unknown.

Xcg strains fall into distinct races (Athinuwat et al., 2009; Hwang et al., 1998; Park et al., 2008) and exhibit varying degrees of virulence

(Kaewnum et al., 2005), which are thought to be associated with distinct profiles of transcription activator-like effectors (TALEs) (Park et al., 2008). The type III secretion system-dependent TALE family is unique to the *Xanthomonas* genus, although some are found in the *Ralstonia solanacearum* species complex (Scholze et al., 2011). The DNA binding domain of TALEs consists of consecutive repeats of highly conserved 33–35 amino acid sequences. The number of repeats present in TALEs varies with an average of 17 and up to 30 repeats (Boch et al., 2010). The amino acids at positions 12 and 13 in the DNA binding domain of TALEs are called the repeat variable di-residues (RVDs), and the sequence of RVDs determines the target specificity of TALEs (Moscou et al., 2009; Gu et al., 2005). TALEs primarily activate the transcription of plant genes involved in disease resistance or susceptibility by binding to effector-binding elements (EBEs) in their promoters (Boch et al., 2010).

Among the *Xanthomonas* TALEs identified so far, the AvrBs3 family has been extensively studied. At least one *avrBs3* gene was identified on a plasmid that is widely present in *Xcg* strains (Kim et al., 2006). The *avrBs3* gene family is associated with mobile elements and transposase genes on the plasmids, suggesting that the *avrBs3* genes might be able to relocate within the bacterial genome. Many *Xanthomonas* strains carry *avrBs3* T3SE genes, yet the copy numbers vary depending on the species. For instance, many *X.*

campestris strains carry only one *avrBs3* gene (Oh et al., 2011), whereas *X. oryzae* strains have 14 to 33 (Oh et al., 2011; Ji et al., 2014; Lee et al., 2005; Yang et al., 2004). Based on hybridization studies, most *Xcg* strains contain three to eight AvrBs3 homologs in their genome (Park et al., 2008; Kim et al., 2006).

Whole-genome sequences of several *Xcg* strains, including strains 12-2, CFBP 2526, CFBP 7119, and BCRC 12609, were previously reported (Chatnaparat et al., 2012; Darrasse et al., 2013; Weng al., 2017). However, these genome assemblies were generated using lllumina-based short-read data. Thus, sequence assembly of TALE genes was likely incomplete due to the presence of many repeat sequences. More recently, the complete genome sequences of three *Xcg* strains, 8ra, 12-2, and EB08 (Lee et al., 2014; Carpenter et al., 2019) were generated using PacBio long-read data. These strains exhibited strong conservation in their overall synteny and repertoire of AvrBs3 family effectors (Carpenter et al., 2019).

Our goal in this work was to compare the genomes of six additional *Xcg* strains that have various levels of virulence in soybean, namely 8ra (a strain from the United States) and the five Korean strains K2, SL1017, SL1018, SL1157, and SL1045. These Korean strains collectively represent *Xcg* isolates collected in the Republic of Korea. We aimed to identify and

compare specific genetic traits associated with pathogenicity, namely the T3SEs and carbohydrate-active enzymes. We particularly focused on the TALE AvrBs3 because of its role in the resistance breeding program in Korea. Our comparative analysis uncovered variability in the sequence and number of AvrBs3 alleles in the Korean *Xcg* strains and in their presumed binding domains.

MATERIALS AND METHODS

1. Bacterial strains and DNA preparation

Four *X. citri* pv. *glycines* (*Xcg*) strains (SL1017, SL1018, SL1157, and SL1045) which was isolated in 1997 selected from the 155 isolates collected from soybean fields in the re-public of Korea. K2 strain was isolated in 2017. The five strains were chosen based on their diverse profiles of infection and aggressiveness on a collection of soybean cultivars and hybridization to an *avrBs3* gene family probe. A US strain, 8ra, was obtained from Dr. Heu of Seoul National University in the Republic of Korea. All strains collected were first tested using the non-host plant tomato to assess their induction of the hypersensitive reaction and the soybean cultivar, Pella, to test for pathogenicity. The taxonomic identification of the strains was verified using fatty acid methyl ester analysis and the Biolog Microbial Identification System according to the manufacturer's instructions. For DNA extraction, *Xcg* strains were cultivated in tryptic soy agar medium (Difco, Franklin Lake, NJ, USA) at 28 °C for 24 h. High-quality, high-molecular-weight genomic DNA was prepared using a DNA extraction kit (iNtRON, Seongnam, Korea) according to the protocol for Gram-negative bacteria. The concentration of the extracted DNA was determined using a BioDrop spectrophotometer

(BioDrop, Cambridge, UK). The integrity of DNA was checked by agarose gel electrophoresis and quantified using a Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA).

2. Genome sequencing and assembly

DNA samples were sequenced using PacBio single-molecule real-time (SMRT) technology. A total of 5 μg for each sample was used as input for library preparation. Libraries were constructed with the SMRTbell Template Prep Kit 1.0 (PN 100-259-100) following the manufacturer's instructions (Pacific Biosciences). Fragments of the SMRTbell template that were less than 20 kb were removed using the Blue Pippin Size selection system. The constructed library was validated to be a large-insert library using an Agilent 2100 Bioanalyzer. After a sequencing primer was annealed to the SMRTbell template DNA, DNA polymerase was bound to the complex using the DNA Polymerase Binding kit P6. This polymerase-SMRTbell-adaptor complex was then loaded onto the SMRT cells. The SMRTbell library was sequenced using one SMRT cell with the MagBead OneCellPerWell v1 Protocol (insert size 20 kb, movie time 1 × 240 min) using P6-C4 chemistry (DNA sequencing Reagent 4.0) on the PacBio RS II (Pacific Biosciences) sequencing platform.

De novo genome assembly was conducted using the hierarchical assembly pipeline of HGAP2 ver.2.3.0 software implemented in PacBio SMRTLink. PRODIGAL version 2.6.2 (Hyatt et al., 2010) was used to predict protein coding sequences (CDSs) based on the bacterial genome database. Rfam and tRNAscan-SE were used to identify rRNA and tRNA sequences, respectively, in the assembled genomes. Each of the *Xanthomonas* genome sequences in this study was assigned an NCBI GenBank accession number as listed in Table 1.

3. Comparative genome analysis

The genome sequences of the six strains were generated and assembled for comparative analysis. To identify homology patterns among the *Xcg* genomes, multiple genome sequences were aligned with progressiveMauve algorithm implemented in the Mauve software3 (Darling et al., 2004). Genome and pan-genome orthologous groups (POGs) of genes were identified by clustering genomes with the unweighted pair group method with arithmetic mean (UPGMA) method, and a heat map was generated to reflect the presence and absence of genes.

For analysis of POGs, a pairwise gene-to-gene comparison of each genome

was conducted using USEARCH (Edgar et al., 2010), and the genes were compared using a reciprocal homology search. A pairwise orthologous group was defined after the initial grouping, and partial genes that were left out of groups due to their short sequence length were targeted for clustering analysis against the POGs using UCLUST (≥95% identity). The CDSs were classified into groups based on their predicted functions, with reference to orthologous groups (EggNOG 4.5, Powell et al., 2014). The annotation of each CDS was conducted by a homology search against the Swis-Prot, EggNOG 4.1, SEED (the database and infrastructure for comparative genomics), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases.

4. CAZyme and TALE analysis

Carbohydrate active enzymes (CAZymes) were evaluated with the online version of automated CAZyme annotation using the dbCAN meta server (http://bcb.unl.edu/dbCAN2/), which implements three independent search modules, i.e., HMMER, Diamond, and Hotpop. All predicted genes identified by the Patric 3.6 annotation pipeline (Wattam et al., 2013) were used to search and identify effectors by running a local BLAST using the Blast2go version 3.2. These predicted genes were further used to identify TALE genes using the HMMER3 program to search the query sequences against the Pfam 27.0

database, in which separate hidden Markov models were constructed using HMMER based on their amino acid sequences. AnnoTALE (Grau et al., 2016) was used for the analysis and annotation of TALE genes and for clustering similar TALEs into classes with agglomerative hierarchical clustering using average linkage. For automated ex-traction of RVDs, the corresponding protein sequence was aligned to the consensus repeat, and the RVDs comprised of the 12th and 13th amino acid positions of each repeat were extracted. For this alignment, a BLOSUM62 substitution matrix and affine gap costs with gap open penalty of 12 and gap extension penalty of 3 were used to account for the existence of aberrant repeat lengths of greater than 35 amino acids. RVDs were aligned using Geneious alignment (Geneious Prime, version 11.1.5). The relationship tree based on RVDs was constructed using a neighbor-joining method implemented in Geneious Prime version 11.1.5. Bootstrapping of 1,000 replicates was applied to check out the robustness of each clade. Putative target sites of a given TALE were determined by the binding specificity of the RVD using a TALE target site finder web application, TALgetter version 2.3, in a local Galaxy server (Grau et al., 2013) (https://www.jstacs.de/index.php/TALgetter).

RESULTS

1. Whole-genome sequence analysis of Korean *X. citri* pv. *glycines* strains

To gain insights into the genome diversity of *Xcg* strains we conducted genome sequence analysis of five Korean *Xcg* strains (K2, SL1017, SL1018, SL1157, and SL1045) and the 8ra strain from the US (Carpenter et al., 2019). Long-read sequence data obtained from the PacBio sequencing platform were used to characterize and compare the genomes of the six strains. *De novo* assembly of the *Xcg* sequences resulted in two to five contigs per strain with an average coverage of 157x. The assembled genome sizes ranged from 5.2 (SL1018) to 5.4 Mbp (8ra) with GC contents ranging from 64.6 to 64.7% (Table 1). The number of predicted total CDSs varied from 4314 in SL1018 to 4597 in 8ra (Table 1).

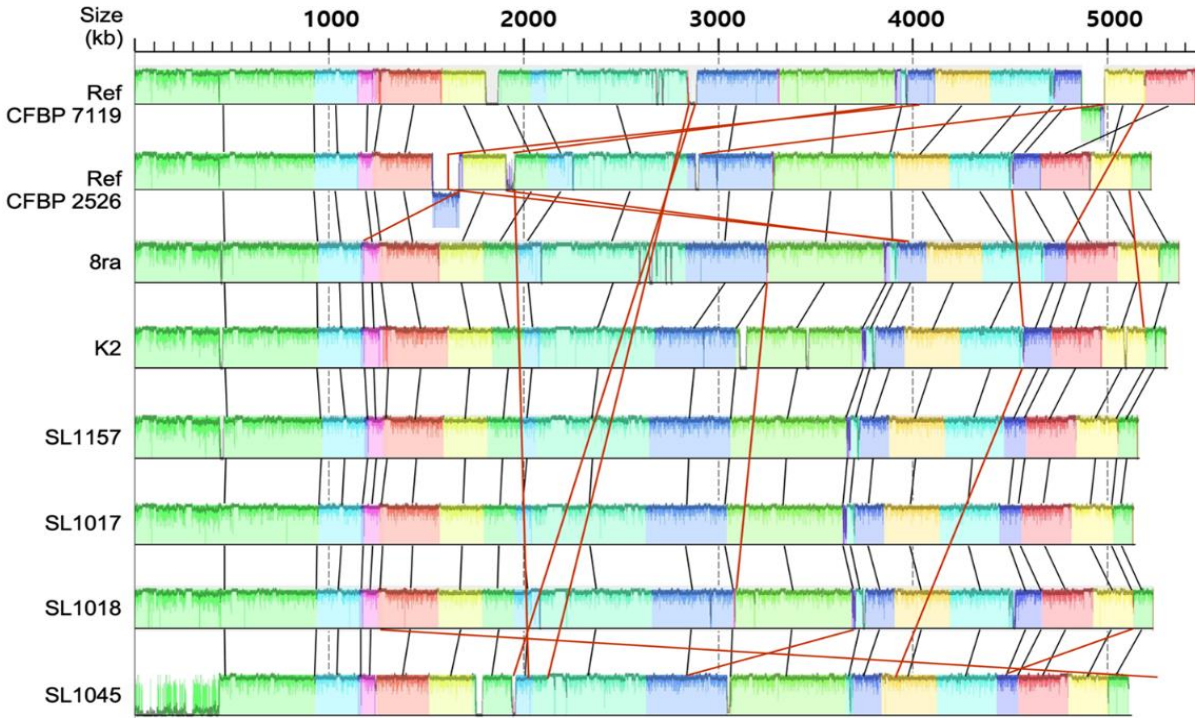
To examine the conservation and variation in the *Xcg* genomes, we performed progressiveMauve alignment (Figure 1), which shows the conserved genomic clusters as similarly colored blocks among the six fully sequenced *Xcg* genomes. The genomes of the sequenced *Xcg* strains in this study (K2, SL1017, SL1018, SL1157, SL1045, and 8ra) are highly homologous to the reference genomes CFBP 7119 and CFBP 2526, although small genomic

rearrangements and inversions were identified (Figure 1). Two other strains, 12-2 and EB08, whose genome sequences are available, were not included because they are nearly identical to strain 8ra (Carpenter et al., 2019). The progressiveMauve alignment demonstrated that the genomes of 8ra, K2, SL1017, SL1018, SL1157, and SL1045 were similar and closely related (Figure 1). These strains were co-linear along the chromosomes; however, SL1045 showed a different genomic arrangement than the other strains.

Table 1. General genomic features of six *X. citri* pv. *glycines* strains.

Strain	No. of Contigs	Depth of Coverage (x)	Genome Size (bp)	% GC	No. of CDSs	No. of Unique CDSs	NCBI Accession Number
8ra	2	178	5,426,838	64.6	4597	144	CP041781, CP041782
K2	2	108	5,322,598	64.6	4519	108	CP041966, CP041967
SL1017	3	198	5,197,177	64.7	4341	5	VMHQ00000000
SL1018	2	163	5,162,305	64.7	4314	5	CP041961, CP041962
SL1157	2	175	5,292,354	64.6	4490	4	CP041963, CP041964
SL1045	5	123	5,169,163	64.7	4332	44	VMHR00000000

Figure 1. Whole-genome comparison of eight *X. citri* pv. *glycines* strains by progressiveMauve alignment. CFBP 7119 and CFBP 2526, isolated in Brazil and Sudan, respectively, were used as reference genomes for comparison (Darrasse et al., 2013). The scale represents the coordinates of each genome. Different color blocks represent LCBs (Local Collinear Blocks), which are conserved segments within the genomes. Within a LCB, the white area represents low similarity regions or regions unique to one genome but absent in another. Colored lines indicate the rearrangement of LCBs between the genomes.


Size
(kb)
1000
2000
3000
4000
5000
Ref
CFBP 7119
Ref
CFBP 2526
8ra
K2
SL1157
SL1017
SL1018
SL1045

2. Comparison of gene content and number of Korean *X. citri* pv. *glycines* strains

We constructed a heatmap and dendrogram based on the gene content of the six *Xcg* strains. The 4700 CDSs in the strains were clustered into pan-genome orthologous groups (POGs) (Figure 2). The heatmap showed that the presence/absence patterns of the POGs in the strain 8ra were different from those in the Korean strains since there were about 100 extra genes in 8ra. Compared to strain 8ra, SL1017 and SL1018 showed 0.38% variation, while K2 showed 2% variation. SL1045 showed 4% variation. In the UPGMA relationship tree based on gene sharing between strains, SL1017 was closely aligned with SL1018, and 8ra was closely aligned with K2.

We also constructed a Venn diagram (Bardou et al., 2014) to compare predicted gene numbers in the five *Xcg* strains and 8ra (Figure 3) and identify the number of shared and unique CDSs. The six *Xcg* strains were closely related to each other at the genome sequence level; they shared 4156 genes (CDSs), with an average of 260 shared genes per genome. SL1017 and SL1018 appeared to have five and four unique genes, respectively. SL1157, SL1045, and K2 had 50, 44, and 108 unique genes, respectively. Strain 8ra had 144 unique genes (Figure 3). Excluding SL1045, the strains shared 39 genes, with 8ra and SL1157 sharing the most, 76 genes.

Figure 2. Comparison of gene content (presence/absence) as shown in a heatmap and a dendrogram (unweighted pair group method with arithmetic mean, UPGMA) of 4700 protein-coding sequences (CDSs) that were placed into pan-genome orthologous groups (POGs). The numbers on the X-axis and columns in the figure represent distinct POGs, which were ordered based on their differential presence (indicated in yellow) or absence (indicated in orange) among the strains.

