



HAL
open science

VNTR Typing of the Bacterial Rice Pathogen *Burkholderia glumae* Reveals the Coexistence of Several Diverging Lineages in a Single Field in Colombia

Gilles Béna, Paola A Fory, Johanna Echeverri Rico, Gloria M Mosquera

► **To cite this version:**

Gilles Béna, Paola A Fory, Johanna Echeverri Rico, Gloria M Mosquera. VNTR Typing of the Bacterial Rice Pathogen *Burkholderia glumae* Reveals the Coexistence of Several Diverging Lineages in a Single Field in Colombia. *Plant Disease*, 2021, 105 (11), pp.3466 - 3473. 10.1094/pdis-02-21-0235-re. hal-03720778

HAL Id: hal-03720778

<https://hal.science/hal-03720778>

Submitted on 12 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL
open science

VNTR Typing of the Bacterial Rice Pathogen *Burkholderia glumae* Reveals the Coexistence of Several Diverging Lineages in a Single Field in Colombia

Gilles Béna, Paola A Fory, Johanna Echeverri Rico, Gloria M Mosquera

► **To cite this version:**

Gilles Béna, Paola A Fory, Johanna Echeverri Rico, Gloria M Mosquera. VNTR Typing of the Bacterial Rice Pathogen *Burkholderia glumae* Reveals the Coexistence of Several Diverging Lineages in a Single Field in Colombia. *Plant Disease*, American Phytopathological Society, 2021, 105, pp.3466 - 3473. 10.1094/pdis-02-21-0235-re . hal-03720778

HAL Id: hal-03720778

<https://hal.archives-ouvertes.fr/hal-03720778>

Submitted on 12 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

VNTR Typing of the Bacterial Rice Pathogen *Burkholderia glumae* Reveals the Coexistence of Several Diverging Lineages in a Single Field in Colombia

Gilles Béna,^{1,†} Paola A. Fory,² Johanna Echeverri Rico,³ and Gloria M. Mosquera²

¹ Plant Health Institute Montpellier (PHIM), Université de Montpellier, French National Research Institute for Sustainable Development (IRD), National Research Institute for Agriculture, Food and Environment (INRAE), French Agricultural Research Centre for International Development (CIRAD), Institut Agro, Montpellier, France

² Agrobiodiversity Program, International Center for Tropical Agriculture (CIAT), Cali 763537, Palmira, Colombia

³ The National Federation of Rice Growers of Colombia (FEDEARROZ), 110911 Bogota, Colombia

Abstract

Burkholderia glumae is responsible for the panicle blight disease of rice. This disease is present worldwide and can result in significant drop in yields. To estimate the genetic diversity of the bacterial strains present in a rice paddy field in Colombia, we sampled 109 strains from infected panicles. To detect fine genetic relationships among related haplotypes, and to overcome a very low nucleotide diversity detected in previous studies, we designed primers to amplify and sequence several highly variable minisatellite loci, or variable number tandem repeats (VNTRs), as well as part of the Toxoflavin *toxA* gene in all strains. Results show that the *toxA* nucleotide diversity defined four lineages and was similar to that detected in several fields in Japan; data suggest that *B. glumae* has

spread from Asia to America without major loss of genetic diversity, and that five VNTR loci discriminated the strains within the field revealing single and multi-infections of the rice panicles with a wide distribution of the haplotypes among the different plots. Even though disease levels vary considerably from year to year, the bacterial genetic diversity is maintained within a field. We do not detect any geographical structuring within the field, nor any effect of the rice cultivar on the observed diversity. The consequences on the origin and evolution of the bacteria are discussed.

Keywords: *Burkholderia glumae*, Colombia, population structure, rice, VNTR

Rice is the staple food for >3.5 billion people worldwide, around half of the world's population. Its cultivation is under threat from many diseases that decrease yields significantly and consequently have dramatic effects on the lives of human beings. Among these diseases, some have been studied for many years, such as rice blast (caused by *Magnaporthe oryzae*; Gladioux et al. 2018) or bacterial blight (caused by *Xanthomonas oryzae* pv. *oryzae*; Jiang et al. 2020). Although less studied, panicle bacterial blight caused by *Burkholderia glumae* appears to be one of the most worrying emerging diseases. *B. glumae* was first described in Japan at the end of the 1950s as causing symptoms of rotting grains and leaf sheaths in rice (Goto and Ohata 1956). This bacterial disease has now been described in many countries in Southeast Asia, and in Latin and North America (Mondal et al. 2015; Nandakumar et al. 2009; Riera-Ruiz et al. 2014; Zhu et al. 2008). The reduction in yield from this disease can reach up to 75% for the most affected plots (Trung et al. 1993), resulting from reduction of the weight of grains or from the sterility of the inflorescences. The determinants of re-emergence of the disease remain unclear. Prolonged periods of warm and wet conditions at the rice flowering time appear to be of primary importance (Lee et al. 2015). The correlation between high prevalence levels and high night-time temperatures, such as those reported in 1995 and 1998 in Louisiana, supports this hypothesis (Nandakumar et al. 2009). In today's context of global warming, such high temperature conditions during flowering are likely to become more frequent in the coming years. This has led authors to suggest that

B. glumae could become in the future the “next major pathogen of rice” (Ham et al. 2011).

Data on the epidemiology of this disease remain very fragmentary. The bacterium is considered by some authors to be seed-borne, wherein it could be located in the epidermis or the parenchyma (Agarwal and Sinclair 1980). However, while *B. glumae* has been detected in seeds (Cottyn et al. 2009; Saylor et al. 2006), seed transmission and the significance of seed-borne initial inoculum for the development of bacterial blight has not been formerly assessed. The bacterium might also be present in the leaf sheaths, forming the source of inoculum for development in the panicles, from which the spread of the disease would be in close proximity to other plants, therefore producing an aggregated profile. The spatial distribution within a field would be a function of the most affected panicles, from which the spread of the disease would be by proximity (Trung et al. 1993). The available data do not go much further, with only microscopy results that monitored tagged bacteria in planta (Li et al. 2016). Moreover, detection of the bacterium in the sheath, leaf blade, boot, or spikelet remains challenging, with drastic variations of population size from sheath to leaf (Tsushima 2011), leading to apparent “disappearance” and “reemergence” between consecutive stages of development (Echeverri et al. 2021). Finally, this bacterium has been shown to interact with other pathogens such as fungi, which may also play a role in the epidemiology of the bacteria (Jung et al. 2018).

The understanding of the origin and emergence of this plant bacteriosis also remains very incomplete. The disease and its causative agent were described and isolated for the first time in Asia (Kurita and Tabei 1967). However, as of this writing, there is no evidence that Asia is actually the center of origin for the disease. Furthermore, at local scale, data on the fine genetic diversity of the bacteria is almost non-existent. One study has looked at the genetic diversity in one country. Maeda et al. (2006) sampled and genetically typed 69 strains of *B. glumae* isolated from rice at 25 different sites in Japan. Characterization based on the sequencing of three loci was, however, not very informative on the genetic structure of the diversity of this species at this geographic scale because of the very low level of nucleotide diversity detected. With 25 isolates sampled in Arkansas in 2003, Saylor et al. (2006) detected 22 unique but closely related DNA fingerprint profiles. Cluster analysis revealed two major groups and

[†]Corresponding author: G. Béna; gilles.bena@ird.fr

Funding: This work was supported by the Consultative Group on International Agricultural Research (CGIAR)'s Program on Rice Agri-food Systems.

*The e-Xtra logo stands for “electronic extra” and indicates there are three supplementary tables published online.

The author(s) declare no conflict of interest.

