

## Comparative Genome Analysis of *Rathayibacter tritici* NCPPB 1953 with *Rathayibacter toxicus* Strains Can Facilitate Studies on Mechanisms of Nematode Association and Host Infection

Jungwook Park<sup>1†</sup>, Pyeong An Lee<sup>2†</sup>, Hyun-Hee Lee<sup>1</sup>, Kihyuck Choi<sup>2</sup>, Seon-Woo Lee<sup>2\*</sup>, and Young-Su Seo<sup>1\*</sup>

<sup>1</sup>Department of Microbiology, Pusan National University, Busan 46241, Korea

<sup>2</sup>Department of Applied Bioscience, Dong-A University, Busan 49315, Korea

(Received on January 25, 2017; Revised on April 12, 2017; Accepted on April 23, 2017)

*Rathayibacter tritici*, which is a Gram positive, plant pathogenic, non-motile, and rod-shaped bacterium, causes spike blight in wheat and barley. For successful pathogenesis, *R. tritici* is associated with *Anquina tritici*, a nematode, which produces seed galls (ear cockles) in certain plant varieties and facilitates spread of infection. Despite significant efforts, little research is available on the mechanism of disease or bacteria-nematode association of this bacterium due to lack of genomic information. Here, we report the first complete genome sequence of *R. tritici* NCPPB 1953 with diverse features of this strain. The whole genome consists of one circular chromosome of 3,354,681 bp with a GC content of 69.48%. A total of 2,979 genes were predicted, comprising 2,866 protein coding genes and 49 RNA genes. The comparative genomic analyses between *R. tritici* NCPPB 1953 and *R. toxicus* strains identified 1,052 specific genes in *R. tritici* NCPPB 1953. Using the BlastKOALA database, we revealed that the flexible genome of *R. tritici* NCPPB 1953 is highly enriched in 'Environmental Information Processing' system and metabolic processes for diverse substrates.

Furthermore, many specific genes of *R. tritici* NCPPB 1953 are distributed in substrate-binding proteins for extracellular signals including saccharides, lipids, phosphates, amino acids and metallic cations. These data provides clues on rapid and stable colonization of *R. tritici* for disease mechanism and nematode association.

**Keywords** : comparative genome, *Rathayibacter toxicus*, *Rathayibacter tritici* NCPPB 1953

**Handling Associate Editor** : Sohn, Kee Hoon

As a Gram positive plant pathogen, the genus *Rathayibacter* is a member of coryneform bacteria belonging to the family *Microbacteriaceae* of phylum *Actinobacteria*, previously attributed to genus *Clavibacter* (Davis et al., 1984). Many coryneform phytopathogenic bacteria were initially classified as the genus *Corynebacterium* (Dowson, 1942), while many properties of the coryneform plant pathogenic bacteria were different from those of *Corynebacterium sensu stricto* (Evtushenko and Dorofeeva, 2012). Later, plant pathogenic *Corynebacterium* species were reclassified into the genus *Clavibacter* due to the presence of DAB, which is a particular component in peptidoglycan cell wall group B (Carlson and Vidaver, 1982; Davis et al., 1984). In 1993, Zgurskaya and colleagues proposed the genus *Rathayibacter* such as *R. tritici*, *R. rathayi*, and *R. iranicus* separating from the genus *Clavibacter* based on DNA-DNA hybridization, chemotaxonomic studies, and numerical analysis of bacterial phenotypes (Evtushenko and Dorofeeva, 2012; Zgurskaya et al., 1993). Separation of *Rathayibacter* species from *Clavibacter* was further supported by the analyses of 16S

<sup>†</sup>These authors contributed equally to this work as first authors.

\*Co-corresponding authors.

SW Lee

Phone) +82-51-200-7551, FAX) +82-51-200-7505

E-mail) seonlee@dau.ac.kr

YS Seo

Phone) +82-51-510-2267, FAX) +82-51-514-1778

E-mail) yseo2011@pusan.ac.kr

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

rRNA gene sequences (Rainey et al., 1994; Takeuchi and Yokota, 1994). The species *Rathayibacter toxicus* was also classified in this genus by Sasaki and colleagues in 1998 (Sasaki et al., 1998). Furthermore, these species of *Rathayibacter* have been shown to differ absolutely in many physiological phenotypes, including cell-wall compositions, multilocus enzyme profiles, and pigmentations (Davis et al., 1984; De Bruyne et al., 1992; Lee et al., 1997; Riley et al., 1988; Zgurskaya et al., 1993).

*R. tritici* has been identified as a causative agent of spike blight, also called yellow ear rot, yellow slime rot, or Tundu disease, in wheat and barley (Paruthi and Gupta, 1987). Plant diseases caused by *R. tritici* result in economic losses in many countries worldwide, including Australia (Riley and Reardon, 1995), China, Cyprus, Iran (Bradbury, 1986; Duveiller and Fucikovsky, 1997; Mehta, 2014) and Pakistan (Akhtar, 1987). Interestingly, *R. tritici* shows an association with *Anguina tritici*, a nematode, which produces seed galls (ear cockles) in certain plant varieties and facilitates spread of *R. tritici* infection (Paruthi and Bhatti, 1985). Intact seed galls produced by nematodes display the greatest grain loss due to development of *R. tritici* (Fattah, 1988). Similarly, a bacterium, *R. toxicus*, causing a gumming disease and ryegrass toxicity is associated with a nematode vector (Riley and Ophel, 1992). Compared to *R. tritici*, *R. toxicus* is commonly found in annual ryegrass with *A. funesta* as a nematode vector, rabbit-foot grass and annual browngress with undescribed *Anguina* species. While host plants for *R. toxicus* are limited to ryegrasses in nature, it is likely that *R. toxicus* is not host specific experimentally (Agarkova et al., 2006). In addition to unique host plants for *R. toxicus*, *R. toxicus* is different from *R. tritici* for the production of glycolipid toxins, known as corynetoxins, in the infected ryegrass, and thus causes a lethal toxicosis in the animal that consumed the infected plants (Agarkova et al., 2006; Jago and Culvenor, 1987). Both in *R. tritici* and *R. toxicus*, bacterial association with a nematode vector could be used as a good model of interaction between plant pathogens and nematodes for successful pathogenesis. Despite significant efforts and interesting association between bacterial pathogen and nematodes, little research is available on the mechanism of disease or bacteria-nematode association of this bacterium due to lack of genomic information.