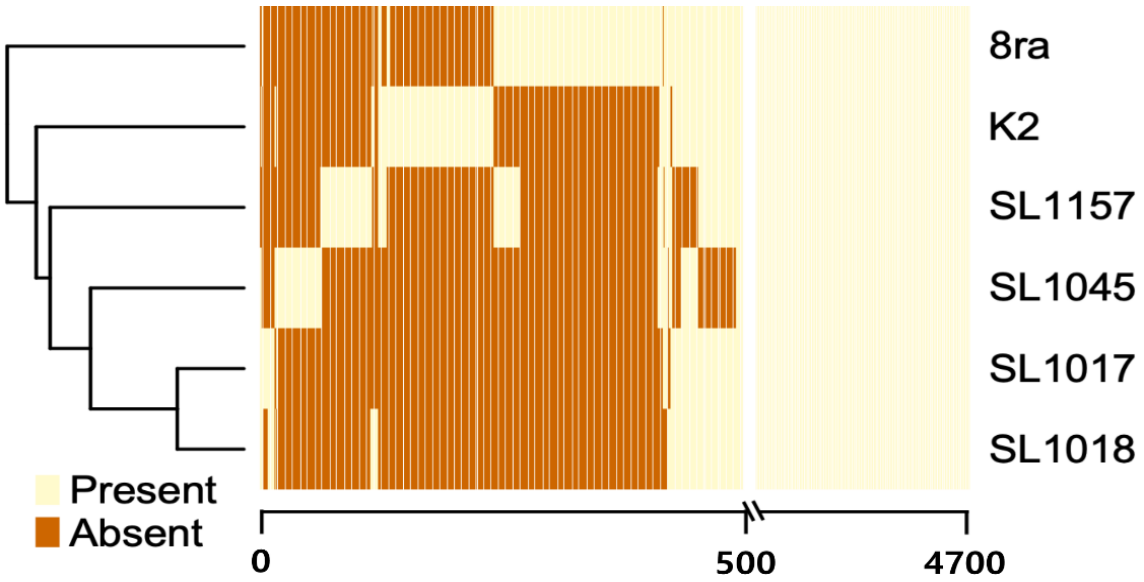
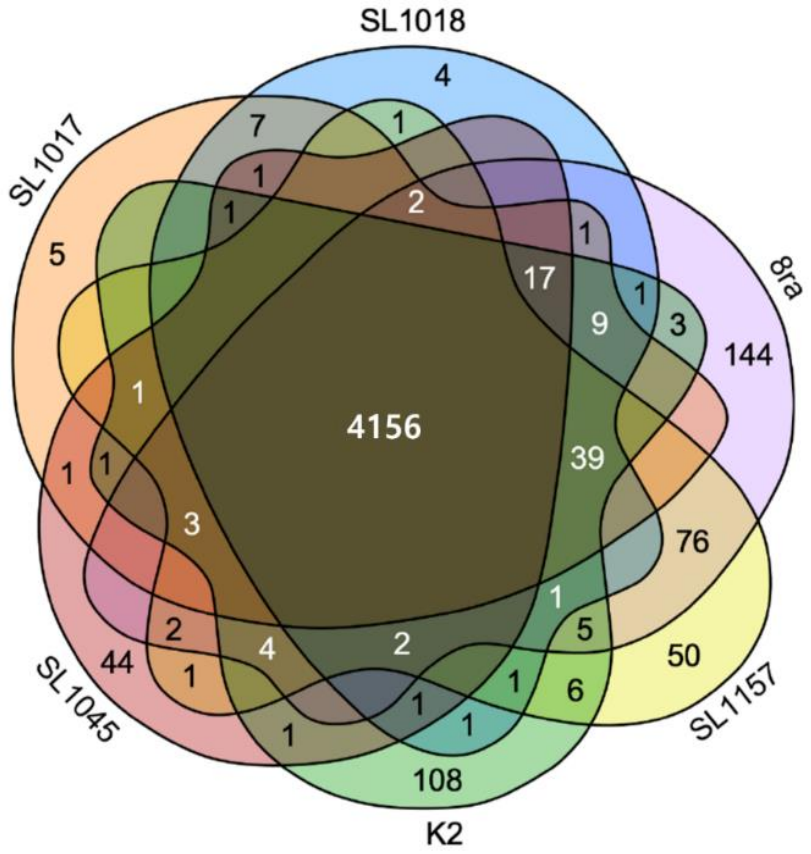

8ra
K2
SL1157
SL1045
SL1017
SL1018
Present
Absent
0
500
4700

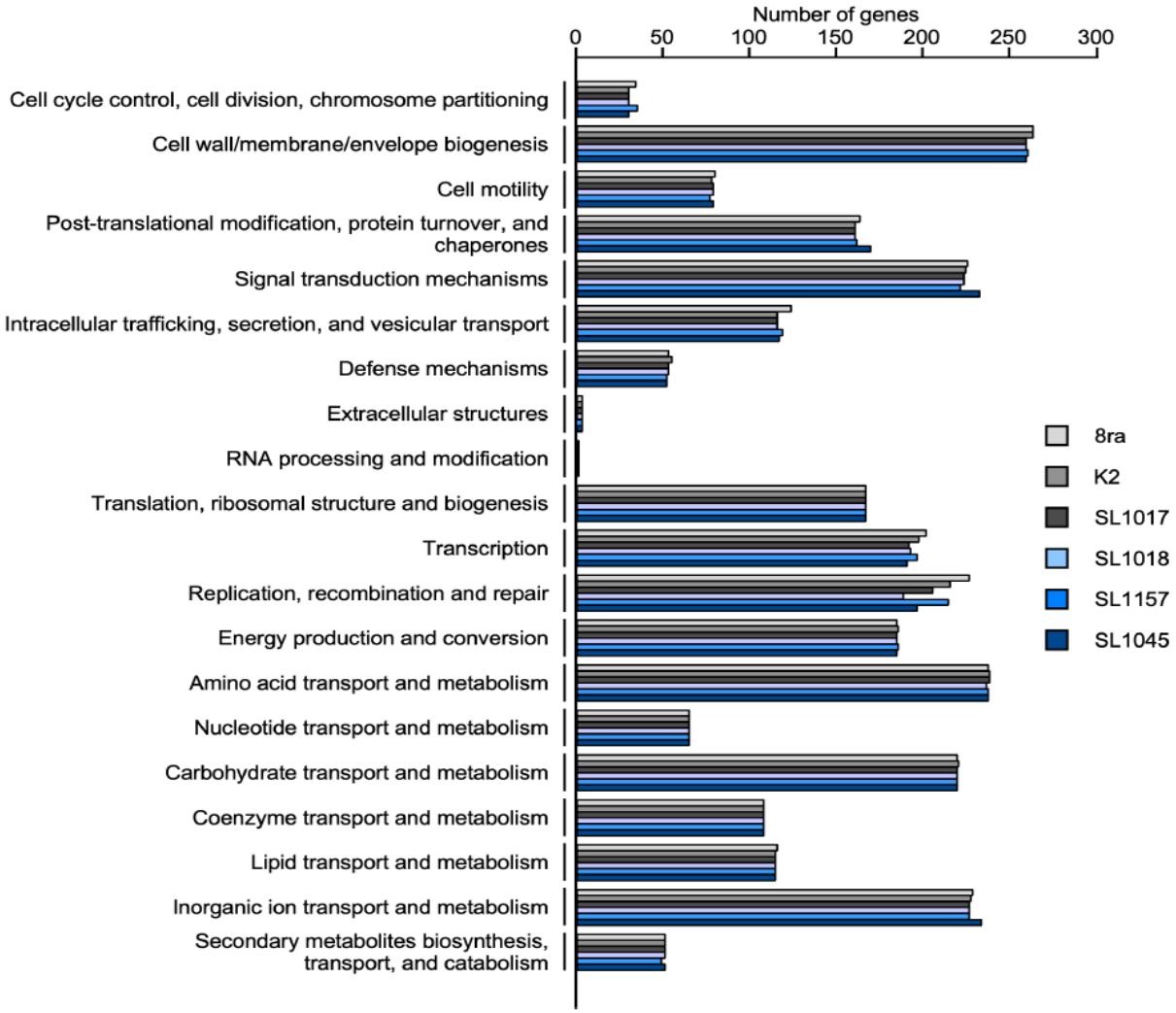
Figure 3. Venn diagram of genes found within the five Korean strains and a US strain, 8ra with numbers in the intersected regions indicating shared genes.


SL1018
SL1017
8ra
SL1045
SL1157
K2
4
7
1
1
2
1
5
17
1
9
3
144
1
4156
1
1
39
3
76
1
2
5
44
4
2
50
1
1
6
1
1
1
108

3. Functional gene distribution in the Korean *X. citri* pv. *glycines* strains and strain 8ra

To compare the distribution of functional genes among the six *Xcg* strains in relation to bacterial virulence, we analyzed the functional gene categories in the Clusters of Orthologous Groups of proteins (COGs) database (Tatusov et al., 2000) (Figure 4). Among the predicted CDSs, about 70% of the genes were classified into one of the 21 COGs categories, and about 30% were of unknown function. In five functional categories, COGs were similarly distributed; these groups were the extracellular structures, RNA processing, and modification group, translation, ribosomal structure and biogenesis group, nucleotide transport and metabolism group, and coenzyme transport and metabolism group. The strains showed variability in gene number, primarily in two categories, transcription group and replication, recombination, and repair group. Strain SL1045 had more genes in several categories such as post-translational modifications, protein turnover, and chaperones group, signal transduction mechanisms group, inorganic ion transport, and metabolism group.

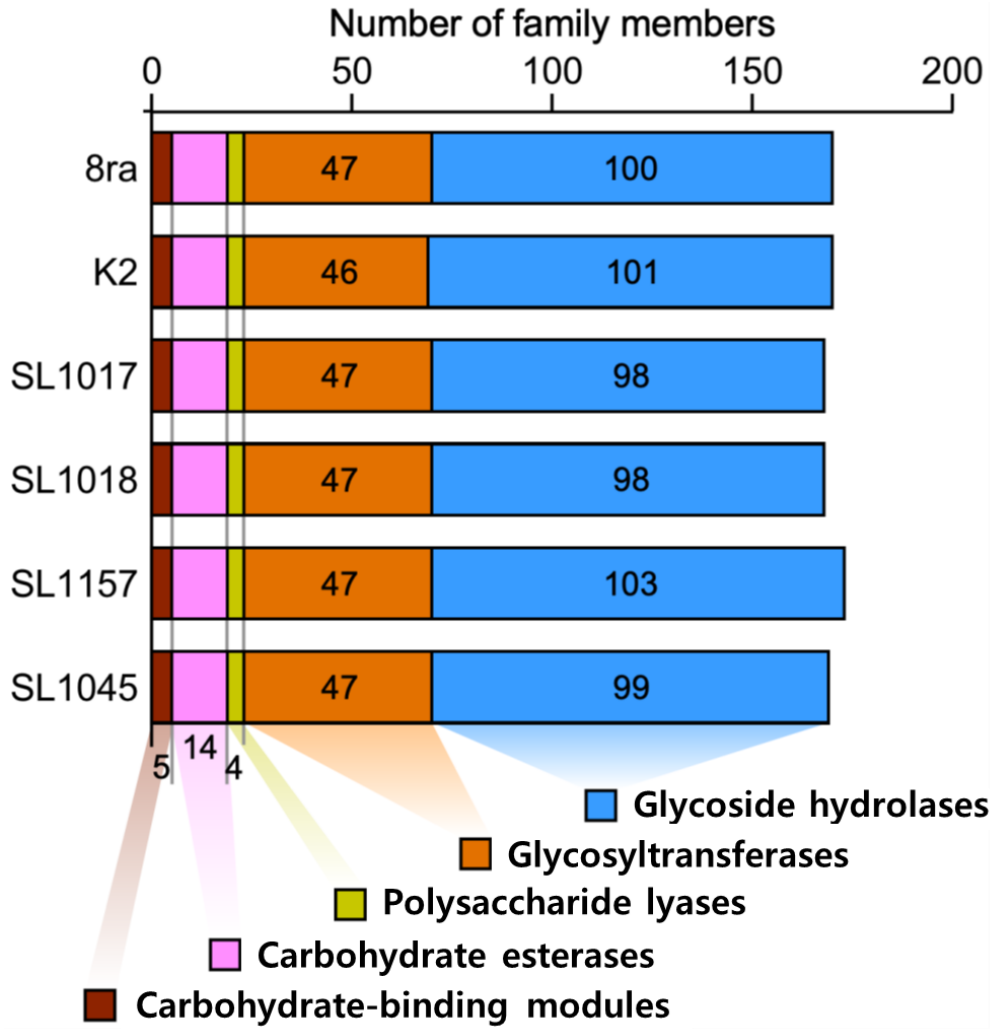
Figure 4. Number of genes in distinct functional categories for the five Korean strains and a US strain, 8ra. All putative genes were aligned to the database of Clusters of Orthologous Groups (COGs) to predict possible functions. The genes were classified into 21 total COGs. The unknown function category is not shown. Values on the X-axis indicate the number of genes, and the Y-axis labels indicate the COG categories.


Number of genes
0
50
100
150
200
250
300
Cell cycle control, cell division, chromosome partitioning
Cell wall/membrane/envelope biogenesis
Cell motility
Post-translational modification, protein turnover, and chaperones
Signal transduction mechanisms
Intracellular trafficking, secretion, and vesicular transport
Defense mechanisms
Extracellular structures
RNA processing and modification
Translation, ribosomal structure and biogenesis
Transcription
Replication, recombination and repair
Energy production and conversion
Amino acid transport and metabolism
Nucleotide transport and metabolism
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Lipid transport and metabolism
Inorganic ion transport and metabolism
Secondary metabolites biosynthesis, transport, and catabolism
8ra
K2
SL1017
SL1018
SL1157
SL1045

4. Comparison of the total number of CAZymes present in the genomes of Korean *X. citri* pv. *glycines* strains and strain 8ra

To examine the repertoire of carbohydrate-active enzymes in the *Xcg* strains, we collected CAZymes from five classes and organized them into families based on their amino acid sequence similarity. We found that the strains had similar proportions of CAZymes belonging to the five major functional classes (Figure 5). Among the CAZymes identified, 58–59% were glycoside hydrolases (GHs), 27.5% were glycosyltransferases (GTs), 2% were polysaccharide lyases (PLs), 8% were carbohydrate esterases (CEs), and 2% were carbohydrate-binding modules (CBMs). All of the strains had four PLs, 14 CEs, and five CBMs, whereas they differed primarily in the number of GHs. SL1157 had the highest number of GHs.

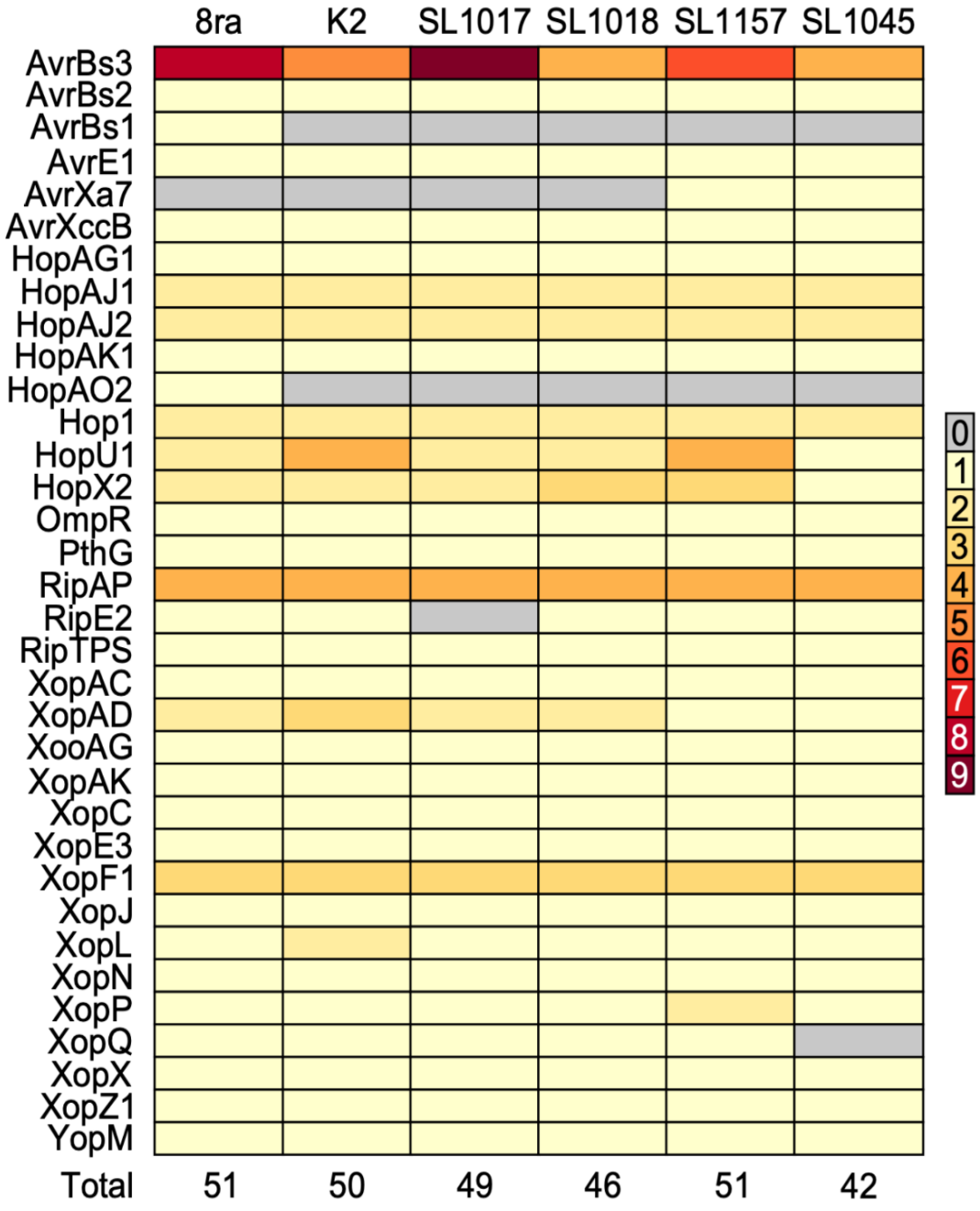
Figure 5. Comparison of total number of CAZymes predicted in the genomes of six *X. citri* pv. *glycines* strains. CAZymes were classified into five classes of enzymatic domains, namely, the glycoside hydrolases domain class, glycosyltransferases domain class, polysaccharide lyases domain class, carbohydrate esterases domain class, and carbohydrate-binding modules domain class.