Accepted for publication 31 March 2021.

their genetic diversity was relatively low as revealed by 16S-23S rDNA internal transcribed spacer sequence analysis. Finally, little genomic data are available. Indeed, at the time this project was initiated, only two full genomes were available from GenBank (including the type strain LMG2196^T; <https://www.ncbi.nlm.nih.gov/genbank/>; Lim et al. 2009), five others being available as draft genomes. However, these strains do not correspond to any logical sampling (in an epidemiological or evolutionary sense), but rather are specific strains from different laboratories. Two studies have compared genomes (Fory et al. 2014; Francis et al. 2013) and suggested a considerable amount of plasticity, largely based on accessory genes and genome islands, including multiple insertion/deletion (indel) events.

The goal in this study was to get the first image of the genetic diversity of *B. glumae* isolates sampled from a single rice field in Colombia, a country where the disease has been present for several years with high incidence rates but with a high variation between years. Considering the lack of variable genotyping markers available for this species, we developed several *B. glumae* variable-number-of-tandem-repeat (VNTR; Bakhtiari et al. 2018) markers to obtain a first snapshot of within field diversity, and thus circumvent the limits of the multi-locus sequence analysis approach. We detected a large genetic diversity, with several independent genetic lineages, raising the possibility of either several introductions of the bacterium into Colombia or considering this country as an alternative diversification zone.

Materials and Methods

Sampling of isolates. We sampled panicles from two rice varieties growing in a single field at the Fedearroz station in Saldaña, Tolima, Colombia (3°54'50.80"N; 74°59'6.76"W). In late summer in 2017, a trial included eight different varieties that were sown for agronomic and resistance to pathogens. Each variety was grown in four replicates, each in a 50 m² plot (Fig. 1). For each of the two varieties known to be either susceptible (F2000) or moderately resistant (FNAL004-1) to panicle blight, we selected in each plot three panicles from three different plants at maturing stage with, whenever possible, two plants with clear symptoms of panicle blight and one plant with no or low level of symptoms. Each panicle was stored in an individual paper bag.

Back at the lab, we isolated strains from seeds. Briefly, from five to 10 seeds were put in a 2-ml Eppendorf tube (Hamburg, Germany) and ground with a small sterile pestle. One ml of sterile water was added, and tubes were left at room temperature for 2 h. Serial dilutions of maceration, from 1/10 to 1/1,000, were plated (100 µl) on King's B agar (King et al. 1954) and incubated at 28°C during 24 h. Each plate was checked under a stereoscope and several colonies were picked up according to their morphology that were similar to *B. glumae* (pictures available upon request to G. Béna). All colonies collected were first streaked on King's B plates to verify their purity, and, starting from a single isolated colony, all were grown in liquid Yeast Mannitol Broth. DNA for each sample was extracted with a classical proteinase K protocol. All strains were stored at -80°C in 20% glycerol.

Molecular characterization of isolates. Toxoflavin, the main phytotoxin produced by both *B. glumae* and *B. gladioli*, is one of the main enzymes involved in the pathogenicity of *B. glumae* (Jeong et al. 2003) and, as of this writing, has never been detected in other species. In a first step, we tested for a positive amplification for all DNA samples targeting the methyltransferase *toxA* gene, which is part of an operon of five genes that codes for the synthesis of the phytotoxin toxoflavin, with primers designed in this study (Table 1). These primers were used both for confirmation of the taxonomic status of the isolate (positive or negative result for PCR amplification) and detection of genetic diversity among our isolates (by sequencing). All *toxA* and subsequent VNTR PCR products were sent to GenoScreen (Lille, France) for purification and forward strand sequencing.

In a second step, we typed all strains of *B. glumae* determined according to their *toxA* sequences. As previous works had shown the low genetic diversity detected among genomes available in GenBank or among isolates sampled all around Japan (Maeda et al. 2006), we used highly variable VNTR loci to characterize each strain. We used the online bioinformatics tool Polloc-V (<http://bioinfo-web.mpl.ird.fr/xantho/utills>) to screen for VNTR loci in five *B. glumae* genomes from the NCBI data bank (LMG2196^T, BGR1, AU3208, 3252-8, and

336gr-1). Polloc-V identifies groups of loci based on the similarity rate of their flanking regions. In a first analysis, we used the following selection criteria:

full loci size from 30 to 1,000 bp;
 pattern size from 5 to 9 bp;
 number of repetitions set to six or above;
 stringency set to maximal values 2 (match) -3 (mismatch) -5 (indel);
 and
 minimum percentage of similarity between repeat units = 80%.

In the first row of analysis, only a few primers could be selected, either because they could not be detected in one or several genomes (from indels or draft status) or because of their lack of diversity. We then

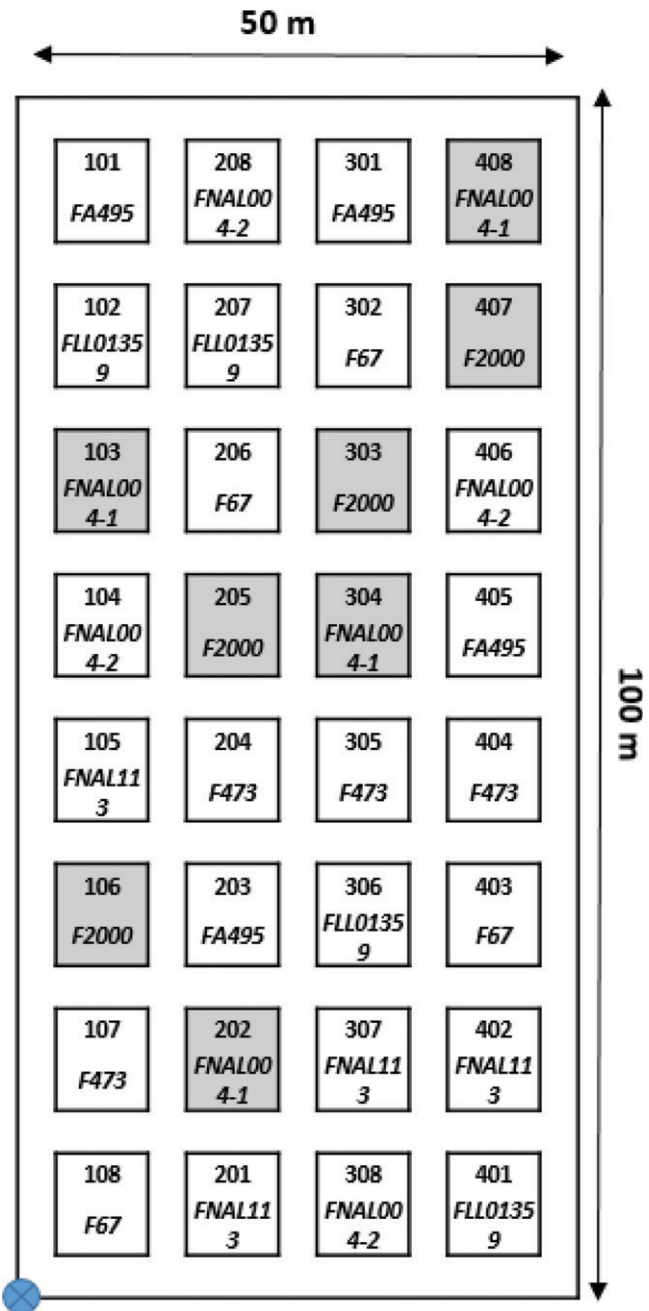


Fig. 1. Map of the trial established during September 2017 in the Experimental Center las Lagunas - Fedearroz in Saldaña, Colombia. Each plot (50 m² each) is numbered from 1 to 8, with four replicates (10 to 40). The whole field was flooded. Rice genotypes shown in each plot are in italics. Plots in gray shading were sampled. The blue cross in a circle is where the global positioning system track was taken (3°54'50.80"N; 74°59'6.76"W).