Characterising the complete genome of *R. tritici* could help to resolve these gaps in our understanding. Here, we report the first whole genome sequence of *R. tritici* NCPPB 1953 isolated from wheat seeds. The whole genome of *R. tritici* NCPPB 1953 harbours one circular chromosome of 3,354,681 bp with 2,866 protein coding genes. The comparative genomic analysis between *R. tritici*

NCPPB 1953 and *R. toxicus* strains showed a lot of 1,052 *R. tritici* NCPPB 1953-specific genes in flexible genome, which refers to genes present in two or more organisms or specific to a single organism in a pan-genome (Medini et al., 2005; Sternes and Borneman, 2016). We revealed that these genes are responsible for 'Environmental Information Processing' system and metabolic processes for diverse substrates, which might explain the rapid and stable growth of *R. tritici* rather than *R. toxicus*. Furthermore, distinctive genetic features of *R. tritici* NCPPB 1953 are distributed in substrate-binding proteins for extracellular signals. More research is required to better understand the pathogenesis of *R. tritici* and association with the nematode, and complete genomic information of *R. tritici* NCPPB 1953 will provide a valuable foundation for diverse biological experiments.

---

## Materials and Methods

### Growth conditions and genomic DNA preparation.

Pure culture of *R. tritici* NCPPB 1953 was grown on nutrient broth yeast extract (NBY) media (8 g nutrient broth, 2 g yeast extract, 2 g K<sub>2</sub>HPO<sub>4</sub>, 0.5 g KH<sub>2</sub>PO<sub>4</sub>, 2.5 g glucose per liter, followed by autoclaving and supplementation with 1 ml of 1 M MgSO<sub>4</sub>·7H<sub>2</sub>O) media for 96 h at 25°C (Zgurskaya et al., 1993). For extraction of genomic DNA, the pellet of bacterial cells grown on the agar plate was harvested with sterilized water. DNA was subsequently isolated using the Promega Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA) following the standard protocol provided by the manufacturer. DNA was visualized in ethidium bromide-stained 0.7% agarose gel, and the concentration and purity of DNA were determined by a NanoDrop™ spectrophotometer.

**Genome sequencing of *R. tritici* NCPPB 1953.** Third generation DNA sequencing of the SMRT technology (Pacific Biosciences, Menlo Park, CA, USA) was used to establish the complete genome sequence of *R. tritici* NCPPB 1953 (Chin et al., 2013). A total of 150,292 raw reads containing an average of 7,236 bp were generated by a 20-kb insert SMRTbell standard library. To correct errors, long reads were selected as seeds, and other short reads were aligned into seeds by the basic local alignment with successive refinement step (Chaisson and Tesler, 2012). After filtration, post-filtered reads included 86,775 reads and represented an average of 11,144 bp with a quality of 0.852. The *de novo* assembly of post-filtered reads was conducted using the HGAP pipeline from SMRT-Analysis with default parameters (Chin et al., 2013). Circularization was verified, and overlapping ends were trimmed (Kopf et al., 2014). Assembly resulted

in a single contig with a circular form of 3,354,681 bp. Subsequently, to evaluate assembly quality, all reads were mapped back to the chromosomal sequence according to the RS\_Resequencing protocol. Final results indicate bases called value of 99.88%, consensus concordance of 100%, and 187.88-fold coverage.

**Gene annotation of *R. tritici* NCPPB 1953.** Genomic rRNAs and tRNAs were predicted using RNAmmer (Lagesen et al., 2007) and tRNAscan-SE (Lowe and Eddy, 1997), respectively. All open reading frames (ORFs) and pseudogenes were predicted using Glimmer (Delcher et al., 1999), GeneMarkHMM (Lukashin and Borodovsky, 1998) and Prodigal (Hyatt et al., 2010). Predicted ORFs with nucleotides were translated into amino acid sequences and then searched using the BLAST algorithm (Altschul et al., 1990) against the NCBI-NR (Benson et al., 2000), Pfam (Finn et al., 2016) and UniProt (UniProt Consortium, 2013) databases for a description of each protein. In addition, the KEGG (Kanehisa and Goto, 2000) and COG (Galperin et al., 2015) databases were used to construct functional categories in accordance with biological systems. Genes in internal clusters were detected using BLASTclust (Alva et al., 2016) with significant cutoffs of 70% coverage and 30% identity. Signal peptides and transmembrane helices were based on SignalP (Petersen et al., 2011) and TMHMM (Krogh et al., 2001), respectively. CRISPRFinder was used to identify the CRISPR (Grissa et al., 2007). Predicted genes were compared with the Pathogen-Host Interaction (PHI) database (Urban et al., 2015) using the BLASTp method and an *e*-value cutoff of  $1.0 \times 10^{-5}$  (Buiate et al., 2017). The annotation results were verified using the Artemis (Rutherford et al., 2000).

**Nucleotide sequence accession number of *R. tritici* NCPPB 1953.** The complete genome of *R. tritici* NCPPB 1953 has been deposited in GenBank (Benson et al., 2000) under accession number CP015515.