Number of family members
0
50
100
150
200
8ra
47
100
K2
46
101
SL1017
47
98
SL1018
47
98
SL1157
47
103
SL1045
47
99
5
14
4
Glycoside hydrolases
Glycosyltransferases
Polysaccharide lyases
Carbohydrate esterases
Carbohydrate-binding modules

5. Comparative analysis of type III Secretion system effector candidate genes

The type III secretion system (T3SS) is critical to the pathogenicity of *Xcg*, as it secretes effectors into the host plant cell. To identify type III secretion system effectors (T3SEs), we used the predicted genes for each of the six strains from Patric 3.6 as queries for BLAST searches of *Xcg* using the Blast2go program. The identified effector names and functions are summarized in Figure 6. The comparative analysis of T3SE candidate genes showed the differences among the strains. A total of 49 effectors were identified in 8ra, 44 effectors in K2, 47 effectors in SL1017, 40 effectors in SL1045, 49 effectors in SL1157, and 44 effectors in SL1018. The US strain 8ra had two effectors, AvrBs1 and HopAO2, that were not found in the five Korean strains. Park et al. (2008) also showed that strains K2, SL1017, SL1018, SL1157, and SL1045 lack AvrBs1. The strains share few Xop effectors with strains causing bacterial spot of tomato and pepper (*X. euvesicatoria* and *X. perforans* strains); this includes both host-specific and non-host-specific effectors (Barak et al., 2016). Additionally, SL1017 and SL1045 did not have RipE2 and XopQ, respectively, and SL1157 and SL1045 had AvrXa7.

Figure 6. The number of type III secretion system effectors in the five Korean *X. citri* pv. *glycines* strains and a US strain, 8ra. The colors represent the number of genes for each effector.


8ra
K2
SL1017
SL1018
SL1157
SL1045
AvrBs3
AvrBs2
AvrBs1
AvrE1
AvrXa7
AvrXccB
HopAG1
HopAJ1
HopAJ2
HopAK1
HopAO2
Hop1
HopU1
HopX2
OmpR
PthG
RipAP
RipE2
RipTPS
XopAC
XopAD
XooAG
XopAK
XopC
XopE3
XopF1
XopJ
XopL
XopN
XopP
XopQ
XopX
XopZ1
YopM
Total
51
50
49
46
51
42
0
1
2
3
4
5
6
7
8
9

6. Comparative analysis of transcription activator-like effector candidate genes

Transcription activator-like effectors (TALEs) are critical for virulence in *Xanthomonas* spp., with some activating key susceptibility genes. Here, we extracted the sequences of TALEs, all of which are homologs of AvrBs3 (Figure 6), to examine the differences among the *Xcg* strains. As TALE specificity primarily depends on the repeat variable di-residues, or RVDs, we aligned the RVD sequences and analyzed the TALEs of each strain with both the HMMER3 program using the Pfam 27.0 database and the AnnoTALE program (Table 3). The HMMER3 program detected more TALEs than the AnnoTALE program (Grau et al., 2016). To compare TALEs in the five Korean *Xcg* strains and 8ra, we renamed all complete TALEs with the guidance of AnnoTALE (Table 2).

Given that some of the TALEs appear to have incomplete forms of RVDs (Table 3), we used only TALE candidates with over 12.5 repeats for RVD analysis. Among these TALEs that were considered complete, the number of repeats varied from 12.5 to 34.5 (Table 2). These TALEs were classified into six groups (Figure 7). Alignment of the RVDs of all of the TALEs showed the presence of a 21-residue consensus region, with-in which N and HDN residues were conserved in all complete TALEs (Figure 8). The TALE proteins of the *Xcg* strains clustered based on the repeat number, with the exception of the TALEs containing 17.5 repeats, which formed two groups (Figure 7). The most common TALEs had 17.5 repeats, followed by those with 19.5 repeats. The sequence HDN recurred frequently within the TALEs, with each TALE group containing at least four HDN repeats (Figure 8).

A phylogenetic tree constructed based on the RVDs of the TALEs showed clustering into three major groups, one with TALEs containing 17.5 repeats (group 1), one with TA-LEs

containing 22.5–34.5 repeats, and one with all of the other TALEs. The phylogenetic tree further supports the division of the TALEs with 17.5 repeats into two distinct clades.

Table 2. The transcription activator-like effectors (TALEs) that were identified in six *Xcg* strains and annotated using AnnoTALE based on the repeat variable di-residues (RVDs).

Strain name	**AnnoTALE name**	**No. of Repeats**	**RVDs** (the 12th and 13th amino acid residues of each repeat region)
8ra	TalGN3 8ra	22.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG-HD-HD-NG
	TalGR3 8ra	18.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGO3 8ra	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGQ3 8ra	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NS-HD-NI-NG-NG
	TalGP3 8ra	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NS-HD-NS-HD-NI-NG-NG
	TalGM3 8ra	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGM4 8ra	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
K2	TalGN3 K2	22.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG-HD-HD-NG
	TalGR3 K2	18.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGO3 K2	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGQ3 K2	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NS-HD-NI-NG-NG
	TalGP3 K2	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NS-HD-NS-HD-NI-NG-NG
SL1017	TalGN3 SL1017	22.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NS-HD-HD-NG
	TalGO3 SL1017	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalHQ1 SL1017	16.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NG-NS-HD-NI-HD-NI-NG-NG
	TalGP3 SL1017	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NS-HD-NS-HD-NI-NG-NG
	TalGR4 SL1017	12.5	NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGM3 SL1017	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGR3 SL1017	18.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGO4 SL1017	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
SL1018	TalGN3 SL1018	34.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NS-HD-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGQ3 SL1018	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NI-HD-NI-NG-NG
	TalGM3 SL1018	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGM4 SL1018	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
SL1157	TalGN3 SL1157	22.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NS-HD-HD-NG
	TalGO3 SL1157	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGQ3 SL1157	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NS-HD-NI-NG-NG
	TalGP3 SL1157	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NI-HD-NS-HD-NI-NG-NG
	TalGM3 SL1157	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGR3 SL1157	18.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
SL1045	TalGM3 SL1045	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGR3 SL1045	18.5	NI-HD-NS-NS-NI-NI-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGQ3 SL1045	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NS-HD-NI-NG-NG
	TalGP3 SL1045	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NS-HD-NS-HD-NI-NG-NG

Figure 7. Protein sequence alignment for the repeat variable di-residues (RVDs) from the transcription activator-like effectors (TALEs) of six *X. citri* pv. *glycines* strains. All RVDs of TALEs were aligned with Geneious Alignment (https://www.geneious.com) (Table 4); only those predicted to be complete are shown here. Red lines above the RVD sequence highlight recurring repeats of the sequence HDN.

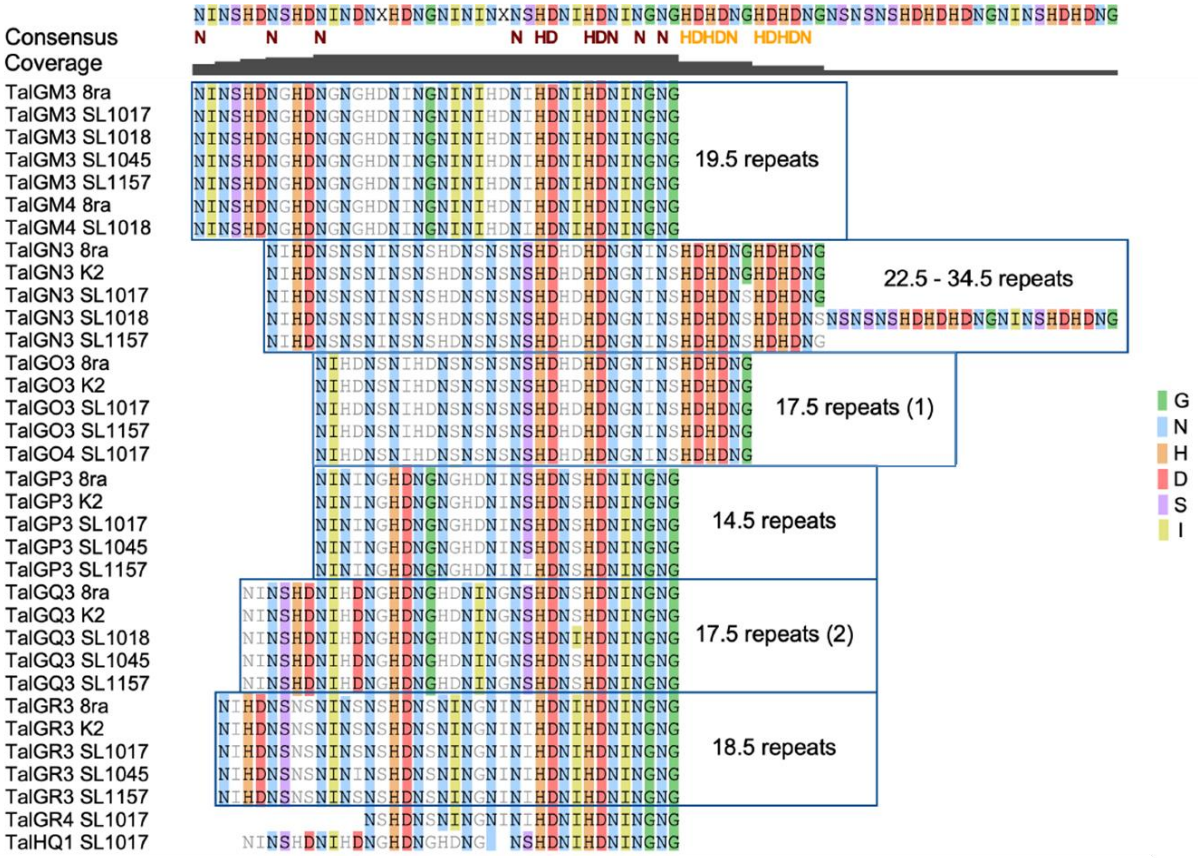

Consensus
Coverage
TalGM3 8ra
TalGM3 SL1017
TalGM3 SL1018
TalGM3 SL1045
TalGM3 SL1157
TalGM4 8ra
TalGM4 SL1018
TalGN3 8ra
TalGN3 K2
TalGN3 SL1017
TalGN3 SL1018
TalGN3 SL1157
TalGO3 8ra
TalGO3 K2
TalGO3 SL1017
TalGO3 SL1157
TalGO4 SL1017
TalGP3 8ra
TalGP3 K2
TalGP3 SL1017
TalGP3 SL1045
TalGP3 SL1157
TalGQ3 8ra
TalGQ3 K2
TalGQ3 SL1018
TalGQ3 SL1045
TalGQ3 SL1157
TalGR3 8ra
TalGR3 K2
TalGR3 SL1017
TalGR3 SL1045
TalGR3 SL1157
TalGR4 SL1017
TalHQ1 SL1017
19.5 repeats
22.5 - 34.5 repeats
17.5 repeats (1)
14.5 repeats
17.5 repeats (2)
18.5 repeats
G
N
H
D
S
I

Table 3. The repeat variable di-residues (RVDs) from all transcription activator-like effectors (TALEs) of six *X. citri* pv. *glycines* strains using both the HMMER3 program with the Pfam 27.0 database and AnnoTALE program[a].

strain name	AnnoTALE name	HMMER3	No. of Repeats	RVDs (the 12th and 13th amino acid residues of each repeat region)
8ra	TalGN3 8ra	8ra_1	22.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG-HD-HD-NG
	TalGM3 8ra	8ra_2	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGM4 8ra	8ra_3	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGR3 8ra	8ra_4	18.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGO3 8ra	8ra_5	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGQ3 8ra	8ra_6	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NS-HD-NI-NG-NG
	TalGP3 8ra	8ra_7	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NS-HD-NS-HD-NI-NG-NG
	TalED3 8ra	8ra_8	4.5	NI-HD-NI-NG-NG
K2	TalGN3 K2	K2_1	22.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG-HD-HD-NG
	TalGR3 K2	K2_2	18.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGO3 K2	K2_3	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGQ3 K2	K2_4	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NS-HD-NI-NG-NG
	TalGP3 K2	K2_5	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NS-HD-NS-HD-NI-NG-NG
		K2_6	7	NI-NI-HD-NI-HD-NI-HD
		K2_7	3.5	NI-NS-HD-NG
		K2_8		
SL1017	TalGN3 SL1017	SL1017_1	22.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NS-HD-HD-NG
	TalGM3 SL1017	SL1017_2	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGR3 SL1017	SL1017_3	18.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGO3 SL1017	SL1017_4	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGO4 SL1017	SL1017_5	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalHQ1 SL1017	SL1017_6	16.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NG-NS-HD-NI-HD-NI-NG-NG
	TalGP3 SL1017	SL1017_7	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NS-HD-NS-HD-NI-NG-NG
	alGR4 SL1017(pseud	SL1017_8	12.5	NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGO5 SL1017	SL1017_9	6.5	NI-HD-NS-NI-HD-NS-NS
		SL1017_10	4.5	NI-HD-NS-NS-NI
		SL1017_11	1.5	HD-HD
		SL1017_12	0.5	NG
SL1018	TalGN3 SL1018	SL1018-1	34.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NS-HD-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGM3 SL1018	SL1018-2	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGM4 SL1018	SL1018-3	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGQ3 SL1018	SL1018-4	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NI-HD-NI-NG-NG
SL1157	TalGN3 SL1157	SL1157_1	22.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NS-HD-HD-NG
	TalGM3 SL1157	SL1157_2	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGR3 SL1157	SL1157_3	18.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGO3 SL1157	SL1157_4	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGQ3 SL1157	SL1157_5	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NS-HD-NI-NG-NG
	TalGP3 SL1157	SL1157_6	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NI-HD-NS-HD-NI-NG-NG
		SL1157_7	14.5	NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
		SL1157_8	0.5	NI
		SL1157_9		
SL1045	GM3 SL1045 (pseud	1045_1	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGR3 SL1045	1045_2	18.5	NI-HD-NS-NS-NI-NI-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	GQ3 SL1045 (pseud	1045_3	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NS-HD-NI-NG-NG
	TalGP3 SL1045	1045_4	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NS-HD-NS-HD-NI-NG-NG
		1045_5	12.5	HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
		1045_6	9.5	NI-HD-NS-NS-NI-NI-NS-HD-NS-N*
		1045_7	6	NI-NS-HD-NG-HD-NG
		1045_8	5	NI-HD-NS-NS-NI
		1045_9	3.5	HD-NI-NG-NG
		1045_10	2.5	NI-HD-N*
		1045_11	2.5	HD-HD-NG
		1045_12	0.5	N*
		1045_13	0.5	NS
		1045_14		
		1045_15		
		1045_16		

[a] RVDs in red indicate TALEs that are predicted to be encoded on plasmids.

Table 4. Clustering of the RVDs from all of the TALEs in Table 3 into six groups based on sequence similarity using the Geneious Alignment program[a,b].