perform a second analysis with the same parameters but with the pattern size extended from 10 to 90 bp (i.e., corresponding to minisatellites rather than microsatellites). Similarly, we only kept sequences that were recovered from the five genomes and for which diversity could be detected. For all loci detected that fulfilled these constraints, we designed a set of primers in the flanking regions of the tandem repeat sequence using the web interface tool, Primer3Plus (<https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>; Untergasser et al. 2007). All pairs of primers designed were tested in silico using the suite of command-line tools NCBI-BLAST+ (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download; Camacho et al. 2009) against *B. glumae* genomes to confirm their specificity for the genomic region containing the locus. Whenever possible, melting temperatures were set around 60°C (Table 1). Moreover, because the genome of *B. glumae* consists of two chromosomes and four plasmids (Lim et al. 2009), we tried as much as possible to have loci on each of the two chromosomes.

All primer pairs designed have been tested for positive amplification with a subsample of 24 randomly selected isolates from our derived sample. To obtain the highest level of diversity, including both the number of repeats and single nucleotide polymorphisms (SNPs), we sequenced these loci for all individuals. We retained for further analysis only those primers for which we obtained amplification for at least 95% of the individuals and for which sequencing revealed variability, both in terms of number of repeats and in terms of SNPs, both within the repeats and in the flanking regions of the locus. All sequences obtained were verified and corrected if necessary using the software ChromasPro v.1.34 (Technelysium Pty Ltd, South Brisbane, Queensland, Australia). The different haplotypes of each locus were detected using the tool Mothur (Schloss et al. 2020). Each different haplotype sequence has been deposited in GenBank. We finally included in the final dataset other *B. glumae* genomes (full or draft) that were released after the typing of all our isolates.

Data analyses. *toxA* sequence diversity. As a first step, using the tool BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch), we recovered all the *B. glumae toxA* sequences of at least 473 bp available in GenBank and we reconstructed a Maximum Likelihood phylogenetic tree using the program MEGA6 (<https://www.megasoftware.net/>; Tamura et al. 2013). The best model of molecular evolution was chosen according to the “Akaike’s information criterion” function implemented in MEGA6. The *toxA* sequences of the sister species *B. gladioli* were only included when we searched for traces of selection along the branch joining the two species (see next paragraph). None of the data extracted from GenBank suggests *toxA* transfer between the two species, which confirms the species determination based on the *toxA* sequence (data not shown). As a second step, to simplify the phylogeny obtained, we used the software package Mothur (<https://mothur.org>). We kept only one or two representatives of each haplotype detected, retaining isolates from different countries when their sequences were identical to reflect the geographical distribution of diversity.

To explore a possible selection that would have shaped and constrained the nucleotide diversity of *toxA* sequences, we estimated the number of synonymous and non-synonymous mutations in the

dataset using the tool DnaSP 6 (<http://www.ub.edu/dnasp/>; Librado and Rozas 2009). We tested for selection along the gene by estimating the dN/dS between all pairs of sequences (i.e., the ratio of the number of non-synonymous substitutions per non-synonymous site to the number of synonymous substitutions per synonymous site). We also tested for a significant departure from a neutral evolution of this ratio along the branch joining the *B. glumae* and the *B. gladioli* clades using the online tool datamonkey (<https://www.datamonkey.org/>; Weaver et al. 2018) with the “aBSREL” test function (<https://www.datamonkey.org/absrel>; Smith et al. 2015).

VNTR diversity and network reconstruction. The different alleles of each VNTR locus were detected on the basis of the total length (and thus the number of repetitions) but also considered point mutations (same length but different sequences). The discriminatory power of each locus was compared with the Hunter–Gaston discriminatory index. This index is based on the probability that two unrelated strains randomly sampled in our population will be placed into different typing groups.

To visualize the different clonal complexes and genetic relationships among haplotypes, we applied a two-step procedure. We considered separately the nucleotide diversity detected in the *toxA* sequences and in the flanking region of each VNTR locus and the variations in the number of repeats for each VNTR locus. To begin, we retrieved all the nucleotide diversity from *toxA* and flanking regions, for which the rate of point mutations must be considerably lower than the rate of mutations within the repeat region. We concatenated these sequences for each isolate. With this data matrix, a neighbor-net split graph was built, using the software SplitsTree4 (<http://www.softsea.com/review/SplitsTree4.html>; Huson and Bryant 2006). This graph allowed us to distinguish between the different groups of sequence type (ST) but also potentially illustrate recombination effects. Then, for the different groups of haplotypes obtained on the graph, we applied a network reconstruction using the program PHYLOViZ v.2 (<https://www.phyloviz.net/>; Nascimento et al. 2017), with the algorithm of global optimal (“goeBURST” function) distances. It was not a straightforward task, as one highly polymorphic VNTR locus (BG02_2) revealed, after sequencing either differences in length involving several indel events or alleles with a different nucleotide sequence (involving punctual mutations within repeats) but the same total number of repeats (i.e., homoplasmy). We thus reconstructed the network based on the haplotypes coding schemes rather than on the VNTR length of each loci, and considered in the network all single and double locus variants.

Structure of the genetic diversity in the field. We tested whether there was a significant effect of the variety or plot on the bacterial genetic diversity recovered from each plant. Initially, we simply mapped on the haplotype network the origin of the isolates, whether it was the variety or the plot. We then tested whether the distribution of the different haplotypes between varieties and between plots deviated from a random distribution, using the program Genepop (<https://cran.r-project.org>; Rousset 2008), with the “population differentiation” option and an unbiased estimate of the *P* value performed on a contingency table. Subsequently, to disentangle plot and variety effects, we tested for a significant differentiation “between variety” and “among plot within each variety” in an analysis of molecular variance function

Table 1. List of the primers used, all of which were designed in this study

Locus	Primer’s name	Primer’s sequences (5’ to 3’)	Tm ^a /elongation time	Location (chromosome/gene) ^b
<i>ToxA</i>	ToxAF	TTTCGGGCGTGAAATCTATC	56°C	Chr2; toxoflavin methyltransferase
	ToxAR	GTCTGCAGGCAGTTGTCTCTG	30 s	
BG09	BG09F	AAACGGTTCCGATTTTCCTC	57°C	Chr1; hypothetical protein
	BG09R	TGGTGGTCGACAACGACAG	1 min	
BGm03	BGm03F	ATCATGGACCGTCTCTCC	57°C	Chr2; intergenic sequence
	BGm03R	TCCGCGTTCTAACTTCTACGA	1 min	
BG12	BG12F	CGGCCGATGTCCTTAAAC	57°C	Chr2; intergenic sequence
	BG12R	TAGTATTGACCGCGTGAAA	1 min	
BG02	BG02F	GTCGGYATATCGTCGATGC	57°C	Chr2; lipoprotein, bacterial collagen
	BG02R	GTGAAAACCGTGCCGAGT	1 min 30 s	

^a Tm, melting temperature.

^b Location on chromosome and gene are given according to LMG2196^T genome sequence and annotation (GCA_000960995.1); all PCR cycling parameters start with an initial denaturation step (95°C for 5 min) followed by 30 three-step cycles (94°C/30 s, Tm/30°, 72°C/elongation time).

performed with the program Arlequin v.3.0 (<http://cmpg.unibe.ch/software/arlequin3/>; Excoffier et al. 2005). The linkage disequilibrium among all pairs of loci was tested with the “log likelihood ratio” statistic implemented in the program Genepop.