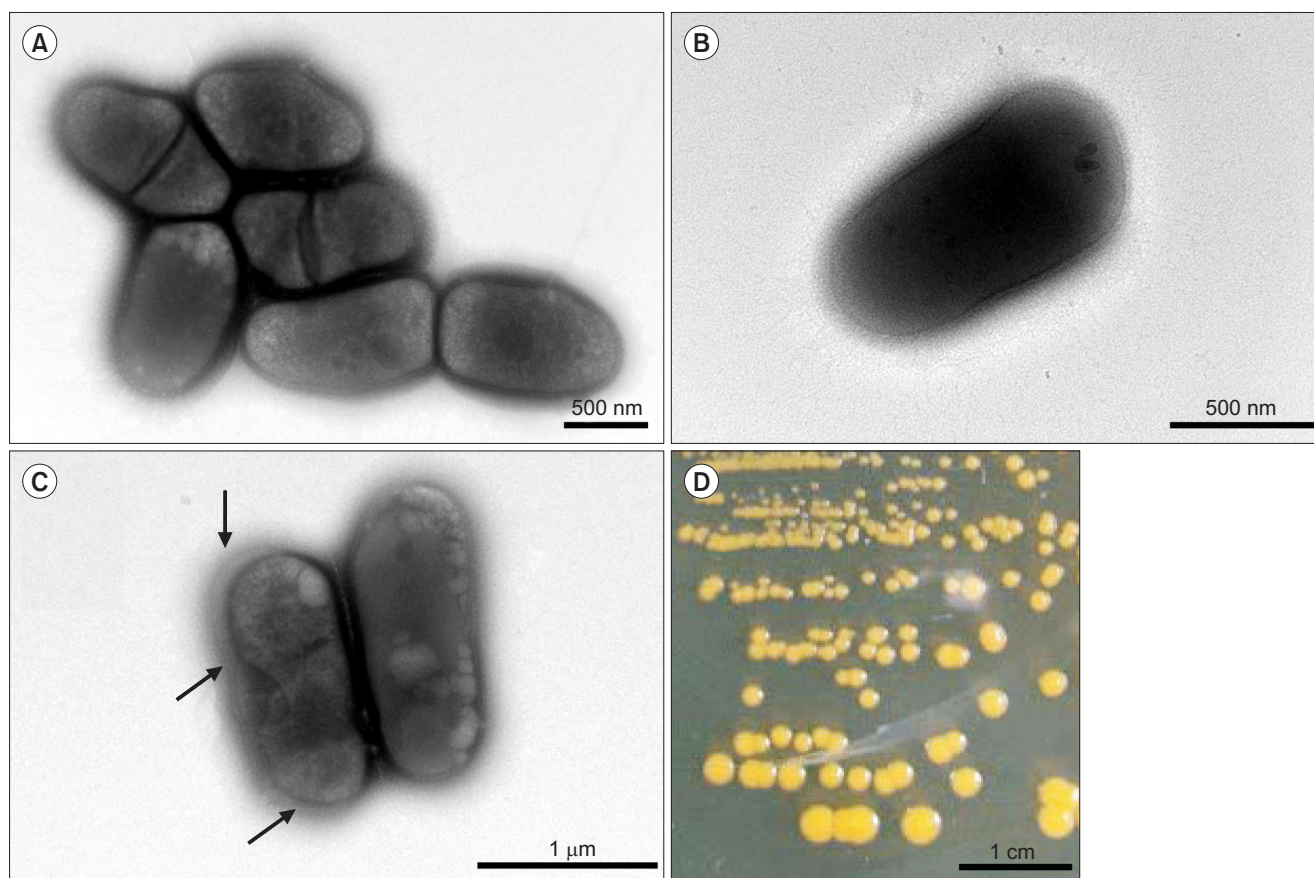
**Phylogenetic analysis.** To evaluate evolutionary distances among *R. tritici* NCPPB 1953 and closely related bacteria in the genus *Rathayibacter*, we collected all 16S rRNA sequences of 26 bacteria from the SILVA database. The ClustalW tool was used to align all sequences with default parameters (Larkin et al., 2007). Phylogenetic tree was conducted by the Maximum Likelihood algorithm based on the Tamura-Nei model as implemented in MEGA 7.0 tool (Kumar et al., 2016). The corresponding parameter of the Maximum Likelihood algorithm was set as ‘complete deletion’. The evolutionary test was performed by bootstrap analysis with 1,000 replications.

**Comparative genome analysis.** For this analyses, three complete genome of *R. tritici* NCPPB 1953 (accession no. CP015515), *R. toxicus* WAC3373 (accession no. CP013292), and *R. toxicus* 70137 (accession no. CP010848) were obtained from GenBank (Benson et al., 2000). We employed the Mauve tool to compare whole genome sequences (Darling et al., 2004). The progressive Mauve algorithm was used to identify genomic rearrangements among whole genome sequences. The alignment was performed using default parameters.

**Pan-genome analysis.** Genomic information of *R. tritici* NCPPB 1953 was formatted for a reference database using the formatdb tool of BLAST (Altschul et al., 1990). The local BLASTp compared total protein genes of *R. toxicus* WAC3373 and *R. toxicus* 70137 to reference database of *R. tritici* NCPPB 1953 through amino acid sequences. The BLAST outputs were filtered with significant criteria of 50% coverage, 50% identity, and an *e*-value cutoff of  $1.0 \times 10^{-5}$  to exclude random hits (Anderson and Brass, 1998). By sorting *e*-value, the best matched subject was selected from total results per each query. Subsequently, to investigate biological meaning of *R. tritici* NCPPB 1953-specific genes, we used the BlastKOALA tool as an automatic annotation server for genome and metagenome sequences (Kanehisa et al., 2016). The BlastKOALA performed KEGG orthology assignments to characterize individual gene functions and reconstruction of KEGG pathways and modules in *R. tritici* NCPPB 1953. Also, we constructed networks of some genes related to transport system using the Cytoscape tool (<http://www.cytoscape.org/>).

## Results and Discussion

**General features of *R. tritici* NCPPB 1953.** *R. tritici* NCPPB 1953 is a Gram positive, non-motile, and rod-shaped bacterium. Its size is 1.1–1.6  $\mu\text{m}$  long and 0.5–0.6  $\mu\text{m}$  wide (Fig. 1). On an aerobic NBY agar plate, *R. tritici* NCPPB 1953 formed round colonies and produced yellow pigment within 96 h at 25°C (Fig. 1D) and could grow in a pH range of 6–8. This strain was a non-motile bacterium and cells of *R. tritici* NCPPB 1953 were short, straight or slightly curved rods with blunt ends. Cells occurred singly, in pairs or sometimes in aggregates (Fig. 1A, B). This is likely because divided cells by binary fission failed to separate after septum formation. Cells were surrounded by a capsule and the capsular material sometimes exhibited distinct extracellular outer layers (Fig. 1C; black arrows). Bacterial colonies were often variable in size, probably because of capsules and bacterial cell aggregates.



**Fig. 1.** Image of transmission electron microscope (TEM) photomicrograph (A–C) and the appearance of colony morphology on solid media (D) of *Rathayibacter tritici* NCPPB 1953. Cells were grown in NBY agar plate for 96 h. The magnification rates of TEM for A–C were 15,000 $\times$ , 25,000 $\times$ , and 10,000 $\times$ , respectively. Bacterial capsules with extracellular outer layer were observed (black arrows).