TALE group	AnnoTALE name	No. of Repeats	RVDs (the 12th and 13th amino acid residues of each repeat region)
2	TalGN3 8ra	22.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG-HD-HD-NG
	TalGN3 K2	22.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG-HD-HD-NG
	TalGN3 SL1017	22.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NS-HD-HD-NG
	TalGN3 SL1018	34.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NS-HD-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGN3 SL1157	22.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NS-HD-HD-NG
6	TalGM3 8ra	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGM4 8ra	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGM3 SL1017	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGM3 SL1018	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGM4 SL1018	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGM3 SL1157	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	TalGM3 SL1045	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	K2_6	7	NI-NI-HD-NI-HD-NI-HD
	1045_5	12.5	HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	1045_7	6	NI-NS-HD-NG-HD-NG
1	TalGR3 8ra	18.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGR3 K2	18.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGR3 SL1017	18.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGR3 SL1157	18.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGR3 SL1045	18.5	NI-HD-NS-NS-NI-NI-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	TalGR4 SL1017	12.5	NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	SL1017_10	4.5	NI-HD-NS-NS-NI
	SL1157_7	14.5	NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	1045_6	9.5	NI-HD-NS-NS-NI-NI-NS-HD-NS-N*
	1045_9	3.5	HD-NI-NG-NG
3	TalGO3 8ra	17.5	NI-HD-NS-NI-HD-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGO3 K2	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGO3 SL1017	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGO3 SL1157	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGO4 SL1017	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
	TalGO5 SL1017	6.5	NI-HD-NS-NI-HD-NS-NS
	K2_7	3.5	NI-NS-HD- -NG
	SL1017_11	1.5	HD-HD
	1045_11	2.5	HD-HD-NG
4	TalGQ3 8ra	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NS-HD-NI-NG-NG
	TalGQ3 K2	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NS-HD-NI-NG-NG
	TalGQ3 SL1018	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NI-HD-NI-NG-NG
	TalGQ3 SL1157	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NS-HD-NI-NG-NG
	TalGQ3 SL1045	17.5	NI-NS-HD-NI-HD-NG-HD-NG-HD-NI-NG-NS-HD-NS-HD-NI-NG-NG
	TalHQ1 SL1017	16.5	NI-NS-HD-NI-HD-NG-HD-NG-HD- -NG-NS-HD-NI-HD-NI-NG-NG
	TalED3 8ra	4.5	NI-HD-NI-NG-NG
5	TalGP3 8ra	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NS-HD-NS-HD-NI-NG-NG
	TalGP3 K2	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NS-HD-NS-HD-NI-NG-NG
	TalGP3 SL1017	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NS-HD-NS-HD-NI-NG-NG
	TalGP3 SL1157	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NI-HD-NS-HD-NI-NG-NG
	TalGP3 SL1045	14.5	NI-NI-NG-HD-NG-NG-HD-NI-NS-HD-NS-HD-NI-NG-NG

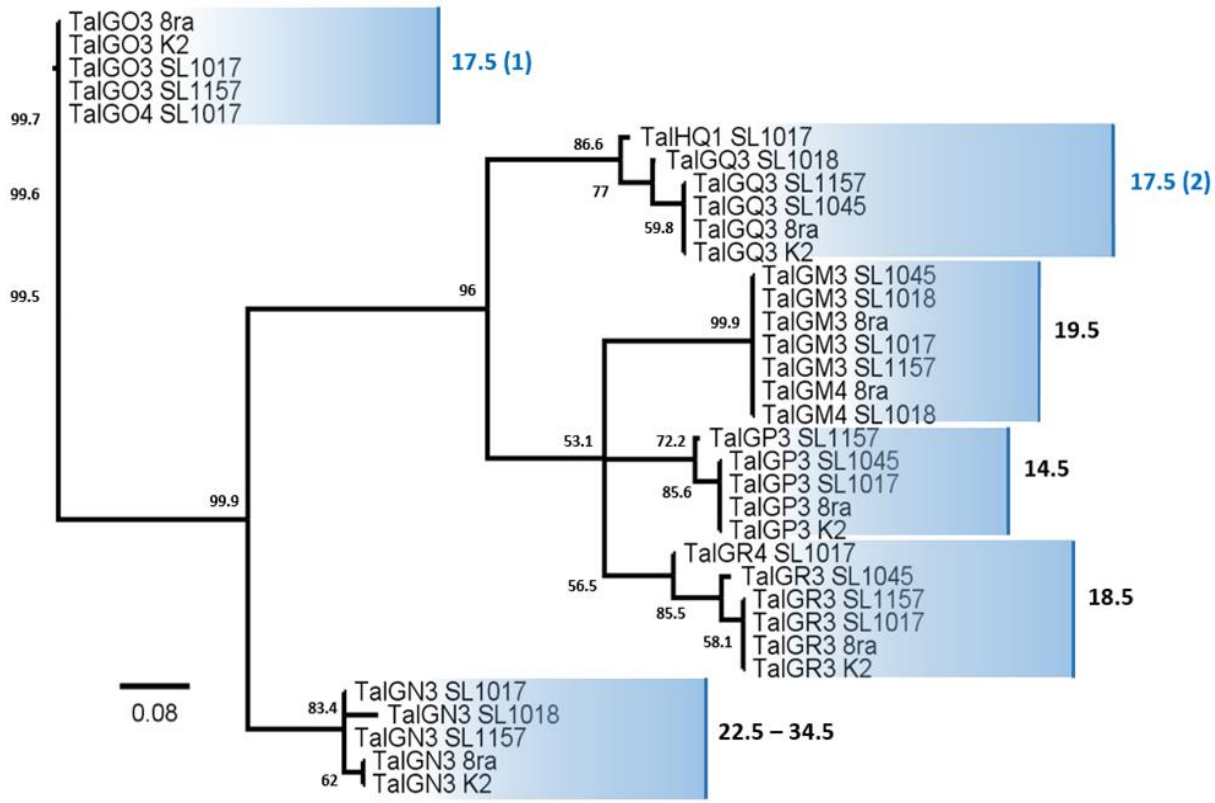
[a] The TALE groups are defined based on the clustering shown in Figure 7, with the addition of shorter and potentially incomplete TALEs from Table 3.

[b] RVDs in blue indicate incomplete TALEs; RVDs in red indicate residues showing differences within the same group.

Table 5. RVDs of *X. citri* pv. *glycine* TALEs found on plasmids

strain name	TALE group	AnnoTALE name	No. of Repeats	RVDs (the 12th and 13th amino acid residues of each repeat region)
8ra P1 (plasmid)	6	TalGM3 8ra	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	6	TalGM4 8ra	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	4	TalED3 8ra	4.5	NI-HD-NI-NG-NG
SL1017 P1 (plasmid)	1	TalGR3 SL1017	18.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
	3	TalGO4 SL1017	17.5	NI-HD-NS-NI-HD-NS-NS-NS-NS-HD-HD-HD-NG-NI-NS-HD-HD-NG
SL1017 P2	6	TalGM3 SL1017	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
SL1018 P1 (plasmid)	6	TalGM3 SL1018	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	6	TalGM4 SL1018	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
SL1157 P1 (plasmid)	6	TalGM3 SL1157	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG
	1	TalGR3 SL1157	18.5	NI-HD-NS-NS-NI-NS-NS-HD-NS-NI-NG-NI-NI-HD-NI-HD-NI-NG-NG
SL1045 P1 (plasmid)	6	TalGM3 SL1045	19.5	NI-NS-HD-NG-HD-NG-NG-HD-NI-NG-NI-NI-HD-NI-HD-NI-HD-NI-NG-NG

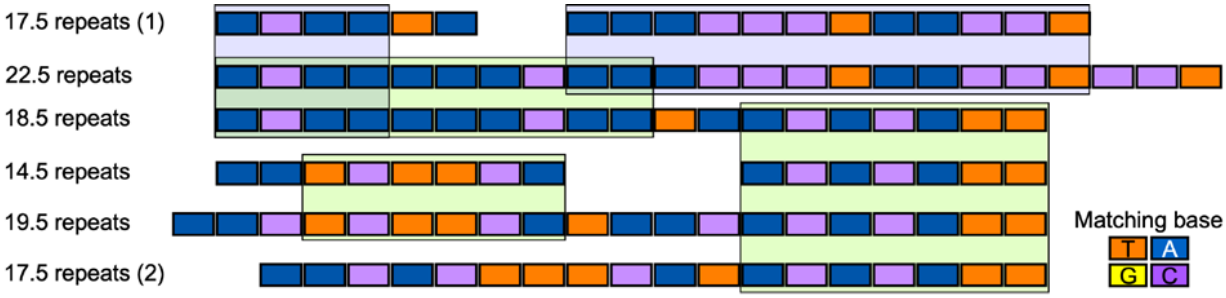
Figure 8. Phylogenetic tree of the transcription activator-like effectors identified in six *X. citri* pv. *glycines* strains based on their repeat variable di-residue sequences. The neighbor-joining tree was based on bootstrapping of 1,000 replications. Bootstrap values above 50% are shown on the node. Bar (0.08) indicates the substitution rate per nucleotide position.


TalGO3 8ra
TalGO3 K2
TalGO3 SL1017
TalGO3 SL1157
TalGO4 SL1017
17.5 (1)
TalHQ1 SL1017
TalGQ3 SL1018
TalGQ3 SL1157
TalGQ3 SL1045
TalGQ3 8ra
TalGQ3 K2
17.5 (2)
TalGM3 SL1045
TalGM3 SL1018
TalGM3 8ra
TalGM3 SL1017
TalGM3 SL1157
TalGM4 8ra
TalGM4 SL1018
19.5
TalGP3 SL1157
TalGP3 SL1045
TalGP3 SL1017
TalGP3 8ra
TalGP3 K2
14.5
TalGR4 SL1017
TalGR3 SL1045
TalGR3 SL1157
TalGR3 SL1017
TalGR3 8ra
TalGR3 K2
18.5
TalGN3 SL1017
TalGN3 SL1018
TalGN3 SL1157
TalGN3 8ra
TalGN3 K2
22.5 – 34.5
0.08
99.7
99.6
99.5
99.9
96
86.6
77
59.8
99.9
53.1
72.2
85.6
56.5
85.5
58.1
83.4
62

7. Comparative analysis of the predicted effector binding elements of the transcription activator-like effector candidate genes

Since the prediction of the effector binding elements (EBEs) is essential to elucidating the biological functions of TALEs, we predicted the EBEs of the TALEs of the six *Xcg* strains. TALE target sites were predicted using TALgetter (Grau et al., 2013). The target gene promoter sequences contained exclusively A, T, and C bases (Figure 9). The predicted EBEs showed five conserved domains (Figure 9). These domains were intermixed among the TALEs. For example, EBEs corresponding to TALEs with 18.5 repeats shared one domain with those containing 22.5 repeats and one domain with those containing 14.5, 19.5, and 17.5 repeats (second group).

Figure 9. The predicted effector binding elements (EBEs) of the transcription activator-like effectors (TALEs) of six *X. citri* pv. *glycines* strains. The EBEs were predicted based on the binding specificity of the repeat variable di-residue sequences with a web-application of TALE target site finder, TALgetter v2.3 in a local Galaxy server (https://www.jstacs.de/index.php/TALgetter). Five conserved domains are highlighted with shaded boxes.


17.5 repeats (1)
22.5 repeats
18.5 repeats
14.5 repeats
19.5 repeats
17.5 repeats (2)
Matching base
T
A
G
C

DISCUSSION

Xcg is an emerging problematic pathogen in soybean in Korea. Genome-wide studies of pathogenicity effectors and analysis of the pathogenic variability of representative strains in Korea should allow us to understand the extent of variation among distinct races of *Xcg*. Here, we characterized potential pathogenicity factors in five Korean *Xcg* strains which were previously reported to exhibit distinct infection profiles and aggressiveness on soybean. Specifically, we analyzed whole genomes sequences of these five Korean strains and the US strain 8ra for two types of pathogenicity factors, plant cell wall-degrading enzymes and type III secretion system effectors.

In a progressiveMauve alignment, the genomes of the six strains, 8ra, K2, SL1017, SL1018, SL1157, and SL1045 were similar and closely related to the genomes of two reference strains, CFBP 7119 and CFBP 2526, although the SL1045 genome was >100 kb smaller than the others. Up to 100 genes are missing in strain SL1045 compared with the other strains, accounting for its heterogeneous gene content (Figures 2 and 4).

Many members of the *Xanthomonas* species are mesophyll pathogens, and their genomes generally encode a range of plant cell wall-degrading enzymes and the associated type II secretion systems encoded by the *xps* and *xcs* gene clusters. Genes for such carbo-hydrate-degrading enzymes, or CAZymes, have been identified previously in strains of bacterial leaf spot-causing *Xanthomonas* species (Potnis et al., 2011). Here, we found that the *Xcg* strains examined had a similar number of CAZymes in each of the distinct domain classes, with some variation in the number of GHs. The GHs in these *Xcg* strains include GHs in the GH9 and

GH5 families. This finding is consistent with the reported role for xyloglucan in the transcriptional activation of virulence factors in *Xanthomonas* spp. (Vieira et al., 2021) since xyloglucan is used by GH9 and GH5 enzymes with xyloglucanase activity. Notably, the sugars released by these enzymes elicit the expression of several key virulence factors in *Xanthomonas* spp., including components of the type III secretion system.

A previous study reported that the Korean *Xcg* strains do not have AvrBs1 homologs whereas strain 8ra does (Park et al., 2008). Here, we found that the Korean *Xcg* strains also do not have HopAO2 homologs whereas strain 8ra does. HopAO2 effectors have phosphatase activity and are involved in the suppression of plant immune responses (Castañeda-Ojeda et al., 2017). The AvrBs1 and HopAO2 effectors may facilitate survival of strain 8ra inside plant cells and thus contribute to its pathogenicity. To date, only a few Korean isolates have been analyzed, so continuous screening of recently isolated *Xcg* strains from the Republic of Korea will help indicate if strains with either of these two effectors are currently present or, in the future, emerge or are introduced into the country.

Distinct host specificity has been reported among strains that carry different numbers of TALE homologs and that differ in their response to diverse soybean cultivars (Park et al., 2008). Athinuwat et al. (2009) also showed that differences in the virulence of *Xcg* strains to soy-bean cultivars is determined by AvrBs3. Since the TALE AvrBs3 determines host and virulence specificity, identifying the target specificity of AvrBs3 variants with distinct sequences and repeats could provide insights into AvrBs3 functions. In this study, a collection of AvrBs3 homologs revealed that these TALEs contain repeat domains of various lengths and fall into six groups based on their RVD sequences. Schandry et al. (2016) showed that TALEs in *Ralstonia solanacearum* could be similarly grouped into subclasses based on RVD sequences.

This study showed that the RVDs in these AvrBs3 homologs were nearly perfectly conserved within a group. Moreover, N and HDN residues make up a conserved core in all of the examined *Xcg* TALEs. The presence of HDN in the core of *Xcg* TALEs is similar to the finding that HDN residues form a conserved core in almost all Avr/Pth proteins in *X. oryzae* pv. *oryzae*, including in AvrBs3 homologs (Wu et al., 2007). The *Xcg* TALE classes with 22.5 repeats and 33.5 repeats had two and four HDHDN repeats, respectively, which suggests that they may have resulted from a duplication event. An analysis of 113 TALEs showed that TALEs with 17.5 repeats were the most common, followed by those with 19.5 and 15.5 repeats (Boch et al., 2010). Our results with the *Xcg* strains were similar in that TALEs with 17.5 repeats were the most common, followed by those with 19.5 and 18.5 repeats; all of the *Xcg* strains had at least one TALE with 18.5 repeats. Also, we found that the phylogenetic tree of TALEs based on RVDs divided the TALEs with 17.5 repeats into two clades, suggesting that two distinct 17.5-repeat-containing TALEs have evolved; these may exhibit specificity to distinct soybean cultivars.

Genes for the *Xcg* TALEs with 19.5 repeats were present on plasmids in the *Xcg* strains (Table 5). Since 85% of isolates from Korea contained genes necessary for plasmid transfer and mobilization (Kim et al., 2006), a plasmid with a TALE gene could potentially transfer it to other strains, with subsequent integration of the TALE gene into the chromosome. Gochez et al. emphasized that since TALE genes are located on plasmids in *Xcg* strains, the variability of TALEs and the shuffling of plasmids among the pathogen population may contribute to pathogen variability and adaptive evolution of the host. An analysis of the variability and distribution of TALEs on plasmids in more strains of *Xcg* may provide a better understanding of the evolution of TALEs in this pathogen.

Following the secretion of TALEs into the cells of a plant host, TALEs may enter the host cell nucleus and activate specific genes by binding to effector binding elements (EBEs) in the promoter [39]. Accurate prediction of EBEs is essential for the robust design of artificial TALE DNA-binding domains in biotechnological applications, and for gaining in-sights into the host cell TALE targets. In this study, we used the sequence of the RVDs to predict the EBEs of the *Xcg* TALEs. We found that, surprisingly, the predicted EBEs of the *Xcg* TALEs contained only A, T and C bases, but not G; this contrasts with the EBEs in other genera such as *Xanthomonas oryzae* pv. *oryzae*, *Xanthomonas oryzae* pv. *oryzicola* (Boch et al, 2010). Another study found that EBE diversity from four *Ralstonia solanacearum* pathotypes was limited (Castañeda-Ojeda et al., 2017). We also found that the EBEs of the *Xcg* TALEs formed six distinct groups that had domains that were intermixed and targeted by multiple TALEs, suggestive of recombination events occurring during the evolution of these AvrBs3 homologs.

In summary, we generated whole-genome sequence assemblies of six *Xcg* strains that exhibit various levels of virulence in soybean. We did not find significant differences among these strains in the genes falling into distinct functional classes or in their carbo-hydrate-degrading enzymes. We found that one strain, the US strain 8ra, had two more T3SEs than the five Korean strains examined, and that overall, the greatest variation in their T3SEs was in the TALE effectors, namely in the AvrBs3 homologs. In looking at the variability in the structure and predicted binding domains of these AvrBs3 homologs, we found multiple classes that vary in the number of repeat regions; furthermore, variation in the predicted binding domains suggests recombination, domain swapping, and likely potential adaptation of these TALEs to their hosts. This comparative genome analysis thus highlights a high overall similarity among *Xcg* strains and a high level of variation in TALE effectors suggestive of duplication and subsequent

adaptation to distinct soybean cultivars.

LITTERATURE CITED

Athinuwat, D., Prathuangwong, S., Cursino, L., Burr, T. 2009. *Xanthomonas axonopodis* pv. *glycines* Soybean Cultivar Virulence Specificity Is Determined by *avrBs3* Homolog a*vrXg1*. *Phytopathol*. 99: 996-1004.

Barak, J. D., Vancheva, T., Lefeuvre, P., Jones, J. B., Timilsina, S., Minsavage, G. V. and Koebnik, R. 2016. Whole-genome sequences of *Xanthomonas euvesicatoria* strains clarify taxonomy and reveal a stepwise erosion of type 3 effectors. *Front. Plant Sci.* 7: 1805.

Bardou, P., Mariette, J., Escudié, F., Djemiel, C. and Klopp, C. 2014. Jvenn: An interactive Venn diagram viewer. *BMC Bioinf.* 15: 293.

Boch, J., Bonas, U. 2010. *Xanthomonas* AvrBs3 family-type III effectors: Discovery and function. *Annu. Rev. Phytopathol*. 48: 419-436.

Buttner, D. and He, S. Y. 2009. Type III protein secretion in plant pathogenic bacteria. *Plant Physiol.* 150: 1656-1664.