Results

Identification of *B. glumae* isolates and *toxA* diversity. We sampled 24 different panicles from eight plots and two varieties. Three panicles showed no symptoms (plots 103, 202, and 407), six showed mild symptoms, and the last 15 showed severe symptoms. We isolated strains based on their morphology from each of the panicles. The number of strains isolated per panicle varied from four to 14, and the number per plot varied from 24 to 31, for a total of 207 strains that were stored in glycerol at -80°C . The *toxA* amplification was positive for 119 samples. We sequenced all 119 of these PCR products, and sequences obtained clearly clustered 10 isolates within a *B. gladioli* clade with all other sequences of this species obtained from GenBank. Although known to be pathogenic on rice, we did not include these strains in the study because they represented $<9\%$ of the entire data sets and were not associated to a variety nor a specific plot. We continued the analyses with the remaining 109 isolates (Supplementary Table S1). These strains were distributed over the eight plots, with a variation from seven (plot 304) to 22 strains (106) per plot. For two panicles (one symptomless panicle from plot 103 and one with mild symptoms from plot 304) no strains for which *toxA* could be amplified were obtained. Conversely, we obtained six strains from a panicle in plot 202 that showed no symptoms (Supplementary Table S1).

We retrieved 88 *toxA* sequences from GenBank, originating from Japan, China, Korea, Vietnam, the U.S.A., Puerto Rico, and Colombia, corresponding to the main areas where the disease has been reported so far. We aligned these sequences with ours, which is 473 bp over the 738 bp of the entire *toxA* gene. Along the 473 bp we only detected five SNPs, two of which were specific to our dataset, giving seven different haplotypes, four of which were detected in our study. The alignment of 23 complete *toxA* sequences, extracted from GenBank with our four haplotypes, revealed only eight different mutations, four being singletons (i.e., limited to a single sequence). The phylogenetic reconstruction resulted in seven different clades, three including strains isolated from different continents (Fig. 2). Of the eight mutations detected, three resulted in amino acid substitution, while the five SNPs detected in our data are all silent. We did not detect any evidence of a significant departure from a neutral selection using the aBSREL test (<https://www.datamonkey.org/absrel>; Smith et al. 2015) implemented with the online tool datamonkey, either when testing the overall *B. glumae* phylogeny, or the branch joining the two *B. glumae* and *B. gladioli* species, or the branch joining the two *Burkholderia* species with two homologous sequences from *Pseudomonas* species (data not shown).

VNTR loci selection. We designed 14 pairs of primers that framed different VNTR loci. The quality and efficiency of these primers were first tested on DNA of the type strain LMG2196^T. Among the 14 pairs of primers, one pair did not produce any amplicon and was discarded. The 13 others each resulted in a single band of the expected size. We then tested these primers on a subsample of 24 *B. glumae* isolates randomly chosen from our study. The rate of success ranged from 42 to 100% for the primers, with only three primer pairs giving 100% amplifications. We retained these last three loci and a fourth primer pair, which failed to amplify two samples. Locus BG09 is located on chromosome 1 whereas the three others (BG02, BGm03, and BG12) are on chromosome 2.

We amplified and sequenced the four loci for all our isolates. Characteristics of each VNTR locus are given in Table 2. The length of each repeat (8 to 12 bp) classify them at the frontier between minisatellites and microsatellites. The BG02 locus is composed of two VNTRs in a row, the first with two to five repetitions of a 12-bp pattern, the second just afterward with 27 to 51 repetitions of a 9-bp pattern. In all subsequent analyses, we treated them separately, respectively as BG02_1 and BG02_2. Loci BG12 and BGm03 are nearly perfect, with only one SNP detected in each; BG02_1 is perfect, whereas BG09 and BG02_2 are compound VNTR loci, with, respectively, four and seven different repeats. However, in each case, the different repeats only

differ from each other by a single nucleotide difference. The 109 typed strains produced 19 multi-locus haplotypes (Supplementary Table S2). The number of alleles, when considering both number of repeats and SNPs, ranged from six to 16 alleles (Table 2). When only considering number of repeats, the number of alleles ranged from four to 14. The HDGI scores ranged from 0.645 to 0.924. The combination of the five loci (*toxA* + the four VNTRs) has a discriminatory power of 0.938.

Within plot and within panicles diversity. Each sampled plot contains from two (plots 103 and 303) to six different haplotypes (plot 106). There is no correlation between genetic distance among haplotypes and their co-occurrence within a plot (Fig. 3). We can also note the low intrapanicle diversity observed. Indeed, of the 21 panicles for which we isolated several strains of *B. glumae*, only six contained several haplotypes (Supplementary Table S2). Among these six panicles, only one case (plot 106 panicle 2) suggests intrapanicular diversification (i.e., haplotype ST11). ST11, composed of only one isolate, is closely related to ST10 ($n = 7$ isolates) as a single locus variant, and further differs by a unique repeat. It probably emerged as a result of a mutation of ST10. The other panicles contain very divergent haplotypes differing by at least three VNTR loci, belonging to different genetic groups (i.e., groups including single- and double-locus variants) according to the network.

The four loci display different mutation step models.

Sequencing suggested that each loci did not evolve by following the same mutation step model. Within the small clonal complexes, each allele differs from another only by a single repeat for VNTR loci BG12, BG09, BGm03, and BG02_01. Conversely, for BG02_02 we detected differences between alleles along the branches of clonal complexes ranging from one to 12 repeats. ST3 and ST12 each differ (by three repeats) at locus BG02_2 (38 and 41, respectively), and the alignment of the sequences shows that this is the result of a single three-repeat indel event. Conversely, ST1 and ST2 also differ by three repeats (34 and 31, respectively), but the alignment shows three different mutation events, involving three different indel events (Fig. 3). We mapped onto the network the different number of mutation events between haplotypes for BG02_2.

Network reconstruction and linkage disequilibrium. We reconstructed network haplotypes in a two-step analysis using the programs SplitsTree4 and Phylovis v.2 (Fig. 3). The concatenated data of VNTR flanking regions and *toxA* sequences resulted in a 1,629-bp matrix with 35 polymorphic sites, including three singletons. In a first step, using the Neighbor-Net method implemented in SplitsTree4, we reconstructed a split network with four clearly separated groups of isolates, showing little evidence of recombination among them. In a second step, for each group, we reconstructed the network considering different alleles of each haplotype rather than the length of each VNTR locus to avoid homoplastic pitfalls (Supplementary Table S3). The tree cut-off, which is the threshold of number of differences (i.e., different alleles at each locus) below which haplotypes are considered to belong to the same clonal complex, was set to three. When considering only VNTR loci that have 19 different haplotypes and 171 potential links between them, 21 display <3 different alleles. The average number of different alleles for the five VNTR loci between haplotypes is 3.78, reflecting the genetic divergence among haplotypes. The network resulted in five groups and two singletons. Three groups each clustered together two haplotypes that differ between them at a single locus. The two other groups are more divergent, with up to two different alleles among haplotypes.

Considering all isolates, we detected a highly significant linkage disequilibrium among all loci ($P < 10^{-5}$ for each locus pair). The analysis of molecular variance showed a lack of variety effect when the plot effect was removed in the model ($P = 0.26$), reflecting the lack of specificity or preference of the various bacterial haplotypes by one of the two varieties.