#### Phylogenetic analysis of *R. tritici* NCPPB 1953 with closely-related bacteria species.

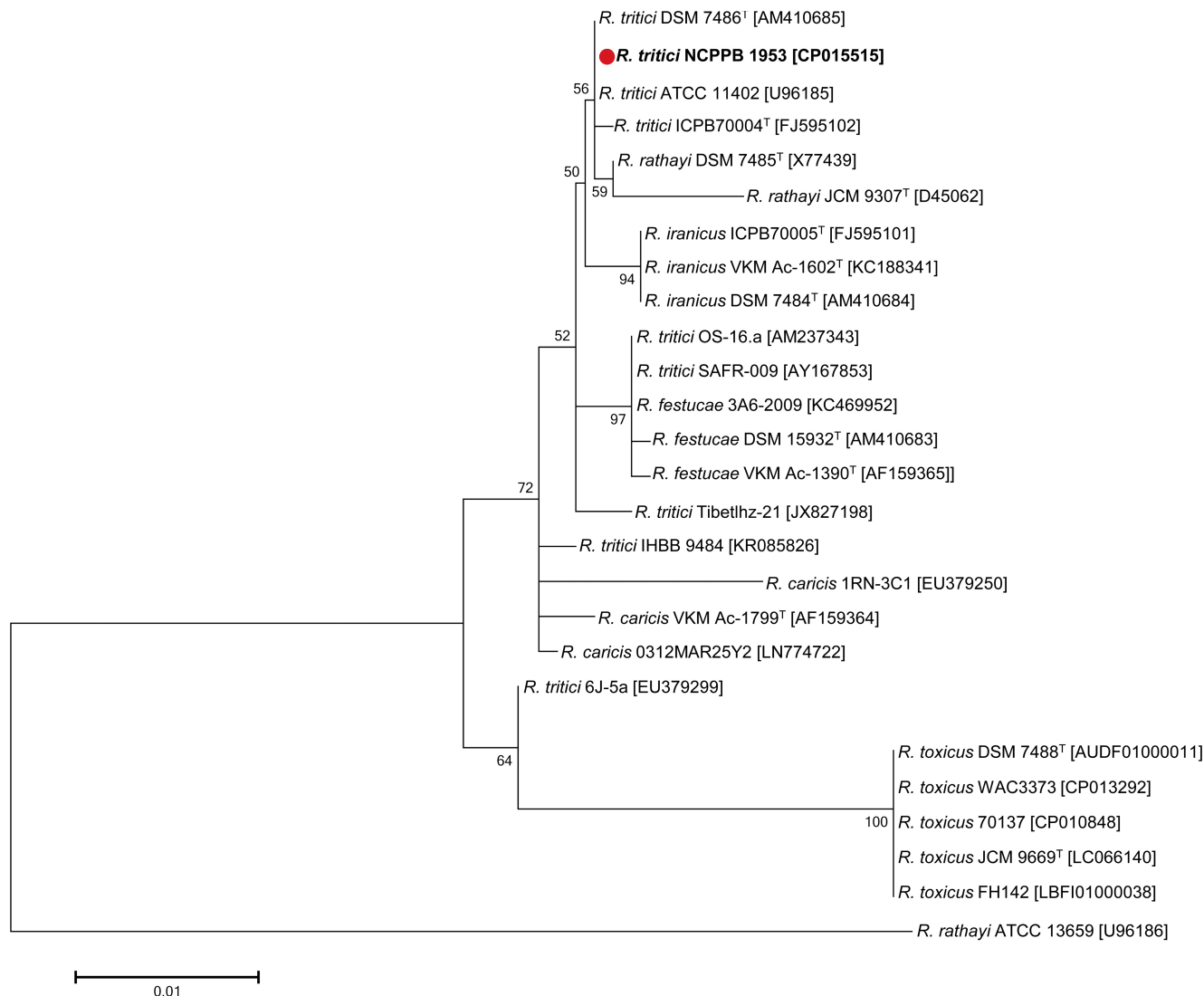
Phylogenetic analysis was performed using the 16S rRNA genes of 26 different bacteria in the genus *Rathayibacter*: 3 strains, *R. caricis*; 3 strains, *R. festucae*; 3 strains, *R. iranicus*; 3 strains, *R. rathayi*; 5 strains, *R. toxicus*; 9 strains, *R. tritici* (Fig. 2). All 16S rRNA sequences were downloaded from the SILVA (Quast et al., 2013), providing information on small (16S/18S, SSU) and large (23S/28S, LSU) subunit ribosomal RNA databases. Pairwise distances were estimated using the Maximum Composite Likelihood approach (Supplementary Table 1). All positions containing gaps and missing data were eliminated. There were a total of 994 positions in the final dataset. As a result, *R. tritici* NCPPB 1953 was grouped with most of *R. tritici* strains. *R. tritici* ATCC 11402 and *R. tritici* DSM 7486<sup>T</sup> represented the closest bacteria as both distance value 0.000, followed by *R. rathayi* DSM 7485<sup>T</sup> (distance value 0.001), *R. tritici* ICPB70004 (distance value 0.001), and *R. iranicus* DSM 7484<sup>T</sup> (distance value 0.003) (Supplementary

Table 1).

#### General genomic features of *R. tritici* NCPPB 1953.

The complete genome of *R. tritici* NCPPB 1953 consists of a single circular form of 3,354,681 bp with 69.48% GC content. There are 2 rRNA operons, 43 tRNAs, and 64 pseudogenes (Table 1). The chromosome contains 2,866 protein-coding genes. Of these, 1,712 proteins are assigned to functional annotations, whereas 1,154 proteins are considered as hypothetical proteins. Table 2 shows the number of genes associated with general COG functional categories. The ‘General function prediction only’ (9.70%) represents the largest category, followed by ‘Amino acid transport and metabolism’ (8.27%), ‘Carbohydrate transport and metabolism’ (8.20%), ‘Transcription’ (6.84%), and ‘Translation, ribosomal structure, and biogenesis’ (6.42%). A complete genomic map of *R. tritici* NCPPB 1953 was constructed using the CGView tool in Fig. 3.