Büttner, D., Bonas, U. 2010. Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol. Rev.* 34: 107-133.

Carpenter, S.C., Kladsuwan, L., Han, S.W., Prathuangwong, S. and Bogdanove, A.J. 2019. Complete genome sequences of *Xanthomonas axonopodis* pv. *glycines* isolates from the United States and Thailand reveal conserved transcription activator-like effectors. *Genome Biol. Evol.* 11: 1380-1384.

Castañeda-Ojeda, M.P., Moreno-Pérez, A., Ramos, C. and López-Solanilla, E. 2017. Suppression of plant immune responses by the *Pseudomonas savastanoi* pv. *savastanoi* NCPPB 3335 type III effector tyrosine phosphatases HopAO1 and HopAO2. *Front. Plant Sci.* 8: 680.

Chatnaparat, T., Prathuangwong, S., Ionescu, M. and Lindow, S.E. 2012. XagR, a LuxR Homolog, contributes to the virulence of *Xanthomonas axonopodis* pv. *glycines* to soybean. *Molecular Plant-Microbe Interact*. 25: 1104-1117.

Constantin, E. C., Cleenwerck, I., Maes, M., Baeyen, S., Van Malderghem, C., De Vos, P. and Cottyn, B. 2016. Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. *Plant Pathol*. 65: 792-806.

Darling, A.C., Mau, B., Blattner, F.R. and Perna, N.T. 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 14: 1394-1403.

Darrasse, A., Bolot, S., Serres-Giardi, L., Charbit, E., Boureau, T., Fisher-Le Saux, M. and Noel, L.D. 2013. High-quality draft genome sequences of *Xanthomonas axonopodis* pv. *glycines* strains CFBP 2526 and CFBP 7119. *Genome Announc.* 1: e01036-13.

Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26: 2460-2461.

Gochez, A.M., Huguet-Tapia, J.C., Minsavage, G.V., Shantaraj, D., Jalan, N., Strauß, A. and Potnis, N. 2018. Pacbio sequencing of copper tolerant *Xanthomonas citri* reveals presence of a chimeric plasmid structure and provides insights into reassortment and shuffling of

transcription activator-like effectors among *X. citri* strains. *BMC Genom.* 19: 16.

Grau, J., Reschke, M., Erkes, A., Streubel, J., Morgan, R.D., Wilson, G.G. and Boch, J. 2016. AnnoTALE: Bioinformatics tools for identification, annotation, and nomenclature of TALEs from *Xanthomonas* genomic sequences. *Sci. Rep.* 6: 21077.

Grau, J., Wolf, A., Reschke, M., Bonas, U., Posch, S. and Boch, J. 2013. Computational predictions provide insights into the biology of TAL effector target sites. *PLoS Comput. Biol.* 9: e1002962

Gu, K., Yang, B., Tian, D., Wu, L., Wang, D., Sreekala, C., Yang, F., Chu, Z., Wang, G.L., White, F.F. and Yin, Z. 2005. R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature* 435: 1122-1125.

Hong, S. J., Kim, Y. K., Jee, H. J., Lee, B. C., Yoon, Y. N. and Park, S. T. 2010. Selection of bactericides for controlling soybean bacterial pustule. *Res. Plant Dis.* 16: 266–273. (In Korean)

Hummel, A.W., Wilkins, K.E., Wang, L., Cernadas, R.A. and Bogdanove, A.J. 2017. A transcription activator-like effector from *Xanthomonas oryzae* pv. *oryzicola* elicits dose-dependent resistance in rice. *Mol. Plant Pathol.* 18: 55-66.

Hwang, I., Lim, S.M. 1998. Pathogenic variability in isolates of *Xanthomonas campestris* pv. *glycines*. *Korean J. Plant Pathol.* 14:19-22.

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W. and Hauser, L.J. 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf.* 11: 119.

Ji, Z.Y., Zakria, M., Zou, L.F., Xiong, L., Li, Z. and Ji, G.H.; Chen, G.Y. 2014. Genetic diversity

of transcriptional activator-like effector Genes in Chinese isolates of *Xanthomonas oryzae* pv. *oryzicola*. *Phytopathol*. 104: 672-682.

Kaewnum, S., Prathuangwong, S., Burr, T.J. 2005. Aggressiveness of *Xanthomonas axonopodis* pv. *glycines* isolates to soybean and hypersensitivity responses by other plants. *Plant Pathol.* 54: 409-415.

Kim, J. G., Choi, S., Oh, J., Moon, J.S. and Hwang, I. 2006. Comparative analysis of three indigenous plasmids from *Xanthomonas axonopodis* pv. *glycines*. *Plasmid* 56: 79-87.

Lee, B. M., Park, Y. J., Park, D. S., Kang, H. W., Kim, J. G., Song, E. S. and Go, S. J. 2005. The genome sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice. *Nucleic Acids Res.* 33: 577-586.

Lee, J.H., Shin, H., Park, H.J., Ryu, S. and Han, S.W. 2014. Draft genome sequence of *Xanthomonas axonopodis* pv. *glycines* 8ra possessing transcription activator-like effectors used for genetic engineering. *J. Biotechnol.* 179: 15-16.

Moscou, M.J., Bogdanove, A.J. 2009. A simple cipher governs DNA recognition by TAL effectors. *Science* 326: 1501.

Narvel, J.M., Jakkula, L.R., Phillips, D.V., Wang, T., Lee, S.H. and Boerma, H.R. 2001. Molecular mapping of *rxp* conditioning reaction to bacterial pustule in soybean. *J. Hered.* 92: 267-270.

Oh, C.S., Lee, S. and Heu, S. 2011. Genetic diversity of avrBs-like Genes in three different *Xanthomonas* species isolated in Korea. *Plant Pathol. J.* 27: 26-32.

Park, H.J., Han, S.W., Oh, C., Lee, S., Ra, D., Lee, S.H., Heu, S. 2008. Avirulence gene

diversity of *Xanthomonas axonopodis* pv. *glycines* isolated in Korea. *J. Microbiol. Biotechnol.* 18: 1500-1509.

Potnis, N., Krasileva, K., Chow, V., Almeida, N. F., Patil, P. B., Ryan, R. P. and Jones, J. B. 2011. Comparative genomics reveals diversity among *Xanthomonads* infecting tomato and pepper. *BMC genomics*. 12: 1-23.

Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J. and Jensen, L.J. 2014. EggNOG v4. 0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.* 42: D231-D239.

Schandry, N., de Lange, O., Prior, P. and Lahaye, T. 2016. TALE-like effectors are an ancestral feature of the *Ralstonia solanacearum* species complex and converge in DNA targeting specificity. *Front. Plant Sci.* 7: 1225.

Scholze, H., Boch, J. 2011. TAL effectors are remote controls for gene activation. *Curr. Opin. Microbiol.* 14: 47-53.

Tatusov, R.L., Galperin, M.Y., Natale, D.A. and Koonin, E.V. 2000. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28: 33-36.

Vieira, P. S., Bonfim, I. M., Araujo, E. A., Melo, R. R., Lima, A. R., Fessel, M. R. and Murakami, M. T. 2021. Xyloglucan processing machinery in *Xanthomonas* pathogens and its role in the transcriptional activation of virulence factors. *Nat. Commun.* 12: 1-15.

Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L. and Machi, D. 2013. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* 42: D581-D591.

Weng, S.F., Luo, A.C., Lin, C.J. and Tseng, T.T. 2017. High-quality genome sequence of *Xanthomonas axonopodis* pv. *glycines* strain 12609 isolated in Taiwan. *Microbiol. Resour. Announc.* 5: e0169516.

Wu, X.M., Li, Y.R., Zou, L.F. and Chen, G.Y. 2007. Gene-for-gene relationships between rice and diverse avrBs3/pthA avirulence genes in *Xanthomonas oryzae* pv. *oryzae*. *Plant Pathol.* 56: 26-34.

Yang, B. and White, F.F. 2004. Diverse members of the AvrBs3/PthA family of type III effectors are major virulence determinants in bacterial blight disease of rice. *Molecular Plant-Microbe Interact*. 17: 1192-1200.

CHAPTER II

Characterization of *Xanthomonas citri* pv. *glycines* population genetics and virulence in a national survey of bacterial pustule of soybean in Korea

Contents of this chapter have been published in The Plant Pathology Journal (Kang et al. 2021. Characterization of *Xanthomonas citri* pv. *glycines* population genetics and virulence in a national survey of bacterial pustule disease in Korea. *Plant Pathol. J.* 37: 652-661).

ABSTRACT

Xanthomonas citri pv. *glycines* (*Xcg*) is a major pathogen of soybean [*Glycine max* (L.) Merr.] in South Korea, despite the availability of soybean varieties with some resistance. We conducted a nationwide survey of the incidence and severity of bacterial pustule caused by *Xcg*. The percentage of infected fields was 7 to 17% between 2015 and 2017. We characterized the diversity of a nationwide collection of 106 *Xcg* isolates based on *avrBs3* banding patterns. The isolates fell into eleven groups, each represented by a type strain; only two of these were similar to isolates collected in 1999 to 2002. The diversity of *Xcg* strains increased and the dominant strains changed between 1999 and 2017, with three new type strains comprising 44% of the isolates examined in 2012 to 2017. Pathogenicity tests did not show evidence for a shift in the races or aggressiveness of *Xcg* strains. Korean soybean cultivars, including the widely-grown Daewon cultivar, were susceptible to the eleven new type strains. The cultivar CNS, which carries the *rxp* resistance gene, was susceptible to most type strains, including two representing 83% of the Korean *Xcg* strains. In contrast, Williams 82, which also carries *rxp*, showed resistance to at least five type strains. Collectively, these results suggest that Williams 82 has resistance loci in addition to *rxp*. The widespread distribution of *Xcg*, the high virulence of the current endemic strains, and the low resistance of most Korean soybean cultivars collectively favor widespread disease in Korea in years that are favorable to pustule development.

INTRODUCTION

Bacterial pustule caused by *Xanthomonas citri* pv. *glycines* (*Xcg*), previously classified as *X. axonopodis* pv. *glycines* (Constantin et al., 2016), is an important disease of soybean and has been reported in most soybean-growing areas of the world (Hartman et al., 2015). The effect of bacterial pustule on the growth and development of soybean largely depends on infection levels, with high infection levels decreasing yields by 20 to 60% following natural or artificial infection (Hokawat et al., 1991; Shukla, 1994; Hong et al., 2011). In Korea, bacterial pustule is one of the most common bacterial diseases of soybean. It occurs nationwide and can be observed any time from July to September. A survey of the distribution of bacterial pustule in Korea in 1997 and 1998 found the disease in 85 of 99 fields examined (Lee et al. 1999), including in fields in the provinces of Gyeonggi, Gangwon, Chungbuk, Chungnam, Jeonbuk, and Gyeongbuk. A similar level of disease incidence, 89.7%, was observed in a survey conducted in 2005 and 2006 in the southern regions of Korea (Hong et al., 2010). Here we provide an updated nationwide survey on the occurrence and severity of bacterial pustule in Korea; this was last examined in 1998 (Lee et al. 1999).

Although bactericides such as copper can be used to control bacterial pustule, bactericides are expensive as a management tool for a field crop and risk enriching for resistant populations of the pathogen and other non-target bacteria. Soybean varieties that are resistant to bacterial pustule are known and have been used to effectively manage this disease. However, only one gene, *rxp* (resistance to *Xanthomonas phaseoli*), is currently known to confer resistance in soybean, and this resistance is only partial (Bernard, 1973). The *rxp* gene was originally found

as a recessive resistance gene in the cultivar CNS (Hartwig and Lehman, 1951). This gene has been introduced into many commercial soybean cultivars, including the popular cultivars Williams and Williams 82 that were developed by the United States Department of Agriculture (Hartwig and Lehman, 1951; Narvel et al., 2001). The *rxp* gene has not yet been cloned, and the availability of bacterial pustule-resistant commercial varieties is still limited. To date, the incorporation of *rxp* into Korean soybean cultivars has not been reported, and the *Xcg* avirulence gene that triggers *rxp*-mediated resistance has not been identified.

The successful deployment of resistant soybean varieties in a geographical region requires knowledge of the population genetics of the pathogen in that region. Strains of the bacterial pustule pathogen have been shown to fall into distinct races (Athinuwat et al., 2009; Hwang et al., 1998; Park et al., 2008) and to differ in their aggressiveness (Kaewnum et al., 2005). Hwang et al. (1998) examined the ability of 63 *Xcg* isolates to induce disease on a set of eleven differential soybean cultivars, including Chippewa, Harosoy, Mukden, Pella, and Williams, and on this basis classified the isolates into five races, designated 1, 2, 3, 4 and 5. The molecular basis of this race specificity, including the identity of potential bacterial effectors interacting with resistance genes, is not yet known. Similarly, Kaewnum et al. (2005) demonstrated that 26 *Xcg* isolates from Thailand differ in their aggressiveness in inducing disease on soybean cultivars and in their ability to induce a hypersensitive response on a range of plant species, including tobacco, cucumber, pea, and sesame. Similar to the previous study, the *Xcg* genes that influence differential aggressiveness on these cultivars are not known. Lastly, Park et al. (2008) classified 155 *Xcg* isolates from Korea into at least six groups based on the number and size of the genomic fragments hybridizing with an avrBs3 family gene probe; such differences in avrBs3 content could be correlated with resistance and/or aggressiveness. This work

identified six type strains that represent these groups. Collectively, these studies highlight natural variation among *Xcg* strains that could underly population and/or genetic shifts that impact the successful use of soybean varieties resistant to bacterial pustule.

Although commercial soybean varieties with at least partial resistance to bacterial pustule have been developed for other areas of the world, this disease remains a problem for soybean cultivation in Korea. The objectives of this study were (i) to survey the current distribution of bacterial pustule throughout Korea, (ii) to evaluate if new groups of *Xcg* isolates as defined by avrBs3 profiles have emerged in Korea since the six type strains were identified in 2008, and (iii) if new avrBs3 groups are identified, to evaluate if they represent new races of Korean *Xcg* isolates. This study is the first to characterize the diversity of Korean *Xcg* strains since 2008 and is intended to provide guidance in developing strategies to deploy soybean varieties resistant to bacterial pustule caused by *Xcg* in Korea.

MATERIALS AND METHODS

1. Survey of disease occurrence

Disease surveys were conducted in 2015, 2016 and 2017 to estimate the incidence and severity of bacterial pustule in soybean fields throughout the Republic of Korea. Fields were selected in at least three regions in each of eight provinces; the regions chosen were those with the greatest area of cultivated soybeans in each of the provinces. Selected fields had soybeans grown in an area over 1,000 m^2. Fields were surveyed once from August to September during each year of the survey. Each field was assessed for the presence of bacterial pustule disease, with a field designated as infected if even one symptomatic plant was observed. For each infected field, the infection rate of bacterial pustule was estimated as the percentage of the total number of plants that exhibited bacterial pustule symptoms. The mean infection rate across all of the fields in a region across all survey dates was determined. For each infected field, the disease severity was estimated as the percentage of the total leaf surface area that was covered with bacterial pustule symptoms. Our methods for estimating disease incidence and severity were in accordance with the Agricultural Science and Technology Research and Analysis Standards of the Rural Development Administration (RDA, 2012).

2. Collection and identification of isolates

To generate a collection of *Xcg* isolates, we collected leaves exhibiting bacterial pustule symptoms from every infected field. Excised pustule-containing leaf tissues were

surface sterilized with 1% hypochlorite for 90s, rinsed once in sterile water, and then macerated with 1 ml of sterile water. The suspension was streaked onto a tryptic soy agar (TSA) medium, and colonies with the characteristic morphology of *Xcg* were selected for isolation. To identify the bacterial pustule pathogens, putative *Xcg* isolates were identified based on the sequences of the 16S ribosomal RNA (rRNA) gene, DNA gyrase subunit B gene (*gyrB*) and the RNA polymerase beta subunit gene (*rpoB*); the primers are shown in Supplementary Table 1. In addition, the isolates were evaluated using the GEN III MicroPlate system (Biolog Inc., Hayward, CA). In short, the isolates were grown for 48 h at 30℃ under aerobic conditions on Biolog Universal growth medium without blood. The identity of the isolates was determined using the manufacturer's instructions with inoculating fluid A, incubation for 24 h at 25°C, and OmniLog MicroArray Data Collection Software 1.2. Isolates that were verified to be *Xcg* were cultured in tryptic soy broth (TSB) with shaking at 28℃ for 24 h and preserved in 20% glycerol (v/v) at –72℃.

3. Isolation of genomic DNA

To extract genomic DNA, *Xcg* cells grown on TSA medium were subcultured into TSB medium (casein peptone 17 g, soya peptone 3 g, sodium chloride 5 g, dipotassium hydrogen phosphate 2.5 g, glucose 2.5 g in 1 liter distilled water) in a 28℃ shaking incubator for 24 h. Bacterial DNA was extracted from 1 ml of cells grown to an optical density at 600 nm (OD_{600}) of 0.8 to 1.0 using a commercial genomic DNA extraction kit (iNtRON Biotechnology, Inc., Seongnam, Korea) and the protocol for Gram-negative bacteria. The concentration of extracted DNA was estimated using a spectrophotometer (BioDrop Duo,

Biochrom Ltd., Cambridge, U.K.).