Discussion

B. glumae is a bacterium that induces the rice panicle blight disease present in Asia and America and can occasionally cause extremely high yield losses. Our study explores for the first time in detail the diversity and genetic structure of bacterial populations in a cultivated rice field. The first step was to isolate bacteria from rice panicles. As

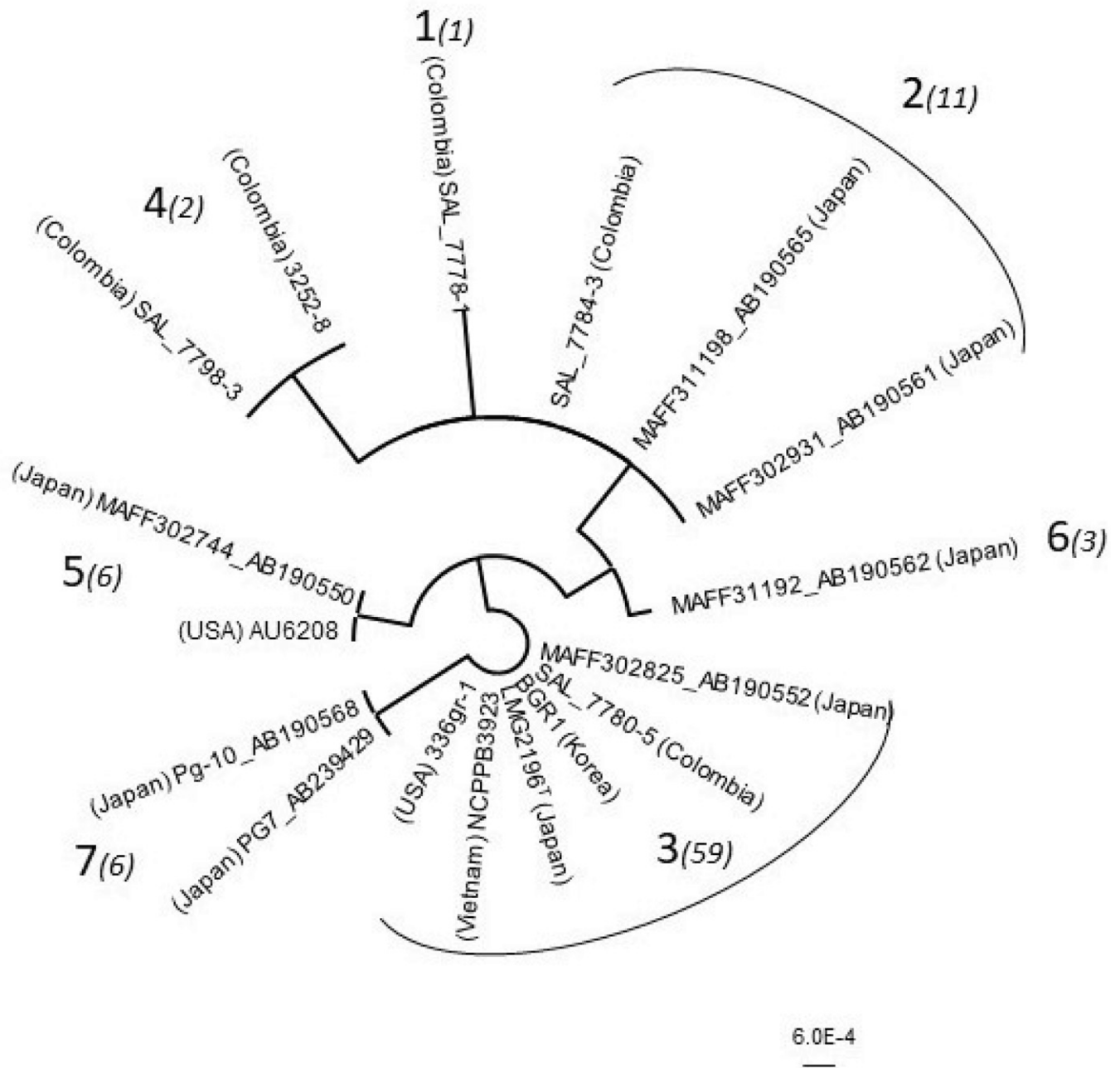


Fig. 2. Maximum likelihood phylogeny of the toxoflavin *toxA* sequence. Tree is based on partial 473-bp sequences. The seven types detected are numbered from 1 to 7, while numbers in parentheses and italics indicate the number of sequences retrieved from GenBank that fall in each clade (thus excluding sequences from our study). Types 2, 3, and 5 clustered sequences are from different countries.

Table 2. Genetic diversity of the five loci retained in the typing scheme of *Burkholderia glumae*^a

Locus	Number of alleles	Number of alleles based only on repeat numbers	Number of SNPs ^b in flanking regions ^c	Allelic range variation (size of the repeat)	Type of VNTR	PCR product size range (bp)	HGDI ^b scores
<i>toxA</i>	4	–	0	–	–	560	0.698
BG09	9	8	13	5–19 (9 bp)	C ^b (4 types)	522–648	0.872
BGm03	6	4	6	2–5 (10 bp)	P ^b (1)	404–445	0.745
BG12	7	5	7	2–11 (8 bp)	P (1)	460–532	0.795
BG02	16	14	4	2–5 (12 bp) and 27–51 (9 bp) ^d	P and C (7)	312–528	0.924

^a The basic statistics were obtained from the Saldaña collection ($n = 119$).

^b SNP, single nucleotide polymorphism; VNTR, variable number of tandem repeats; HGDI, Hunter–Gaston discriminatory index; P, perfect (numbers of SNPs detected within VNTR); C, compound (number of types detected).

^c From 68 to 394 bp around the VNTR loci.

^d Locus BG02 includes two VNTR loci in a row.

expected, there is a strong correlation between the severity of the disease and our ability to recover isolates, even if we managed to recover a few isolates from asymptomatic panicles, as done in a previous study from China (Zhu et al. 2008). Bacteria invaded the panicles and the grain and increased in density during infection (Pedraza et al. 2018), which clearly explains our success in recovering more easily isolates from diseased panicles. We were, however, unable to detect any correlation between the genotype of the isolates and the health status of the panicle, although our sampling was not strictly adapted to detect such a link. Pathogenicity tests should be carried out to detect the variation in symptoms induced by each bacterial genotype and to see if some of them are more aggressive than others.

We recovered four different *toxA* haplotypes, involving four SNPs among the 109 isolates. Quite surprisingly, the study from Maeda et al. (2006) typed 67 strains of *B. glumae* samples recovered from all over Japan, and in examining the same portion of the sequence as we did, only three SNPs and four haplotypes were found. The diversity of the partial *toxA* gene found in a single field in Colombia is therefore similar, if not higher, than the diversity found throughout Japan. Moreover, by reconstructing the phylogeny (although again based on a limited number of mutations), it appears clear that there is no global geographical pattern of *toxA* diversity (Fig. 2). For instance, *toxA* type 3 is shared by isolates recovered from the U.S.A. (336gr1), Vietnam (NCPB3923), Japan (LMG2196 and others), Korea (BGR1), and Colombia (this study). It is also the case for type 2 (Japan and Colombia) and type 5 (Japan, and a lung isolate taken from a hospital in the U.S.A.). This finding demonstrates that, so far, the toxoflavin diversity is widespread over the different regions from which *B. glumae* isolates have been recovered. Because the majority of the

mutations detected are silent (i.e., no change in the amino acid), we may reject the hypothesis that they have arisen independently by convergence and selection (i.e., are homoplastic). We can then suggest that if the disease really originated in Asia, the bacterium has spread from Asia to America, either once or repeatedly, without significant genetic bottleneck. Conversely, we can also suggest that Asia may not be the center of origin of the disease, which could be America instead. This strong hypothesis, however, requires a much larger sampling, both in Asia and America, to be tested. Finally, it can be noted that the majority of sequences recovered belong to haplotype 3 (59 sequences out of 87 in total), and that this haplotype is the one with the widest geographical distribution. Whether this predominance is the result of greater competitiveness in its ability to infest the host or to infest a wider host spectrum, to better survive in the environment (which could have led to a recent expansion), or is simply random should be tested by artificial inoculations and large sampling in other parts of the world.