So far, it is hard to find virulence factors of *R. tritici* NCPPB 1953, since there is a lack of research on the



**Fig. 2.** 16S rRNA phylogenetic tree showing the relationships of *Rathayibacter tritici* NCPPB 1953 (shown in bold print) with other members in the genus *Rathayibacter*. Sequences were aligned using ClustalW and the phylogeny was inferred from the alignment of 1,267 bp of 16S rRNA using the Maximum Likelihood method based on the Tamura-Nei model with MEGA 7.0 software. The bootstrap consensus tree from 1,000 replicates was performed to assess the support of the clusters. The scale bar indicates 0.01 nucleotide change per nucleotide position. The GenBank accession numbers are shown in parentheses.

pathogenesis and virulence mechanisms of *R. tritici*. We therefore performed detection of putative virulence genes in the *R. tritici* NCPPB 1953 genome using the PHI database that archives the list of experimentally verified pathogenicity, virulence, and effectors from diverse microbes (Urban et al., 2015). A total of 676 putative virulence genes were identified by the significant cutoff of  $1.0 \times 10^{-5}$  *e*-value (Table 1). *R. tritici* NCPPB 1953 had about 200 more virulence genes than *R. toxicus* strains (447 virulence genes in WAC3373 and 445 virulence genes in 70137). Detailed information with putative virulence genes of *R. tritici* NCPPB 1953 was listed in Supplemen-

tary Table 2. Most of putative virulence genes in *R. tritici* NCPPB 1953 were distributed in many plant pathogenic bacteria and fungi. *Magnaporthe oryzae*, which causes a serious disease affecting rice (Wilson and Talbot, 2009), represented the largest category matched with 58 genes of *R. tritici* NCPPB 1953, followed by *Xanthomonas citri* (matched with 36 genes), *Fusarium graminearum* (matched with 32 genes), and *Pseudomonas cichorii* (matched with 21 genes). In particular, the *X. citri* mutant of the gene *pstB* (encoding the ABC phosphate transporter subunit), which are matched with 30 genes of *R. tritici* NCPPB 1953, revealed a complete absence of symptoms

**Table 1.** Genome statistics in *Rathayibacter tritici* NCPPB 1953 and *Rathayibacter toxicus* strains

Attribute	<i>R. tritici</i> NCPPB 1953	<i>R. toxicus</i> WAC3373	<i>R. toxicus</i> 70137
Genome size (bp)	3,354,681	2,346,032	2,328,288
DNA coding (bp)	2,830,330	1,970,833	1,957,820
DNA G + C (bp)	2,330,883	1,442,241	1,431,482
DNA scaffolds	1	1	1
Total genes	2,979	2,137	2,120
Protein coding genes	2,866	1,999	1,993
RNA genes	49	51	51
Pseudo genes	64	75	70
Genes in internal clusters	780	309	303
Genes with functional prediction	1,712	1,174	1,158
Genes assigned to COGs	2,238	1,497	1,499
Genes with Pfam domains	2,159	1,575	1,572
Genes with signal peptides	218	117	117
Genes with transmembrane helices	694	497	495
CRISPR repeats	0	2	2
Putative virulence genes	676	447	445

**Table 2.** A distribution of genes associated with COG categories in *Rathayibacter tritici* NCPPB 1953

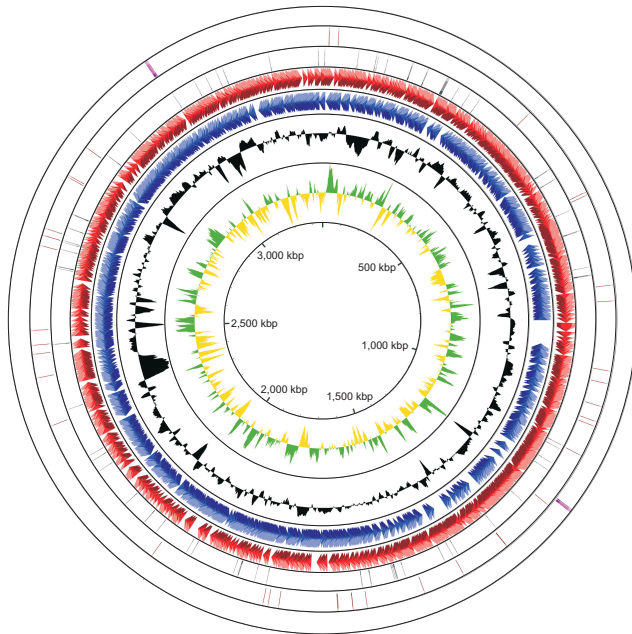
Code	Value	Percentage	Description
J	184	6.42	Translation, ribosomal structure and biogenesis
A	1	0.03	RNA processing and modification
K	196	6.84	Transcription
L	126	4.40	Replication, recombination, and repair
B	0	0.00	Chromatin structure and dynamics
D	35	1.22	Cell cycle control, cell division, chromosome partitioning
V	68	2.37	Defense mechanisms
T	113	3.94	Signal transduction mechanisms
M	149	5.20	Cell wall/membrane biogenesis
N	16	0.56	Cell motility
U	20	0.70	Intracellular trafficking and secretion
O	109	3.80	Posttranslational modification, protein turnover, chaperones
C	110	3.84	Energy production and conversion
G	235	8.20	Carbohydrate transport and metabolism
E	237	8.27	Amino acid transport and metabolism
F	91	3.18	Nucleotide transport and metabolism
H	147	5.13	Coenzyme transport and metabolism
I	116	4.05	Lipid transport and metabolism
P	130	4.54	Inorganic ion transport and metabolism
Q	76	2.65	Secondary metabolites biosynthesis, transport, and catabolism
R	278	9.70	General function prediction only
S	130	4.54	Function unknown
-	628	21.91	Not in COGs

(Moreira et al., 2015). It was also reported that the gene *ABC4* (matched with 16 genes of *R. tritici* NCPPB 1953) of ABC superfamily of membrane transporters is required for the pathogenicity of *M. oryzae*, helping the fungus to

cope with the cytotoxic environment within host (Gupta and Chattoo, 2008). This data suggests that many virulence genes in *R. tritici* NCPPB 1953 are implicated to play a role in the membrane transporters to control efflux

and influx of host metabolites for their pathogenicity.