4. Southern blot analysis of *avrBs3* repeats

For Southern blot analysis, genomic DNA of *Xcg* was digested with the restriction enzyme *Bam*HI (New England Biolabs, Ipswich, MA, U.S.). Electrophoretic separation and transfer to nylon membranes were performed as described (Sambrook et al., 1989). For DNA hybridization, a digoxigenin (DIG)-labeled *avrBs3* gene probe was used; this probe was comprised of a 3.3-kb internal *Bam*HI fragment of *avrBs3* cloned into pBlueScript, as described previously (Park et al., 2008), using a commercial protocol for DIG-labeled probes and membrane hybridization (Roche Diagnostics, Switzerland). Detection of hybridization was performed using an anti-DIG antibody (Roche Diagnostics, Switzerland) and subsequent exposure to an X-ray film (Agfa CP-BU new).

5. Assay for resistance to bacterial pustule disease

Seeds of soybean (*Glycine max* L.) of American and Korean varieties were obtained from the Korean Agricultural Culture Collection at the RDA in Korea. Plants were grown in the greenhouse to vegetative stage 3. *Xcg* strains were grown on TSA for 2 days at 28℃, then colonies were suspended in sterilized 10 mM $MgCl_2$ and adjusted to an OD_{600} of 0.5, which corresponded to a cell density of ~5×10^8 CFU/ml. Bacterial suspensions were inoculated into the fully expanded two leaves by spraying with an air compressor (SMH10L, SMASH, China) with a 2.5-KPa injection pressure. All plant inoculation experiments were repeated 3 times, and

each replicate consisted of three plants. Disease was assessed at 7 days post-inoculation. Disease severity was assessed on three leaflets per plant using a disease severity index that was scored, on a per leaflet basis, as 0 for no pustules, 1 for 1-25 pustules, 2 for 26-50 pustules, 3 for 51-75 pustules, 4 for 76-100 pustules, and 5 for >100 pustules. To provide a conservative estimate of resistance, we reported the average of the two highest severity indices scored.

6. Assay for induction of the hypersensitive response (HR)

Isolates were tested for their ability to induce an HR on four species of tobacco (*Nicotiana rustica* L., *Nicotiana benthamiana* L., *Nicotiana tabacum* L. cv. Xanthi and *Nicotiana tabacum* L. cv. Samsun), pepper (*Capsicum annuum* L. cv. Supermanita), sesame (*Sesamum indicum* L. cv. Milsung) and tomato (*Solanum lycopersicum* M. cv. Berry King F). Inocula of *Xcg* strains were prepared as described above for the resistance assay, but were adjusted to an OD_{600} of 0.3 (~5×10^7 CFU/mL). Bacterial suspensions were infiltrated into the leaf mesophyll using a 1-mL hypodermic syringe without a needle. Infiltrated zones were observed for development of tissue collapse and cell death within 24 h post-infiltration. A negative response was indicated by some yellowing but no visible collapse or cell death in the infiltrated zone, whereas a positive response was indicated by complete collapse and cell death in the infiltrated.

RESULTS

1. Recent trends in the disease incidence and severity of bacterial pustule in Korea

A nationwide survey of the incidence and severity of bacterial pustule caused by *X. citri* pv. *glycines* was conducted from 2015 to 2017. Bacterial pustule was observed widely across Korea (Table 1). Nationwide, the incidence of bacterial pustule was similar in 2015 and 2016, but increased in 2017 (Table 2). In particular, the percentage of fields nationwide that were infected increased from 7 to 8% in 2015 and 2016, respectively, to over 17% in 2017, thus the number of fields with bacterial pustule doubled in 2017. Whereas the percentage of fields with bacterial pustule increased in Korea in 2017, the average within-field infection rate per region did not (Table 2). The average disease severity in those fields that were infected increased slightly (Table 2). Increased incidence of bacterial pustule was also observed when the percentage of infected fields was examined on a per-province basis (Table 3). The average within-field infection rate did not parallel the increased percentage of fields with bacterial pustule when examined on a per-province basis, and the severity increased slightly (Table 3). Among the provinces, Gyeonggi was notable for the dramatic increase in bacterial pustule in 2017, having had almost no bacterial pustule detected in the nearly 40 fields examined in 2015 and 2016 (Table 1). Nationwide, however, the endemic nature of bacterial pustule was demonstrated by the detection of the disease every year of the survey in 27% of the regions that were surveyed in all three years.

The increase in disease incidence in 2017 could be due to climatic factors, shifts in the pathogen populations in Korea, shifts in soybean varieties used, and a combination of these or

other factors. Bacterial pustule is generally favored by high temperatures and frequent precipitation (Hartman et al., 2015; Moffett and Croft, 1983; Bradbury, 1986). Bacterial pustule in Korea can be observed from July when the plants are in vegetative stages to September when the plants have passed flowering, with disease severity often correlated with the amount of precipitation in August (Hong et al., 2011). Weather data collected in 2015, 2016, and 2017 indicate that precipitation across Korea in August of 2017 was much higher than in August of 2015 and 2016, with 241 mm of total precipitation in 2017 versus only 111.3 and 76.2 mm in 2015 and 2016, respectively (https://data.kma.go.kr/cmmn/main.do). Thus, conditions conducive to disease are likely to be one factor contributing to the high incidence of disease in 2017. Below, I explore the possibility of population shifts in the *Xcg* pathogen.

Table 1. Bacterial pustule incidence and severity in three annual surveys in Korea.

Province	Region	2015				2016				2017			
		No. fields[a]	No. infected fields[a]	Infection Rate[b]	Severity[c]	No. fields	No. infected fields	Infection Rate	Severity	No. fields	No. infected fields	Infection Rate	Severity
Gyeonggi	Yeoncheon	15	0	-	-	15	1	10%	+	15	1	10%	+
	Suwon	12	0	-	-	5	0	-	-	15	4	10%	+++
	Paju	10	0	-	-	10	0	-	-	15	2	30%	++
	Pocheon	0	-	-	-	10	0	-	-	7	0	-	-
Gangwon	Yeongwol	15	2	10%	++	15	0	-	-	15	2	10%	+++
	Pyeongchang	17	2	10%	++	17	0	-	-	15	0	-	-
	Hoengseong	5	0	-	-	0	-	-	-	0	-	-	-
	Inje	10	1	10%	+	0	-	-	-	0	-	-	-
	Jeongseon	0	-	-	-	5	1	10%	++	5	1	10%	++
Chungbuk	Goesan	10	4	50%	++++	10	0	-	-	15	2	10%	++
	Danyang	20	2	10%	+	20	0	-	-	10	0	-	-
	Jecheon	30	1	10%	++	20	1	10%	+	10	3	10%	++
	Chungju	3	3	10%	++	3	1	10%	++	5	1	10%	++
Chungnam	Seosan	10	2	15%	++	8	1	10%	++	10	2	10%	++
	Taean	20	0	-	-	15	1	10%	+	20	3	10%	+++
	Gongju	20	0	-	-	10	2	10%	+	15	3	10%	++
	Hongseong	0	-	-	-	3	0	-	-	10	0	-	-
	Cheongyang	10	0	-	-	4	1	10%	++	10	3	10%	++
Jeonbuk	Gimje	10	0	-	-	12	1	10%	++	20	6	15%	+++
	Sunchang	10	1	10%	+	10	2	10%	++	20	4	10%	+++
	Gochang	10	0	-	-	10	2	10%	++	10	3	10%	++
	Jeongeup	0	-	-	-	0	-	-	-	10	2	10%	++
Jeonnam	Jangheung	5	2	25%	++	0	-	-	-	10	3	10%	+++
	Muan	58	4	10%	++	25	3	30%	++++	25	8	30%	++
	Hampyung	0	-	-	-	10	1	10%	++	0	-	-	-
Gyeongbuk	Andong	20	1	10%	++++	15	1	10%	++	20	2	10%	++
	Cheongdo	10	1	10%	++	0	-	-	-	0	-	-	-
	Chilgok	6	0	-	-	0	-	-	-	0	-	-	-
	Gyeongsan	5	0	-	-	0	-	-	-	0	-	-	-
	Mungyung	0	-	-	-	15	1	30%	++++	25	3	30%	++++
	Sangju	0	-	-	-	5	0	-	-	8	0	-	-

	Yecheon	10	0	-	-	10	1	10%	++	10	2	10%	++
Gyeongnam	Hapcheon	10	0	-	-	7	1	20%	+++	7	1	10%	++
	Geochang	5	0	-	-	15	1	20%	++++	8	1	10%	++
	Changyeong	6	0	-	-	5	1	20%	+++	7	1	10%	++

[a] Number of fields examined, and Number of infected fields.

[b] Infection rate is the average infection rate (% of plants infected) in infected fields in a region.

[c] Severity reflects the average percentage of leaf area that was symptomatic in infected fields in a region, where + is 1-10%, ++ is 11-30%, +++ is 30-50%, and ++++ is >50%.

Table 2. Summary of annual survey data of bacterial pustule in Korea.

Parameter	**2015**	**2016**	**2017**
Total no. of fields examined	372	309	372
Disease incidence			
Total no. of infected fields in Korea	26	24	63
% of infected fields in Korea	7.0	7.8	17.0
Within-field infection rate per region[a]	12.7 ± 3.0	13.7 ± 1.6	12.7 ± 1.4
Disease severity			
Within-field disease index per region[a]	1.8 ± 0.3	2.2 ± 0.2	2.3 ± 0.1

[a] The mean within-field infection rate and mean within-field disease index per region were determined only for fields with the disease. Data shown are the mean ± standard error of the mean. The mean within-field disease index was calculated by converting the severity data shown in Table 1 to numerical scores of 1 to 4, where 1 represents + and 4 represents ++++.

Table 3. Summary of annual survey data of bacterial pustule in Korea by province[a].

	No. fields			% infected fields			Within-field infection rate			Severity		
Province	2015	2016	2017	2015	2016	2017	2015	2016	2017	2015	2016	2017
Gyeonggi	37	40	52	--	2.5	13.5	--	10	17	--	1.0	2.0
Gangwon	47	37	35	10.6	2.7	8.6	10	10	10	1.7	2.0	2.5
Chungbuk	63	53	40	15.9	3.8	15.0	20	10	10	2.3	1.5	2.0
Chungnam	60	40	65	3.3	12.5	16.9	15	10	10	2.0	1.5	2.3
Jeonbuk	30	32	60	3.3	15.6	25.0	10	10	11	1.0	2.0	2.5
Jeonnam	63	35	35	9.5	11.4	31.4	18	20	20	2.0	3.0	2.5
Gyeongbuk	51	45	63	3.9	6.7	11.1	10	17	17	3.0	2.7	2.7
Gyeongnam	21	27	22	--	11.1	13.6	--	20	10	--	3.3	2.0
Means	*46.5*	*38.6*	*46.5*	*5.8*	*8.3*	*16.9*	*13.8*	*13.3*	*13.1*	*2.0*	*2.1*	*2.3*
± SE	*± 5.6*	*± 2.8*	*± 5.6*	-	-	-	-	-	-	*± 0.3*	*± 0.3*	*± 0.1*

[a] Data shown are the sum of the total no. of fields, the percentage of the total number of fields that were infected, and the within-field infection rate and severity for those fields showing disease, where the latter were calculated by converting the severity data shown in Table 1 to numerical scores of 1 to 4, where 1 represents + and 4 represents ++++.

Table 4. Type strains of the *X. citri* pv. *glycines* isolates obtained in this study[a].

Strain name	**No. *avrBs3* repeats**	**% of isolates**	**Origin**
K1	7	0.9	Goesan-gun, Chungcheongbuk-do, isolated in 2017
K10	6	35	Taean-gun, Chungcheongnam-do, isolated in 2017
K13	4	0.9	Jangheung-gun, Jeollanam-do, isolated in 2017
K20	5	0.9	Cheongyang-gun, Chungcheongnam-do, isolated in 2016
K29[b]	5	48	Jecheon-si, Chungcheongbuk-do, isolated in 2015
K34	5	3.8	Goesan-gun, Chungcheongbuk-do, isolated in 2015
K46	4	0.9	Pyeongchang-gun, Gangwon-do, isolated in 2015
K47	4	4.7	Seosan-si, Chungcheongnam-do, isolated in 2015
K53	5	1.9	Inje-gun, Gangwon-do, isolated in 2015
K79	6	0.9	Sacheon-si, Gyeongsangnam-do, isolated in 2013
K100[c]	6	1.9	Cheongyang-gun, Chungcheongnam-do, isolated in 2016

[a] All of the strains were pathogenic on soybean.

[b] Type strain K29 is the same as type strain OcsF (Park et al., 2008).

[c] Type strain K100 is the same as type strain SL1017 (Park et al., 2008).

2. *Xcg* strain diversity based on profiles of *avrBs3* repeats in the genome

Strains of *Xcg* have been found to carry multiple genes encoding homologs of the AvrBs3 effector family (Park et al. 2008; Kim et al. 2006). Effectors in this family are transcription activator-like (TAL) effectors that are secreted by the bacterial type III secretion system and influence the expression of host genes involved in disease resistance or susceptibility (Hueck, 1998). The presence of multiple copies of this gene family in *Xcg* may be due to their association with plasmids and mobile genetic elements (Kim et al. 2006). The presence of multiple copies enables strain profiling based on the size and number of genomic fragments that contain a copy of an avrBs3 allele. To investigate whether the diversity of the *Xcg* strains in soybean cultivation fields in Korea have changed since isolates collected in 1999 to 2002 were examined (Park et al. 2008), we characterized total 106 *Xcg* strains including the strains isolated from >150 bacterial pustule leaf lesions collected during our nationwide survey from 2015 to 2017 (Supplementary Table 2). These isolates were confirmed to be *Xcg* strains using Biolog data (Supplementary Table 3) and phylogenetic analyses based on 16S rRNA, *gyrB* and *rpoB* gene sequences (Supplementary Fig. 1). In contrast to Park et al. (2008) who identified six *Xcg* groups based on avrBs3 banding patterns, we found that the 106 strains examined fell into eleven groups (Fig. 1). We selected strains from each group to serve as type strains (Fig. 1, Table 4). Only two of these groups, K29 and K100, had a similar *avrBs3* banding pattern to previously identified groups (Table 4).

In addition to an increase in the number of strain groups identified between 1999-2002 and 2012-2017, we identified a major shift in the relative abundance of these strain groups (Fig. 2). The K29 group exhibited the same *avrBs3* banding pattern as the previously identified OcsF group (Park et al., 2008). In both survey periods, strains with this banding pattern were

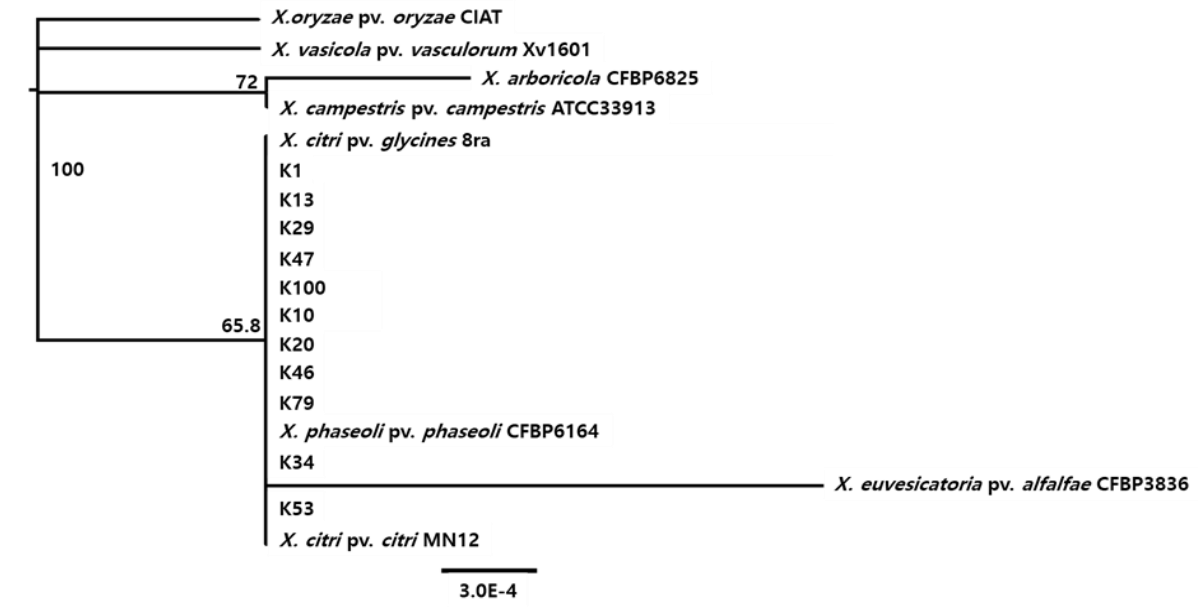
predominant among the *Xcg* isolates; however, the proportion of these isolates decreased from over 80% in 1999-2002 to 48% in 2012-2017 (Fig. 2 and Table 4). The K100 group also exhibited the same avrBs3 banding pattern as a previous group, the SL1017 group, but this group was present at a low relative abundance, ≤ 3%, in both survey periods. Notably, a new strain group designated K10, emerged as a second predominant group in 2012-2017; this group comprised 35% of the isolates in the most recent survey and was not identified in the isolates collected in 1999-2002. Three new type strains, K10, K47, and K34, collectively comprised 44% of the isolates examined, demonstrating a major shift in the pathogen population over about a two-decade period. The K10 group may have been present at a low frequency in the earlier survey period. Whether it was present in that period or emerged in the intervening years, the high abundance of this strain group indicates that it was favored by the soybean genotypes, cultivation conditions, and/or climatic conditions in Korea.