It is finally interesting to note that the overall diversity, even over the whole sequence of *toxA*, remains extremely low. This low nucleotide diversity could be the result of high purifying selection because of its role in pathogenicity. We tried to detect evidence of such selection, but the test we applied was not significant; perhaps this was from the very low number of mutations along the sequences, which could restrict the power of this test. Whether these mutations have an effect or not, however, should be studied in depth in the future and their impact in the infection process should be assessed.

At the beginning of this project, we faced one main pitfall to select primers: only two genomes (LMG2196^T and BGR1) were completed at the time of this analysis; the three others remained at the draft level (from 96 to 2,457 scaffolds). The fragmentations of genome sequences

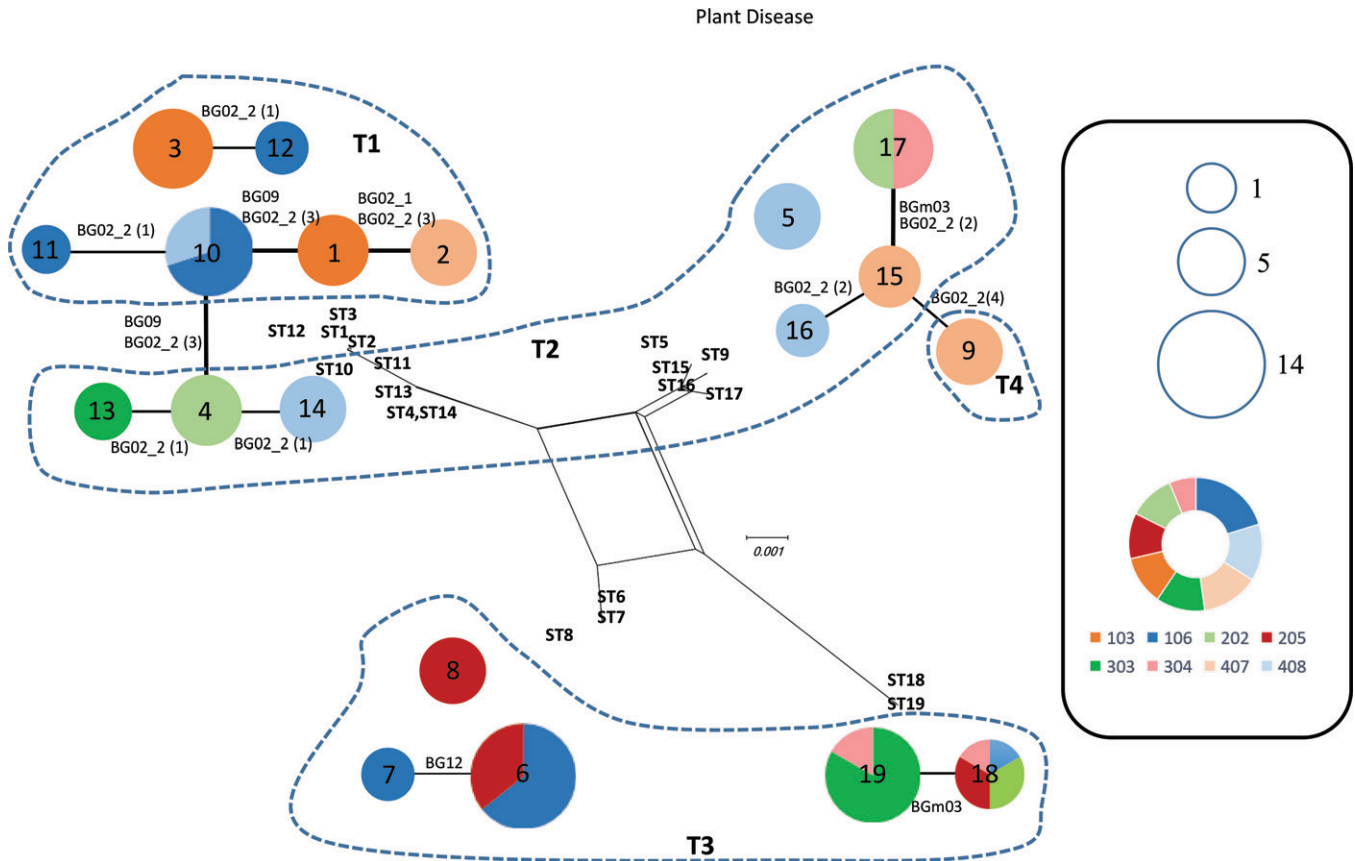


Fig. 3. Network of the 19 haplotypes, reconstructed using the software programs SplitsTree4 and Phyloz v.2. Within each subgroup detected in the SplitsTree software analysis, the network of haplotype was reconstructed using the different haplotypes of all variable numbers of tandem repeat (VNTR) loci and considering the two VNTR loci detected for the BG02 locus independently. The tree cut-off was set up at 3 (that is, we connected only sequence type [ST], which differed by one or two alleles). The thick line connecting STs indicates two differences (i.e., two different VNTR alleles) among them, while the thin line indicates a single locus difference among the five. Each number within a circle is the ST's haplotype code. Circle size is proportional to the number of isolates, with the range of variation of size given in the legend, from one to 14 individuals per haplotype. Colors indicate plot of origin of each haplotype, following the code given in the key (right). The VNTR code along branches between haplotypes indicate, in parentheses, the number of insertion/deletions between the two STs for BG02_2. For the other VNTR loci, it indicates the locus that differ by one-repeat variation between the two haplotypes. The dotted lines frame isolates that share the same *toxA* alleles (numbered from T1 to T4).

are often positioned within the repeated sequences, limiting the possibility of searching for sequences conserved on either side of the VNTR loci to design the primers. We did succeed in designing primers for 14 different loci, but many failed to amplify all of our subsample of isolates. Such a low level of success with our primers was quite disappointing. The primers were not simply designed on the basis of five genomes available in GenBank; we also verified that they were present in three other genomes added after our initial analyses. These eight genomes had been sampled from 1967 (for type strain LMG2196) to 2017 (Bp9029), taken from two continents and seven different countries. We were therefore expecting them to be representative of the species diversity, which was clearly not the case. As suggested earlier (Francis et al. 2013), one explanation for this failure could be that *B. glumae* has a high degree of plasticity in its genomic structure, with several genomic islands. These rearrangements and indels could explain some of the amplification failures.

In a first stage of development of these VNTR loci, we chose to sequence all those that were usable to detect different variants although of the same length. As expected, considering all SNPs along sequences, rather than only the number of repeats, brought more information. Sequencing makes it possible to reveal undetected variants by simple analysis of amplified fragment sizes and thus increase the power of discrimination of the loci. This is visualized by the increase of the HDGI index. Although not sufficient to reconstruct fine genetic relationships among isolates, especially at the field scale, our typing scheme appears useful and was adapted to retrieve a first and reliable image of the genetic diversity detected from a single rice field. However, the combination of the four loci based on repeats number resulted in 18 different multi-locus haplotypes, whereas the full typing including SNPs resulted in 19 different haplotypes. The gain of information when considering point mutations is therefore limited in our study, although the number of different haplotypes increases for each VNTR locus. This is partly from the redundancy of information between loci as well as the very low level of recombination between loci (see below). We can, however, note that this set of VNTR loci is sufficient to visualize a one-off diversification event—with the ST11 haplotype that seems to have emerged within a panicle from the ST10 haplotype, which is detected within another plot as well.