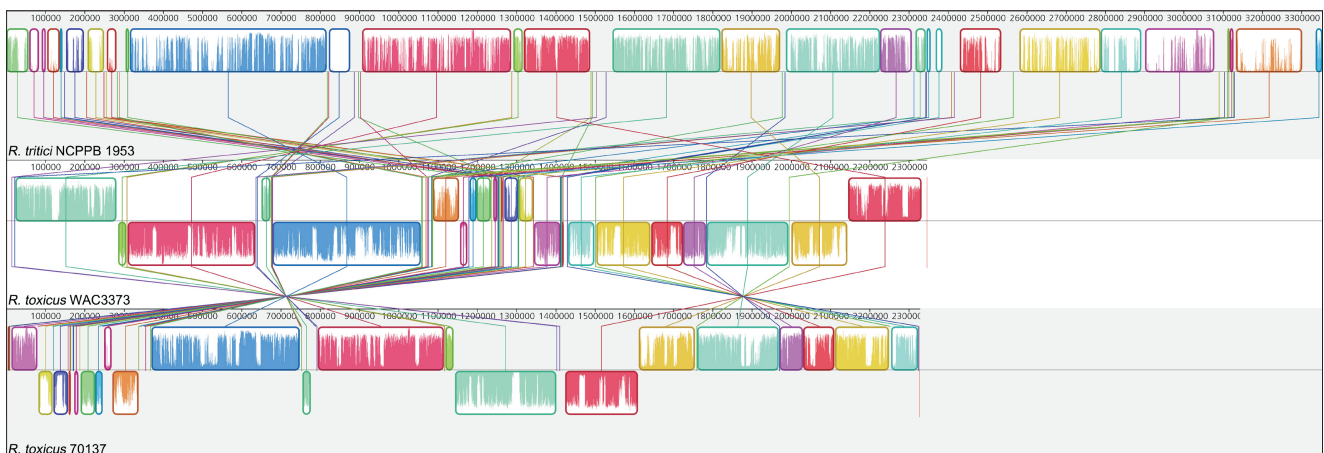
**The comparative genomic analysis between *R. tritici* NCPPB 1953 and *R. toxicus* strains.** *R. toxicus* shares



**Fig. 3.** Graphical circular map of *Rathayibacter tritici* NCPPB 1953 chromosome displaying relevant genetic features. From centre to outside: 1, positive (green) and negative (yellow) of GC skew; 2, GC content (black); 3, CDSs on reverse strand (blue); 4, CDSs on forward strand (red); 5, pseudogenes (grey); 6, tRNAs (brown); 7, rRNAs (pink). This figure was built using the CGView tool.

major biological properties compared to *R. tritici*. For examples, a gumming disease was reported as a common plant disease caused by both *R. tritici* and *R. toxicus* (Agarkova et al., 2006; Arif et al., 2016). Also, *R. toxicus* uses several species of *Anguina* as the vector system to carry themselves into the plant host like *R. tritici* (Price et al., 1979; Riley, 1992). Riley and Reardon (1995) discussed the potential of *R. tritici* as a biological control for *R. toxicus* because of co-occurrence of two bacteria. Thus, *R. toxicus* is the important source for comparative analysis with *R. tritici* and we intended to find distinctive genetic features of *R. tritici* NCPPB 1953 through comparative genomic analysis between *R. tritici* and *R. toxicus*. Among *R. toxicus* strains, there are two complete genomes in the GenBank (Benson et al., 2000) including *R. toxicus* WAC3373 and *R. toxicus* 70137 under accession number CP013292 and CP010848, respectively. Based on the pairwise BLAST comparison of the 16S rRNA gene sequences, *R. tritici* NCPPB 1953 was closely related to *R. toxicus*, resulting in 100% coverage and 98% identity in both comparisons with *R. toxicus* WAC3373 and *R. toxicus* 70137. A similar distribution of rRNA and tRNA genes was maintained as follows: 6 rRNAs and 43 tRNAs, *R. tritici* NCPPB 1953; 6 rRNAs and 45 tRNAs, *R. toxicus* WAC3373; 6 rRNAs and 45 tRNAs, *R. toxicus* 70137.

However, there are different features between *R. tritici* and *R. toxicus* in the genome-wide level. GC contents of whole genome sequence showed a large discrepancy between *R. tritici* (GC content 69.5%) and *R. toxicus* (GC contents 61.5% in WAC3373 and 61.5% in 70137). Above



**Fig. 4.** Whole genome alignment of *Rathayibacter tritici* NCPPB 1953, *Rathayibacter toxicus* WAC3373, and *Rathayibacter toxicus* 70137. The top, middle, and bottom sequences represent each genomic sequence of *R. tritici* NCPPB 1953, *R. toxicus* WAC3373, and *R. toxicus* 70137, respectively. Identically coloured boxes by same lines are locally collinear blocks, indicating homologous and conservative regions in both genomes. Coloured bars inside boxes are related to the level of sequence similarities. Numbers above the map show nucleotide positions.

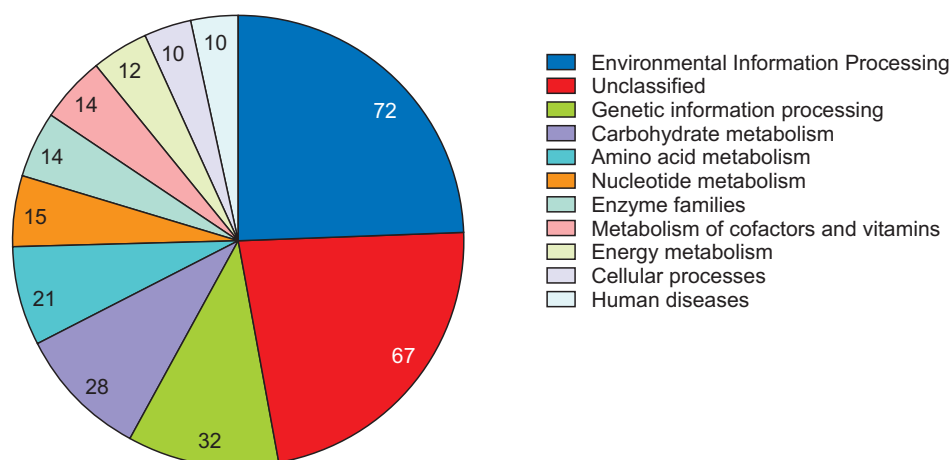
all, *R. tritici* NCPPB 1953 harbours a larger genomic size (about 3.35 Mbp) and more genes (2,979 genes), when compared to *R. toxicus* WAC3373 (about 2.35 Mbp and 2,137 genes) and *R. toxicus* 70137 (about 2.33 Mbp and 2,120 genes). Indeed, we revealed the dynamic genomic rearrangement of whole genome sequence between *R. tritici* and *R. toxicus* using the Mauve aligner (Darling et al., 2004). In Fig. 4, *R. tritici* NCPPB 1953 shared a total of 58 locally collinear blocks (LCBs), which are highly homologous regions among genome sequences, in comparison to *R. toxicus*. Although these LCBs covered most parts of the whole genome in *R. tritici* NCPPB 1953, most of LCBs showed sporadic inversion and irregular genomic rearrangement caused by local and/or large-scale changes in genetic loci. In addition to the variable rearrangement, we confirmed additional genetic differences in *R. tritici* NCPPB 1953 such as non-matched regions and some LCBs of low similarities. Medini et al. (2005) reported that conserved core regions of microbial genome are responsible for basic aspects of the biology related to essential phenotypes and growth, whereas each flexible genome allows bacteria to obtain supplementary biochemical functions as selective advantages. Consequently, this data suggests that *R. tritici* NCPPB 1953 have needed additional genetic elements and functions, which could provide effective physiological activity in accordance with specific environment and conditions.