Figure 1. Distinguishing features for *Xanthomonas citri* (PG I), based on Biolog GEN III data. Values are the percentage of intermediate and positive (i.e. / or +) strain.

	K29	K34	K46	K53	K47	K100	K1	K10	K13	K20	K79	8ra
D-melibiose	26(-)	29(-)	21(-)	34(-)	23(-)	34(-)	30(-)	31(-)	26(-)	29(-)	31(-)	43(-)
***N*-Acetyl-D-glucosamine**	80(+)	93(+)	75(+)	86(+)	69(+)	71(+)	68(+)	62(+)	49(+)	73(+)	84(+)	85(+)
α-D-glucose	59(+)	90(+)	75(+)	83(+)	55(+)	68(+)	69(+)	67(+)	49(+)	70(+)	85(+)	83(+)
Glycerol	21(-)	31(-)	33(-)	33(-)	19(-)	20(-)	21(-)	23(-)	16(-)	19(-)	27(-)	33(-)
D-glucose-6-PO_4	14(-)	17(-)	16(-)	20(-)	19(-)	19(-)	23(-)	25(-)	18(-)	20(-)	18(-)	21(-)
D-fructose-6-PO_4	22(-)	22(-)	25(-)	23(-)	24(-)	26(-)	27(-)	38(-)	22(-)	23(-)	25(-)	29(-)
Rifamycin.SV	100	100	100	100	100	100	92(+)	71(+)	74(+)	100	100	100
Gelatin	78(+)	98(+)	65(+)	93(+)	69(+)	64(+)	76(+)	56(+)	44(+)	76(+)	88(+)	86(+)
D-saccharic acid	16(-)	16(-)	16(-)	18(-)	15(-)	17(-)	17(-)	16(-)	17(-)	15(-)	16(-)	18(-)
L-lactic acid	23(-)	24(-)	22(-)	29(-)	19(-)	25(-)	28(-)	18(-)	22(-)	22(-)	28(-)	28(-)
Nalidixic acid	12(-)	15(-)	11(-)	12(-)	11(-)	11(-)	12(-)	20(-)	11(-)	12(-)	19(-)	14(-)
Aztreonam	90(+)	78(+)	50(+)	96(+)	46(/)	53(+)	51(+)	75(+)	31(/)	53(+)	84(+)	76(+)

Figure 2. A phylogenetic analysis of *X. citri* pv. *glycines* type strains based on 16S rRNA gene, DNA gyrase subunit B gene (*gyrB*) and RNA polymerase beta subunit gene (*rpoB*) sequences. The tree in (A) is based on the 16S rRNA gene, and the tree in (B) is based on a concatenation of *gyrB* and *rpoB*, with *X. oryzae* pv. *oryzae* used as an outgroup. The phylogenetic trees were generated by the neighbor-joining (NJ) method with 1000 bootstrap replicates and the bootstrap values are displayed on the trees. Bars (0.01) indicate the substitution rate per nucleotide position.

A.


X.oryzae pv. oryzae CIAT
X. vasicola pv. vasculorum Xv1601
72
X. arboricola CFBP6825
X. campestris pv. campestris ATCC33913
X. citri pv. glycines 8ra
100
K1
K13
K29
K47
K100
65.8
K10
K20
K46
K79
X. phaseoli pv. phaseoli CFBP6164
K34
X. euvesicatoria pv. alfalfae CFBP3836
K53
X. citri pv. citri MN12
3.0E-4

B.

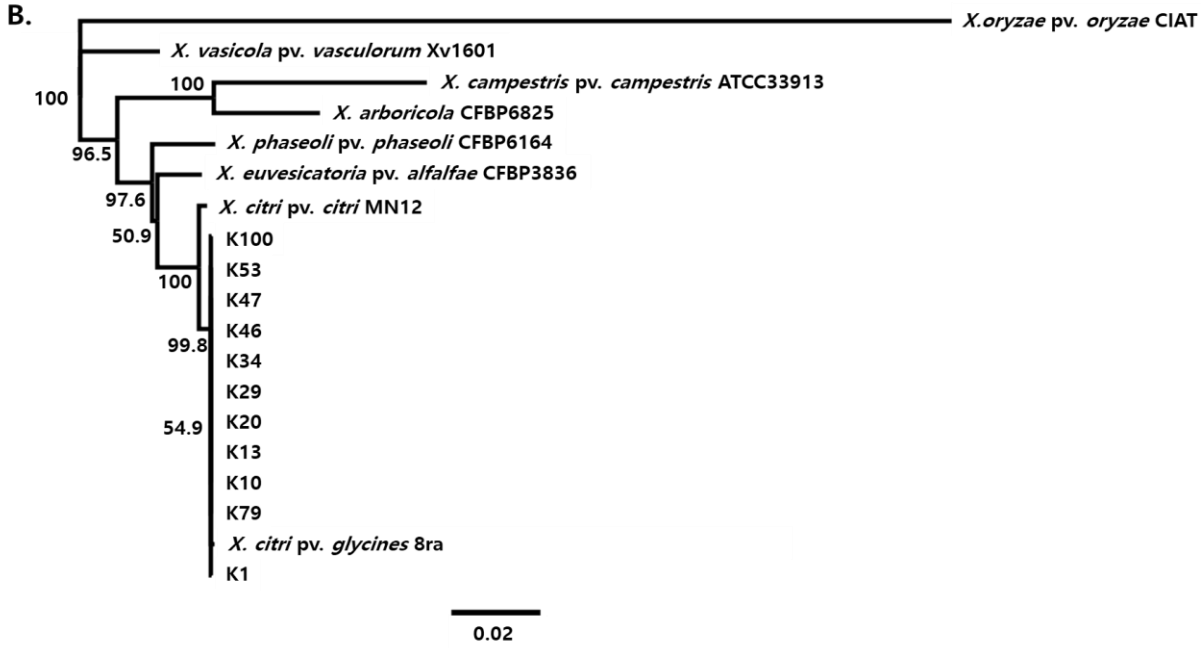

X.oryzae pv. oryzae CIAT
X. vasicola pv. vasculorum Xv1601
100
100
X. campestris pv. campestris ATCC33913
X. arboricola CFBP6825
96.5
X. phaseoli pv. phaseoli CFBP6164
X. euvesicatoria pv. alfalfae CFBP3836
97.6
X. citri pv. citri MN12
50.9
K100
100
K53
K47
99.8
K46
K34
K29
54.9
K20
K13
K10
K79
X. citri pv. glycines 8ra
K1
0.02

Figure 3. Southern blot analysis of *X. citri* pv. *glycines* strains isolated in Korea from 2012 to 2017. Genomic DNA was digested by *Bam*HI, run on a gel, hybridized with the pBSavrXa10 to probe for *avrBs3*, and detected using X-ray film. DNA ladder with sizes labeled on the left (in kb). Distinct banding patterns are labeled with numbers above the lanes, and those corresponding to the designated type strains are labeled with the strain names. Bands that are particularly dense reflect *avrBs3* alleles that putatively are on plasmids (Park et al., 2008).

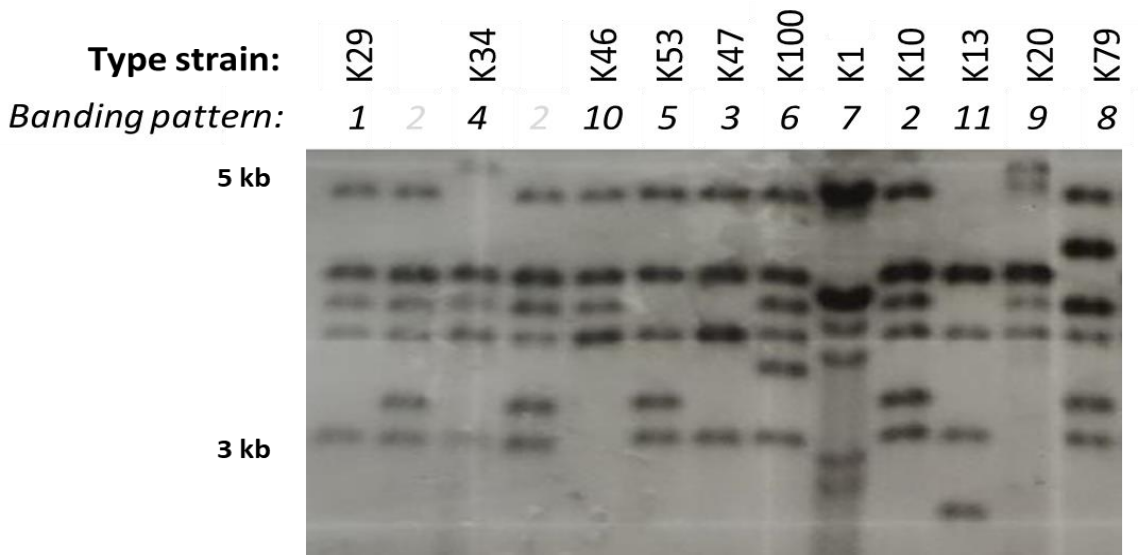

Type strain:
K29
K34
K46
K53
K47
K100
K1
K10
K13
K20
K79
Banding pattern:
1
2
4
2
10
5
3
6
7
2
11
9
8
5 kb
3 kb

Figure 4. Shift in the relative abundance of *Xcg* strain groups in Korea over a 13- to 15-year period. The strain groups were identified on the basis of their *avrBs3* banding pattern among 155 isolates collected in 1999 to 2002 (Park et al., 2008) and 106 isolates collected in 2012 to 2017 (this study). Two groups exhibited similar banding patterns in the two surveys: the groups labeled OcsF and SL1017, as labeled by Park et al., which corresponded to the groups labeled K29 and K100, respectively, in this study.

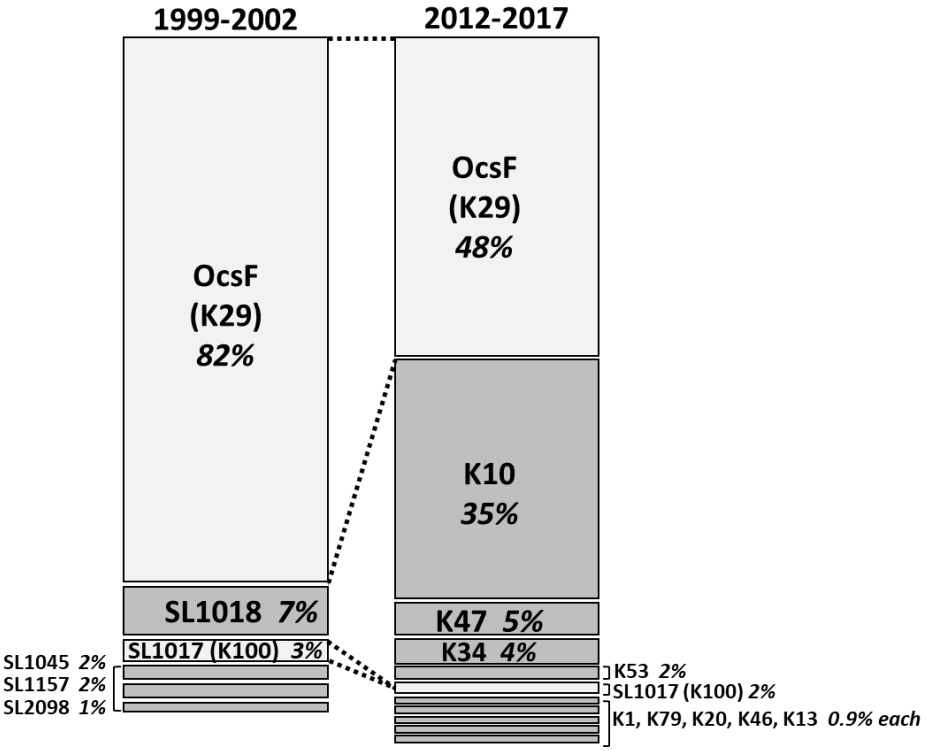
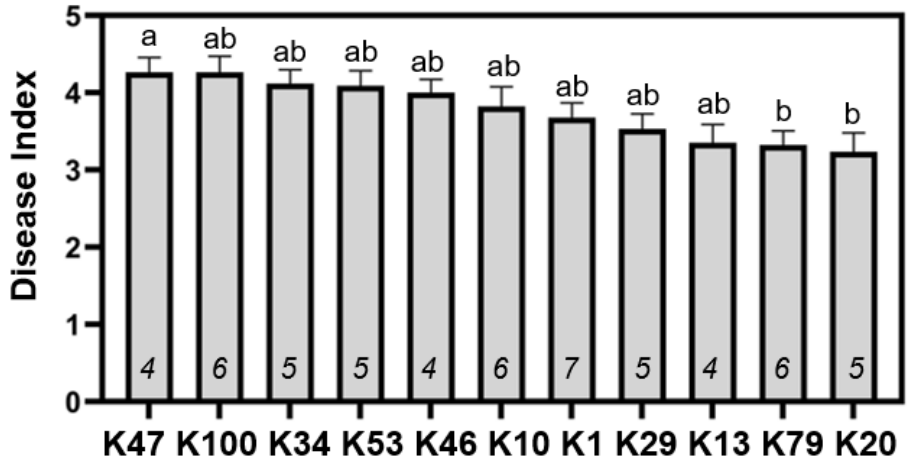

1999-2002
2012-2017
OcsF
(K29)
82%
OcsF
(K29)
48%
K10
35%
SL1018 7%
SL1017 (K100) 3%
K47 5%
K34 4%
SL1045 2%
SL1157 2%
SL2098 1%
K53 2%
SL1017 (K100) 2%
K1, K79, K20, K46, K13 0.9% each

Figure 5. Variation in the aggressiveness of the *Xcg* type strains in inducing bacterial pustule on soybean. The average disease index across 17 soybean cultivars tested (data were excluded for cultivars Williams and Williams 82) is shown for the 11 type strains. Italicized numbers within the bars are the *avrBs3* family copy number. Values shown are the mean + standard error of the mean, with significant differences based on analysis of variance and a Kruskal-Wallis multiple comparison test ($P < 0.1$).


Disease Index
5
4
3
2
1
0
a
ab
ab
ab
ab
ab
ab
ab
ab
b
b
4
6
5
5
4
6
7
5
4
6
5
K47 K100 K34 K53 K46 K10 K1 K29 K13 K79 K20

3. Responses of soybean cultivars to the *Xcg* type strains

Xcg strains have been divided into races based on their ability to induce disease in a panel of five differential cultivars (Williams, Harosoy, Pella, Chippewa, and Mukden) (Hwang et al., 1998). The Korean *Xcg* type strains isolated in 1999 to 2002 were screened on these five cultivars and 14 Korean soybean cultivars, and some resistance to these strains was identified in both sets of cultivars (Park et al., 2008). Here, the *Xcg* type strains isolated in 2012 to 2017 were screened on a similar set of cultivars to evaluate if shifts have occurred in Korean *Xcg* populations in the last decades. The response of 21 soybean cultivars to the 11 type strains resulted in a higher disease index across all cultivars and strains (mean of 3.8) than was observed in the study by Park et al. (2008) (mean of 2.3) (Table 5), suggesting that the strains were generally more virulent or the conditions were more conducive to disease during this recent screen.

The cultivars Williams 82 and Williams that carry the *rxp* resistance gene showed the most resistance to the most strains (Table 5). With the exception of type strain K79 on Williams, the disease indices on these cultivars were in the bottom quartile of the disease indices (i.e., ≤ 3.5) for all cultivar × type strain combinations screened. However, some strains were moderately virulent on Williams (K47, K79) and Williams 82 (K46, K47, K53 and K100) (Fig. 3). Although Hwang et al. (1998) identified a *Xcg* strain with virulence to Williams, this is the first report of *Xcg* strains showing virulence to Williams 82. Cultivar CNS also carries the *rxp* gene, but this cultivar exhibited less resistance than Williams or Williams 82 to all of the type strains except K79. Park et al. (2008) found that cultivars with the *rxp* gene were resistant to the type strain that represented the majority (83%) of the Korean *Xcg* strains in 1999 to 2002, suggesting that *rxp* could be useful if deployed for resistance in Korean soybean cultivars,

although the causal nature of *rxp* to resistance to these strains is not known. Interestingly, we found that CNS (PI 548445), which carries *rxp*, was susceptible to the two type strains that represented the majority (84%) of the Korean *Xcg* strains, K29 and K10, as well as to most of the other type strains, whereas Williams 82 and Williams, which also have *rxp*, were resistance to at least five Korean strains, indicating that two cultivars might have resistance loci in addition to *rxp*.

The *Xcg* strains in this study could not be classified into races based on their profile of virulence to the five cultivars used in previous studies (Chippewa, Harosoy, Mukden, Pella, and Williams) (Hwang et al., 1998). Although Williams showed some variation in resistance to the new Korean type strains, the cultivars Chippewa, Harosoy, Mukden, Pella, and Peking were highly susceptible to all of the new Korean type strains. Similarly, the *Xcg* type strains did not exhibit notable differences in aggressiveness on these cultivars, as the disease indices were almost all ≥ 4.0.