The locus BG02 is highly variable, with 16 alleles detected (based on sequencing). BG02 is atypical compared with others in several features. First, as mentioned before, it is composed of two VNTR loci in a row, separated by a stretch of 13 nucleotides. It is also the single locus among the four that fall within a gene. It is part of a collagen-like protein, characterized by a Gly-Xaa-Yaa repeating amino acid sequence. This gene is similar to that found in animal collagens and underlies their unique triple helical structure (Yu et al. 2014). Collagen-like proteins in the two human pathogenic species, *Burkholderia pseudomallei* and *B. mallei*, have already been described by Bachert et al. (2015). It was suggested that these proteins participate in pathogenesis, based on their non-collagenous domains that are associated with the components of a type-3 secretion apparatus. A type-3 secretion system is also present in *B. glumae*. However, even if the less virulent T3SS-deficient *B. glumae* mutant suggested that the effectors of T3SS are required and must play a role in the pathogenicity, to our knowledge none of them has ever been reported so far in *B. glumae* (Cui et al. 2016). Whether this collagen-like protein is involved in pathogenicity, and the repeated sequence is under selection, remains to be explored.

By aligning sequences for this locus, we detected several examples of homoplasy between haplotypes (showing the same number of repeats, but in different sequences) but found that differences in size between alleles may have resulted from several indel events. For instance, the two haplotypes ST5 and ST9 both have 45 repeats for the BG02 locus but with three indel events (located at different positions within the repeats) between them; this shows that their identical length is homoplastic and emerged independently. Similarly, ST4 and ST6 share 40 repeats but with a highly diverging pattern of indels between them. Conversely, ST13 and ST14 both display 39 repeats but with a single punctual nucleotide difference between them, suggesting their close evolutionary relationship and that the single nucleotide mutation has appeared recently in one isolate. A high number of loci analyzed may counterbalance homoplasy artifacts. However, in our case it could

be easier and more useful to sequence this highly polymorphic and informative locus to enlighten us as to the evolutionary relationship among isolates rather than develop and type several other loci.

Based on five loci (splitting BG02 in two VNTR loci), we recovered 19 different haplotypes that could be clustered in seven different genetic groups as shown in Figure 3. However, the average number of locus variants among haplotypes taken from different groups reach 4.05, while the within group divergence is only 1.22. This large difference illustrates the discontinuity of genetic divergence observed between the different strains according to whether they belong to the same group or to different groups, reflecting the coexistence of several distinct genetic lines in this field. These values, together with the network structure and the linkage disequilibrium estimated, suggest, for this field, a lack of, or a very low level of, recombination among isolates. As we lack data from other sites, it is not clear to what extent this might be from an admixture of several independent lineages newly introduced into the field that will gradually recombine and homogenize over time and generations, or whether *B. glumae* is a species with a very low ability to recombine. One point that could partly resolve the issue is the geographical distribution of haplotypes in the field. After mapping the origin of each haplotype on the network, we found that there is no correlation between genetic proximity and the plot of origin. A plot may contain several divergent haplotypes, just as some haplotypes have been recovered from different plots (Fig. 3). Bacteria may disperse rapidly in the field (which is certainly possible). Alternatively, seeds used for the trials may have been heavily contaminated at the time of sowing. Indeed, previous studies have shown that seeds may contain *B. glumae* isolates (Cottyn et al. 2009; Saylor et al. 2006). However, because there is no effect of the rice variety on the bacterial diversity recovered, we had to hypothesize that the two lots of seeds have been produced under the same bacterial population contamination pressure. Finally, this distribution on the different plots could suggest that these different haplotypes have been present for a long time, and therefore would mean that the degree of recombination among strains must remain very low.

These results suggest key elements in the presence of the bacterium in Colombia. Colombia is a region of major diversity and diversification of *B. glumae* as a rice pathogen, with the genetic diversity observed being the result of migration events from Asia (the region assumed to be the center of origin of the pathogen) and also possibly from other regions where the disease is present, such as northern America. Indeed, the presence of *toxA* haplotypes shared between Asia and Colombia, coupled with the high conservation of sequences of this gene, supports the hypothesis of recurrent migration events between the two continents. This is not surprising, based on the high exchange of rice germplasm for breeding purposes. At the same time, the high number of multi-loci haplotypes and VNTR alleles reflect the diversification process that occurs in Colombia.

If we set aside the hypothesis that the seed lots used were already heavily contaminated, the coexistence of several genetically highly divergent bacterial lineages within a field shows that these bacterial lineages are locally maintained and can reinfest a field year after year, either by wind or by irrigation water. Such a dynamic must play a great role in the epidemiology of the disease and its ability to maintain itself over years in a geographic area. It would be particularly relevant to study the extent to which such diversity may have a role in the dynamics of the disease and its adaptability to the different rice varieties that can be deployed in the field. We have not yet tested for differing virulence or aggressiveness in rice cultivars. However, it certainly might be the case, because previous studies have shown that different strains may have variable toxoflavin production, which affects their ability to colonize the plant, and their particular virulence (Karki et al. 2012). It has been also reported that *B. glumae* can act as a biocontrol agent (some strains are producers of antibiotics) and that may help for survival in the seeds and other tissues (Cottyn et al. 2009). Colombia experiences large year-to-year variations in disease incidence and severity. Our results suggest that this is not from annual variations in the diversity of strains present but more likely the result of environmental fluctuations; previous studies have shown that the disease is dependent on particular environmental conditions for its development.

Acknowledgments

We are very grateful to L. Gagnevin and two anonymous reviewers for their helpful comments on an earlier version of this manuscript and to Isabelle Rimbault for technical help in the lab work.