**Functional analyses of flexible genome in *R. tritici* NCPPB 1953 compared to *R. toxicus* strains.** To investigate the biological meaning of distinctive genetic elements in *R. tritici* NCPPB 1953, we performed the pan-genome analysis of *R. tritici* NCPPB 1953 compared to *R. toxicus* WAC3373 and *R. toxicus* 70137, leading to identification of 1,052 *R. tritici* NCPPB 1953-specific genes (Supplementary Fig. 1). Furthermore, 1,052 genes

were identified through the BlastKOALA algorithm to infer high-level functions of gene clusters (Kanehisa et al., 2016). As a result, the highest enriched system was ‘Environmental Information Processing’ (72 genes) related to sensing signals and responses to diverse stimuli (Fig. 5). Also, many metabolic processes for substrates were observed in enriched systems as follows: ‘Carbohydrate metabolism’, 28 genes; ‘Amino acid metabolism’, 21 genes;

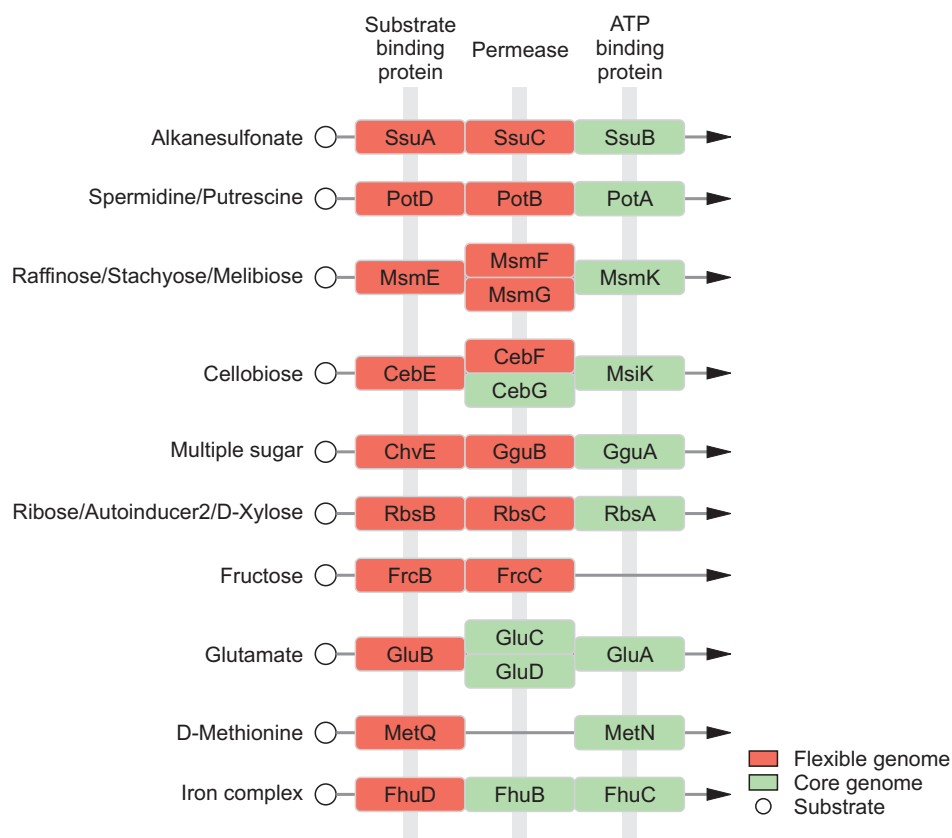
**Table 3.** Modules related to ‘Environmental Information Processing’ in *Rathayibacter tritici* NCPPB 1953-specific genes

KEGG ID	Module	Gene counts
rtn_M00196	Multiple sugar transport system	8
rtn_M00206	Cellobiose transport system	2
rtn_M00207	Putative multiple sugar transport system	1
rtn_M00209	Osmoprotectant transport system	1
rtn_M00212	Ribose transport system	2
rtn_M00216	Multiple sugar transport system	4
rtn_M00218	Fructose transport system	2
rtn_M00221	Putative simple sugar transport system	3
rtn_M00233	Glutamate transport system	1
rtn_M00236	Putative polar amino acid transport system	1
rtn_M00238	D-Methionine transport system	1
rtn_M00240	Iron complex transport system	1
rtn_M00243	Manganese/iron transport system	1
rtn_M00252	Lipooligosaccharide transport system	2
rtn_M00254	ABC-2 type transport system	6
rtn_M00299	Spermidine/putrescine transport system	2
rtn_M00436	Sulfonate transport system	2
rtn_M00813	Lantibiotic transport system	2



**Fig. 5.** Summary of biological systems in *Rathayibacter tritici* NCPPB 1953-specific genes. A distribution of the KEGG systems is displayed as a pie-chart. The number of gene counts is shown inside each pie of categories. Total 11 KEGG terms are displayed.