The Korean cultivars were, in general, also highly susceptible to the eleven *Xcg* type strains (Table 5), including the cultivar Daewon, which is widely grown in Korea. However, several cultivars exhibited a differential response to distinct type strains. The highest levels of resistance were Danbaek and Sinpaldal2 to strain K13, and Pungsannamul to strain K20. The next highest levels of resistance were observed in only four cultivars, Gwangan to strain K79, Pungsannamul to strains K1, K10, K29 and K79, Sinpaldal2 to strains K20, K47 and K79, and Somyung to strains K10 and K20. If these resistance responses reflect cultivar-specific recognition of avirulence genes in the type strains, then cultivars Pungsannamul and Sinpaldal2 and type strains K10, K20 and K79 are the best candidates for future studies exploring these host-specific interactions. In the absence of effector-triggered resistance, the variation in

disease indices among the cultivars (Fig. 3) may reflect variation in the aggressiveness of the strains on the cultivars tested. Consistent with the findings of Park et al. (2008), we did not find evidence that the copy number of the avrBs3 family genes influenced *Xcg* aggressiveness on these soybean cultivars (Fig. 3).

Table 5. Bacterial pustule disease severity induced by the type strains of *X. citri* pv. *glycines*[a]

	K1	**K10**	**K13**	**K20**	**K29**	**K34**	**K46**	**K47**	**K53**	**K79**	**K100**
Soybean cultivars used globally[b]											
Chippewa	4.0	5.0	4.0	3.5	4.5	5.0	4.5	5.0	5.0	4.0	5.0
CNS	4.0	4.0	3.5	3.0	3.5	4.5	4.0	4.5	4.0	3.5	4.5
Harosoy	4.0	4.0	4.0	3.0	4.5	4.0	4.0	4.5	4.0	3.5	5.0
Mukden	4.0	5.0	4.0	4.5	4.5	5.0	5.0	5.0	5.0	4.0	5.0
Peking	4.0	4.5	4.0	4.5	4.0	5.0	5.0	5.0	5.0	4.0	5.0
Pella	4.0	3.5	4.0	4.0	4.5	4.0	4.5	4.5	5.0	4.0	5.0
Williams	2.0	3.0	2.0	1.0	2.0	3.5	3.0	3.5	3.0	4.0	3.0
Williams 82	2.0	1.0	1.0	1.0	1.0	2.5	3.5	3.5	3.5	3.0	3.5
Soybean cultivars used in Korea											
Daepung2	4.0	3.0	3.5	2.5	3.5	4.0	4.5	5.0	3.5	4.0	3.5
Danbaek	2.5	3.0	1.0	2.5	2.5	3.5	3.0	4.0	2.5	3.0	2.5
Dawon	4.5	5.0	4.0	4.0	4.0	5.0	4.5	4.0	5.0	2.5	5.0
Gwangan	3.0	2.5	3.0	4.0	2.5	4.0	4.0	4.0	3.5	2.0	4.5
Hannam	4.0	4.0	4.0	4.0	4.0	4.5	4.5	5.0	4.5	4.0	4.5
Manri	5.0	5.0	4.0	4.0	4.0	5.0	4.0	5.0	5.0	3.5	4.5
Moohan	4.5	5.0	3.5	4.0	3.5	4.0	4.0	4.0	3.0	3.5	3.0
Pungsannamul	2.0	2.0	2.5	1.0	2.0	3.0	2.5	3.5	3.5	2.0	3.0
Pureun	3.0	4.0	3.5	2.5	2.5	3.5	3.5	4.0	4.0	3.5	4.0
Sinpaldal2	3.0	3.5	1.0	2.0	3.0	3.0	3.0	2.0	3.5	2.0	5.0
Somyung	3.0	2.0	3.5	2.0	3.0	3.0	3.5	3.5	3.5	3.5	3.5
Daewon	4.5	5.0	4.5	5.0	5.0	5.0	4.5	4.5	4.5	4.5	4.5
Taekwang	4.5	5.0	5.0	5.0	5.0	5.0	4.5	5.0	4.5	4.0	4.5

[a] Disease severity was evaluated on a 0 to 5 scale, with 0 indicating no pustules and 5 representing >100 pustules per leaflet.

[b] Cultivars Williams 82, Williams and CNS are known to carry the *rxp* resistance gene

4. Responses of nonhost plant species to the *Xcg* type strains

In a previous study, Kaewnum et al. (2005) demonstrated that 26 *Xcg* isolates from Thailand demonstrated not only variability in aggressiveness in their pathogenicity on soybean, but also variability in their ability to induce HR on nonhost plant species. Here, we evaluated the ability of the eleven *Xcg* type strains to induce the HR on diverse nonhost plant species. We did not observe variation among the *Xcg* type strains, as none of them induced a HR on two species of *Nicotiana* or on two cultivars of tobacco (*Nicotiana tabacum* L.). Similarly, none of them induced an HR on pepper (*Capsicum annuum* L.) or sesame (*Sesame indicum* L.), and they all induced an HR on tomato (*Solanum lycopersicum* M.).

In previous studies, a U.S. strain of *Xcg*, strain 8ra, similarly failed to induce HR on tobacco and induced it on tomato, but strain 8ra also induced HR on pepper, unlike all of the type strains in this study. Our findings support the ability of *Xcg* strains to induce a HR in a plant-specific manner, although the basis for this specificity is not known, but they do not provide further evidence for genetic variation among *Xcg* strains in their ability to induce HR on any of the nonhost species examined.

Table 6. Hypersensitive response (HR) induced by the type strains of *X. citri* pv. *glycines* (*Xcg*) on seven non-host species.

Plant species	**HR induction by the *Xcg* type strains**[a]
Nicotiana rustica L.	-
Nicotiana benthamiana L.	-
Nicotiana tabacum L. cv. Xanthi	-
Nicotiana tabacum L. cv. Samsun	-
Capsicum annuum L. cv. Supermanita	-
Sesamum indicum L. cv. Milsung	-
Solanum lycopersicum M. cv. Berry King F	+

DISCUSSION

In contrast to previous surveys of bacterial pustule in Korea that estimated that 86% of Korean soybean cultivation fields had bacterial pustule (Lee et al., 2010), we found only 7 to 8% of these fields were infected in 2015 and 2016. The recent survey varied from the previous ones in being nationwide and in surveying at least three times more fields. Despite this broadening of the survey, the factors responsible for the large decrease in bacterial pustule between 1997-1998 (Lee et al., 1999) and 2015 are not clear. The decrease could be related to a transition in the soybean cultivars that were planted. In the 1990s, the cultivars Taekwang and Daewon were the most widely grown cultivars in Korea, but since then, other varieties have displaced Taekwang. This decrease in disease could also be associated with shifts in the pathogen population toward reduced virulence or changes in climate. Given the favorability of high temperatures and heavy rainfall to bacterial pustule development, decreased disease would more likely be an outcome of water restriction and drought events in agricultural areas than warming temperatures (Lee and Shin 2021). Our survey data also indicated that the incidence of bacterial pustule in Korea over doubled between 2016 and 2017. This increase could be due to similar factors, in this case an increase in the precipitation in August of 2017, decreased use of resistant soybean cultivars and/or shifts toward increased virulence of the predominant *Xcg* strains. Our data indicate that significant population shifts occurred in the *Xcg* strains in soybean cultivation fields in Korea between 1999-2002 and 2012-2017. The dominant strains changed, and the diversity of the populations increased based on the greater number of distinct type strains and the greater evenness in their relative abundance. Although the *Xcg* strains

varied in their virulence on a set of 21 host cultivars, they generally exhibited strong virulence, and the Korean soybean cultivars exhibited little resistance to these strains. Collectively, these data do not support the possibility that shifts in the *Xcg* population contributed to the decreased incidence in bacterial pustule in Korea in the last decade. However, they predict that in years that are favorable to pustule development, the widespread distribution of the pathogen, the high virulence of the current endemic strains, and the low resistance of the Korean soybean cultivars are likely to create opportunities for widespread disease. Consequently, continued efforts to identify and integrate genetic resistance into commercial cultivars for soybean cultivation in Korea are important, but based on the susceptibility of CNS to all of the *Xcg* type strains representing the *Xcg* populations across Korea, this genetic resistance should target resistance beyond that mediated by *rxp*.

LITTERATURE CITED

Athinuwat, D., Prathuangwong, S., Cursino, L. and Burr, T. 2009. *Xanthomonas axonopodis* pv. *glycines* soybean cultivar virulence specificity Is determined by *avrBs3* homolog *avrXg1*. *Phytopathol*. 99: 996-1004.

Bernard, R. L. and Weiss, M. U. 1973. Qualitative genetics. In: *Soybeans: Improvement, production, and uses*, ed. by B. E. Caldwell, pp. 117-164. American Society of Agronomy, Madison, WI, USA.

Bradbury, J.F. 1984. Genus II. *Xanthomonas* Dowson 1939, 187. In: *Bergey's Manual of Systematic Bacteriology*, Vol. 1, ed. by N. R. Krieg and J. G. Holt, pp.199-210. Williams & Wilkins. Baltimore, MD, USA

Constantin, E.C., Cleenwerck, I., Maes, M., Baeyen, S., Van Malderghem, C., De Vos, P. and Cottyn, B. 2016. Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. *Plant Pathol.* 65: 792-806.

Hartman, G. L., Rupe, J. C., Sikora, E. J., Domier, L. L., Davis, J. A. and Steffey, K. L. 2016. *Compendium of soybean diseases and pests*. 5th ed. American Phytopathological Society, St. Paul, MN, USA.

Hartwig, E. and Lehman, S. G. 1951. Inheritance of resistance to bacterial pustule disease in soybeans. *Agron J*. 43: 226-229.

Hokawat, S. and Rudolph, K. 1991. Variation in pathogenicity and virulence of strains of *Xanthomonas campestris* pv. *glycines*, the incitant of bacterial pustule of soybean. *J. Phytopathol*. 131: 73-83.

Hong, S. J., Hong, Y. K., Lee, B. C., Lim, M. J., Yoon, Y. N., Hwang, J. B. and Park, S. T. 2007. Detection of *Xanthomonas axonopodis* pv. *glycines* and survey on seed contamination in soybean seeds using PCR assay. *Res. Plant Dis.* 13: 145-151. (In Korean)

Hong, S. J., Kim, Y. K., Jee, H. J., Lee, B. C., Yoon, Y. N. and Park, S. T. 2010. Selection of bactericides for controlling soybean bacterial pustule. *Res. Plant Dis.* 16: 266-273.

Hong, S. J., Kim, Y. K., Jee, H. J., Shim, C. K., Kim, M. J., Park, J. H. and Lee, B. C. 2011. Influence of disease severity of bacterial pustule caused by *Xanthomonas axonopodis* pv. *glycines* on soybean yield. *Res. Plant Dis.* 17: 317-325. (In Korean)

Hwang, I. and Lim, S. M. 1998. Pathogenic variability in isolates of *Xanthomonas campestris* pv. *glycines*. *Korean J. Plant Pathol*. 14: 19-22.

Hueck, C. J. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol. Biol. Rev*. 62: 379-433

Kaewnum, S., Prathuangwong, S. and Burr, T.J. 2005. Aggressiveness of *Xanthomonas axonopodis* pv. *glycines* isolates to soybean and hypersensitivity responses by other plants. *Plant Pathol*. 54: 409-415.

Kim, J. G., Choi, S., Oh, J., Moon, J. S. and Hwang, I. 2006. Comparative analysis of three indigenous plasmids from *Xanthomonas axonopodis* pv. *glycines*. *Plasmid*. 56:79-87.

Moffett, M.L. and Croft, B.J. 1983. *Xanthomonas.* In: *Plant bacterial diseases: A diagnostic guide,* eds. by P. C. Fahy and G. J. Persley, pp. 189-228. Academic Press, New York, NY, USA.

Narvel, J. M., Jakkula, L. R., Phillips, D. V., Wang, T., Lee, S. H. and Boerma, H. R. 2001. Molecular mapping of *Rxp* conditioning reaction to bacterial pustule in soybean. *J. Hered*. 92: 267-270.

Lee, S. D. 1999. Occurrence and characterization of major plant bacterial diseases in Korea. Ph.D. thesis. Seoul National University, Seoul, Korea.

Lee, J. and Shin, H. 2021. Assessment of future climate change impact on an agricultural reservoir in South Korea. *Water.* 13: 2125.

Park, H. J., Han, S. W., Oh, C., Lee, S., Ra, D., Lee, S. H. and Heu, S. 2008. Avirulence gene diversity of *Xanthomonas axonopodis* pv. *glycines* isolated in Korea. *J. Microbiol. Biotechnol*. 18: 1500-1509.

Sambrook, J., Fritsch, E. F. and Maniatis, T. 1989. *Molecular cloning: a laboratory manual,* 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA. 1546 pp.

Shukla, A. K. 1994. Pilot estimation studies of soybean (*Glycine max*) yield losses by various levels of bacterial pustule (*Xanthomonas campestris* pv. *glycines*) infection. *Int. J. Pest Manage*. 40: 249-251.

한국의 콩 불마름병균 *Xanthomonas citri* pv. *glycines*의 유전체 비교분석 및 레이스 동정

강인정

요약(국문초록)

Xanthomonas citri pv. *glycines*에 의한 콩 불마름병은 한국에서 가장 문제가 되고 있는 콩[Glycine max (L.) Merr.]의 세균병 중 하나이다. 먼저, 그동안 국내 불마름병균의 타입레이스로 알려진 5개 균주와 미국 균주인 8ra에 대해 전체 게놈 염기서열 및 유전자들의 유사성 분석을 토대로 병원성 관련 유전자를 분석하였다. 6개 균주에서 예측되는 type III secretion system effector인 transcription activator-like effectors는 다양한 반복 크기를 가지고 있었고, 특히 최소 6가지 형태로 repeat variable di-residue 서열이 나뉘어짐을 확인할 수 있었다. 또한 repeat variable di-residue 서열을 정렬한 후 그것을 기반으로 한 계통 발생을 확인해 본 결과, 17.5 repeat variable di-residue가 두 개의 별개 cluster로 구분됨을 확인할 수 있었다. 추가로, Transcription activator-like effectors의 effector

binding elements 역시 6개 그룹으로 분류되었고, 서열이 매우 보존적이어서 콩 품종에서 effector binding elements 특이성이 진화하고 있다고 추측되었다. 이러한 결과들은 불마름 균주들의 type III secretion system effector의 가변성과 적응성을 보여주고 있으며, 불마름 균주들의 병원성에 대한 이해를 높일 수 있을 것으로 기대한다.

두 번째로, *avrBs3* profile을 기반으로 2008년에 6개의 타입들의 대표 strain이 보고된 이후, 새로운 타입을 가진 불마름 균주가 한국에서 출현했는지 확인하고, 새로운 *avrBs3* 타입이 확인되면 이들이 한국 불마름 균주의 새로운 레이스를 대표할 수 있는지에 대한 병원성 등을 분석하였다. 전국에서 수집된 106개의 불마름 균주들의 다양성을 *avrBs3* banding pattern을 기반으로 확인한 결과, 최근에 분리된 균주들은 새롭게 11개 타입으로 나뉘며 이들 중 2개 타입만이 1999년에서 2002년 사이에 보고된 타입과 유사함을 확인하였다. 이는 불마름 균주들의 다양성이 증가하였음을 보여주는 결과이며, 특히 1999년과 2017년의 기간 동안 우세한 타입이 바뀌었으며, 2012년부터 2017년에 분리된 균주들에서 확인된 3개의 새로운 유형의 타입이 전체의 44%를 차지함을 알 수 있었다. 새롭게 분리된 타입들의 대표 균주들의 병원성을 검정한 결과, 현재 국내에서 널리 재배되는 대원 품종을 포함한 한국 콩 품종은 11개 새로운 유형의 대표 균주들에 감수성이었다. 또한 *rxp* 저항성 유전자를 보유하고 있는 품종 CNS는 한국 불마름 균주의 83%를 차지하는 2개의 대표 균주들을 포함하여 대부분의 새로운 대표 균주들에 감수성

이었다. 대조적으로, *rxp*를 보유하고 있다고 알려진 또다른 품종 Williams 82는 적어도 5개의 대표 균주들에 대해 저항성 반응을 보였다. 종합적으로, 이러한 결과는 Williams 82가 *rxp* 외에도 다른 저항성 유전자를 가지고 있다는 가능성을 시사한다. 종합적으로, 새로운 유형의 불마름 균주들의 광범위한 분포, 현재 분포하고 있는 불마름 균주 레이스들의 높은 병원력, 새롭게 확인된 레이스들에 대한 대부분의 한국 콩 품종의 감수성 반응 등이 보여주는 것은 콩 불마름병 발생이 유리한 조건에서 불마름병원균이 널리 확산될 수 있는 위험이 있음을 시사한다.

본 연구는 한국 불마름 대표 균주들에 대해 전체 게놈 수준에서 transcription activator-like effectors 및 그 effector binding elements를 특성화하고, 최근 한국에서 분리된 불마름 균주들에서 새로운 레이스를 확인하여 국내에서 재배되고 있는 주요 콩 품종에 대한 병원성을 확인한 첫번째 연구이다. 본 연구를 통해 한국에서 콩 불마름병에 내성을 가진 저항성 품종 개발에 대한 전략을 세울 수 있을 것으로 기대한다.

주요어: *Xanthomonas citri* pv. *glycines*, 콩 불마름병, 유전체 비교분석, 신규 레이스, 병원성

학 번: 2016-38299