Literature Cited

- Agarwal, V. K., and Sinclair, J. B. 1980. Principles of Seed Pathology. CRC, Boca Raton, FL.
- Bachert, A. B., Choi, S. J., Snyder, A. K., Rio, R. V. M., Durney, B. C., Holland, L. A., Amemiya, K., Welkos, S. L., Bozue, J. A., Cote, C. K., Berisio, R., and Lukomski, S. 2015. A unique set of the *Burkholderia* collagen-like proteins provides insight into pathogenesis, genome evolution and niche adaptation, and infection detection. PLoS One 10:e0137578.
- Bakhtiari, M., Shleizer-Burko, S., Gymrek, M., Bansal, V., and Bafna, V. 2018. Targeted genotyping of variable number tandem repeats with adVNTR. Genome Res. 28:1709-1719.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10:421.
- Cottyn, B., Debode, J., Regalado, E., Mew, T. W., and Swings, J. 2009. Phenotypic and genetic diversity of rice seed-associated bacteria and their role in pathogenicity and biological control. J. Appl. Microbiol. 107:885-897.
- Cui, Z. Q., Zhu, B., Xie, G. L., Li, B., and Huang, S. W. 2016. Research status and prospect of *Burkholderia glumae*, the pathogen causing bacterial panicle blight. Rice Sci. 23:111-118.
- Excoffier, L., Laval, G., and Schneider, S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinform. Online
- Echeverri-Rico J, Petro E, Fory PA, Mosquera GM, Lang JM, Leach JE, et al. (2021) Understanding the complexity of disease-climate interactions for rice bacterial panicle blight under tropical conditions. PLoS ONE 16(5): e0252061. <https://doi.org/10.1371/journal.pone.0252061>
- Fory, P. A., Triplett, L., Ballen, C., Abello, J. F., Duitama, J., Aricapa, M. G., Prado, G. A., Correa, F., Hamilton, J., Leach, J. E., Tohme, J., and Mosquera, G. M. 2014. Comparative analysis of two emerging rice seed bacterial pathogens. Phytopathology 104:436-444.
- Francis, F., Kim, J., Ramaraj, T., Farmer, A., Rush, M. C., and Ham, J. H. 2013. Comparative genomic analysis of two *Burkholderia glumae* strains from different geographic origins reveals a high degree of plasticity in genome structure associated with genomic islands. Mol. Genet. Genomics 288:195-203.
- Gladieux, P., Ravel, S., Rieux, A., Cros-Arteil, S., Adreit, H., Milazzo, J., Thierry, M., Fournier, E., Terauchi, R., and Tharreau, D. 2018. Coexistence of multiple endemic and pandemic lineages of the rice blast pathogen. MBio 9: e01806-17.
- Goto, K., and Ohata, K. 1956. New bacterial diseases of rice (brown stripe and grain rot). Ann. Phytopathol. Soc. Jpn. 21:46-47.
- Ham, J. H., Melanson, R. A., and Rush, M. C. 2011. *Burkholderia glumae*: next major pathogen of rice? Mol. Plant Pathol. 12:329-339.
- Huson, D. H., and Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254-267.
- Jeong, Y., Kim, J., Kim, S., Kang, Y., Nagamatsu, T., and Hwang, I. 2003. Toxoflavin produced by *Burkholderia glumae* causing rice grain rot is responsible for inducing bacterial wilt in many field crops. Plant Dis. 87:890-895.
- Jiang, N., Yan, J., Liang, Y., Shi, Y.-L., He, Z.-Z., Wu, Y.-T., Zeng, Q., Liu, X.-L., and Peng, J.-H. 2020. Resistance genes and their interactions with bacterial blight/leaf streak pathogens (*Xanthomonas oryzae*) in rice (*Oryza sativa* L.): An updated review. Rice (N. Y.) 13:3.
- Jung, B. N., Park, J. W., Kim, N. G., Li, T. Y., Kim, S. Y., Bartley, L. E., Kim, J. Y., Kim, I. Y., Kang, Y. H., Yun, K. H., Choi, Y. H., Lee, H.-H., Ji, S.-Y., Lee, K. S., Kim, B. Y., Shon, J. C., Kim, W. C., Liu, K.-H., Yoon, D., Kim, S.-K., Seo, Y.-S., and Lee, J.-K. 2018. Cooperative interactions between seed-borne bacterial and air-borne fungal pathogens on rice. Nat. Commun. 9:31.
- Karki, H. S., Shrestha, B. K., Han, J. W., Groth, D. E., Barphagha, I. K., Rush, M. C., Melanson, R. A., Kim, B. S., and Ham, J. H. 2012. Diversities in virulence, antifungal activity, pigmentation and DNA fingerprint among strains of *Burkholderia glumae*. PLoS One 7:e45376.
- King, E. O., Ward, M. K., and Raney, D. E. 1954. Two simple media for the demonstration of pyocyanin and fluorescein. J. Lab. Clin. Med. 44:301-307.
- Kurita, T., and Tabei, H. 1967. On the pathogenic bacterium of bacterial grain rot of rice. Ann. Phytopathol. Soc. Jpn. 33:111.
- Lee, Y. H., Ko, S. J., Cha, K. H., and Park, E. W. 2015. BGRcast: A disease forecast model to support decision-making for chemical sprays to control bacterial grain rot of rice. Plant Pathol. J. 31:350-362.
- Li, L., Wang, L., Liu, L., Hou, Y., Li, Q., and Huang, S. 2016. Infection process of *Burkholderia glumae* before booting stage of rice. J. Phytopathol. 164:825-832.
- Librado, P., and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452.
- Lim, J., Lee, T. H., Baek, H. N., Yang, D. C., Kim, M., and Hwang, I. 2009. Complete genome sequence of *Burkholderia glumae* BGR1. J. Bacteriol. 191: 3758-3759.
- Maeda, Y., Shinohara, H., Kiba, A., Ohnishi, K., Furuya, N., Kawamura, Y., Ezaki, T., Vandamme, P., Tsushima, S., and Hikichi, Y. 2006. Phylogenetic study and multiplex PCR-based detection of *Burkholderia plantarii*, *Burkholderia glumae* and *Burkholderia gladioli* using *gyrB* and *rpoD* sequences. Int. J. Syst. Evol. Microbiol. 56:1031-1038.
- Mondal, K. K., Mani, C., and Verma, G. 2015. Emergence of bacterial panicle blight caused by *Burkholderia glumae* in North India. Plant Dis. 99:1268.
- Nandakumar, R., Shahjahan, A. K. M., Yuan, X. L., Dickstein, E. R., Groth, D. E., Clark, C. A., Cartwright, R. D., and Rush, M. S. 2009. *Burkholderia glumae* and *B. gladioli* cause bacterial panicle blight in rice in the southern United States. Plant Dis. 93:896-905.
- Nascimento, M., Sousa, A., Ramirez, M., Francisco, A. P., Carriço, J. A., and Vaz, C. 2017. PHYLOViZ 2.0: Providing scalable data integration and visualization for multiple phylogenetic inference methods. Bioinformatics 33: 128-129.
- Pedraza, L. A., Bautista, J., and Uribe-Vélez, D. 2018. Seed-born *Burkholderia glumae* infects rice seedling and maintains bacterial population during vegetative and reproductive growth stage. Plant Pathol. J. 34:393-402.
- Riera-Ruiz, C., Vargas, J., Cedeño, C., Quirola, P., Escobar, M., Cevallos-Cevallos, J. M., Ratti, M., and Peralta, E. L. 2014. First report of *Burkholderia glumae* causing bacterial panicle blight on rice in Ecuador. Plant Dis. 98:988.
- Rousset, F. 2008. Genepop'007: A complete re-implementation of the Genepop software for Windows and Linux. Mol. Ecol. Resour. 8:103-106.
- Sayler, R. J., Cartwright, R. D., and Yang, Y. 2006. Genetic characterization and real-time PCR detection of *Burkholderia glumae*, a newly emerging bacterial pathogen of rice in the United States. Plant Dis. 90:603-610.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., van Horn, D. J., and Weber, C. F. 2020. Introducing Mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75:7537-7541.
- Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S. L. 2015. Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol. Biol. Evol. 32:1342-1353.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30: 2725-2729.
- Trung, H. M., Van, N. V., Vien, N. V., Lam, D. T., and Lien, M. 1993. Occurrence of rice grain rot disease in Vietnam. Int. Rice Res. Notes 18:30.
- Tsushima, S. 2011. Study on control and epidemiology of bacterial grain rot of rice. Jpn. Agric. Res. Q. 77:358.
- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J. A. M. 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. 35 (Web Server):W71-W74.
- Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., and Kosakovsky Pond, S. L. 2018. Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. Mol. Biol. Evol. 35:773-777.
- Yu, Z., An, B., Ramshaw, J. A., and Brodsky, B. 2014. Bacterial collagen-like proteins that form triple-helical structures. J. Struct. Biol. 186:451-461.
- Zhu, B., Lou, M.-M., Yan, H., Xie, G.-L., Luo, J.-Y., and Xu, L.-H. 2008. Isolation and identification of *Burkholderia glumae* from symptomless rice seeds. Rice Sci. 15:145-149.