**Fig. 6.** The network of *Rathayibacter tritici* NCPPB 1953 in KEGG transporter. Pathway maps of KEGG transporter system were constructed by the Cytoscape tool. The KEGG database was used to download the data source. There are two color types in the rectangle node: red, genes belonging to flexible genome of *R. tritici* NCPPB 1953; green, genes belonging to the core genome among *R. tritici* NCPPB 1953, *R. toxicus* WAC3373, and *R. toxicus* 70137.

‘Nucleotide metabolism’, 15 genes; ‘Metabolism of co-factors and vitamins’, 14 genes; ‘Energy metabolism’, 12 genes (Fig. 5). These results indicate that flexible genome in *R. tritici* NCPPB 1953 have been mainly developed to sense and process diverse substrates from the environment.

In general, the environment with host condition is far away from the optimal growth conditions for bacteria than experimental conditions with rich media due to multiple stresses (Shimizu, 2013; Yadeta and Thomma, 2013). The acquisition of nutrients of plant pathogens in hostile environment is a fundamental challenge against an extremely low level of organic and inorganic compounds (Yadeta and Thomma, 2013). Pathogens have to need the ability to sense, process, and respond to a variety of substrates and stresses for survival (Shimizu, 2013). Until now, diverse mechanisms for sensing and processing exogenous metabolites within host have been intensively studied in many pathogens (Divon et al., 2006; Jones and Wildermuth, 2011; Wooldridge and Williams, 1993). Our module enrichments showed that most genes in the ‘Environmental Information Processing’ system are associated with recognition and transport of many substrates, including saccharides, lipids, phosphates, amino acids, metallic cations, minerals, organic ions, and vitamins (Table 3).

Flexible genome of *R. tritici* NCPPB 1953 was particularly distributed in substrate-binding proteins of transport system (Fig. 6), which is a key determinant of the extracellular signals and high affinity of uptake systems (Maqbool et al., 2015). In this regard, there are interesting reports on the flexible growth and colonization of *R. tritici* than *R. toxicus* (Bradbury, 1973; Riley and Ophel, 1992). In 1992, Riley and Ophel reported that *R. tritici* grows more rapidly than *R. toxicus* in culture (Riley and Ophel, 1992). It was reported that *R. tritici* successfully colonizes in plant hosts such as the gummosis of seed heads rather than *R. toxicus* (Bradbury, 1973). Therefore, we suggest that distinctive genetic features in flexible genome help *R. tritici* NCPPB 1953 to establish the stable growth and colonization in diverse environments through effective recognition of substrates. Further studies could lead to a better understanding of interactions between *R. tritici* and nematode as well as *R. tritici* and plant hosts.

### Acknowledgments

This research was supported by the Research of Animal and Plant Quarantine Agency, South Korea and with the support of “Cooperative Research Program for Agriculture Science & Technology Development (PJ01110502)”.

Rural Development Administration, South Korea.

## References

- Agarkova, I. V., Vidaver, A. K., Postnikova, E. N., Riley, I. T. and Schaad, N. W. 2006. Genetic characterization and diversity of *Rathayibacter toxicus*. *Phytopathology* 96:1270-1277.
- Akhtar, M. A. 1987. Outbreaks and new records. Pakistan. Bacterial gumming disease of wheat. *FAO Plant Prot. Bull.* 35: 102.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- Alva, V., Nam, S. Z., Söding, J. and Lupas, A. N. 2016. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res.* 44:W410-W415.
- Anderson, I. and Brass, A. 1998. Searching DNA databases for similarities to DNA sequences: when is a match significant? *Bioinformatics* 14:349-356.
- Arif, M., Busot, G. Y., Mann, R., Rodoni, B., Liu, S. and Stack, J. P. 2016. Emergence of a new population of *Rathayibacter toxicus*: an ecologically complex, geographically isolated bacterium. *PLoS One* 11:e0156182.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A. and Wheeler, D. L. 2000. GenBank. *Nucleic Acids Res.* 28:15-18.
- Bradbury, J. F. 1973. *Corynebacterium tritici*. In: *C.M.I. descriptions of pathogenic fungi and bacteria*, eds. by Commonwealth Mycological Institute and C.A.B. International Mycological Institute, pp. 371-380. Commonwealth Mycological Institute, Kew, UK.
- Bradbury, J. F. 1986. Guide to plant pathogenic bacteria. CAB International Mycological Institute, Slough, UK. 332 pp.
- Buiate, E. A., Xavier, K. V., Moore, N., Torres, M. F., Farman, M. L., Schardl, C. L. and Vaillancourt, L. J. 2017. A comparative genomic analysis of putative pathogenicity genes in the host-specific sibling species *Colletotrichum graminicola* and *Colletotrichum sublineola*. *BMC Genomics* 18:67.
- Carlson, R. R. and Vidaver, A. K. 1982. Taxonomy of *Corynebacterium* plant pathogens, including a new pathogen of wheat, based on polyacrylamide gel electrophoresis of cellular proteins. *Int. J. Syst. Bacteriol.* 32:315-326.
- Chaisson, M. J. and Tesler, G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* 13:238.
- Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W. and Korlach, J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10:563-569.
- 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.
- Davis, M. J., Gillaspie, A. G., Vidaver, A. K. and Harris, R. W. 1984. *Clavibacter*: a new genus containing some phytopathogenic coryneform bacteria, including *Clavibacter xyli* subsp. *xyli* sp. nov., subsp. nov. and *Clavibacter xyli* subsp. *cynodontis* subsp. nov., pathogens that cause ratoon stunting disease of sugarcane and bermudagrass stunting disease. *Int. J. Syst. Bacteriol.* 34:107-117.
- De Bruyne, E., Swings, J. and Kersters, K. 1992. Enzymatic relatedness amongst phytopathogenic coryneform bacteria and its potential use for their identification. *Syst. Appl. Microbiol.* 15:393-401.
- Delcher, A. L., Harmon, D., Kasif, S., White, O. and Salzberg, S. L. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27:4636-4641.
- Divon, H. H., Ziv, C., Davydov, O., Yarden, O. and Fluhr, R. 2006. The global nitrogen regulator, FNR1, regulates fungal nutrition-genes and fitness during *Fusarium oxysporum* pathogenesis. *Mol. Plant Pathol.* 7:485-497.
- Dowson, W. J. 1942. On the generic name of the gram-positive bacterial plant pathogens. *Trans. Br. Mycol. Soc.* 25:311-314.
- Duveiller, E. and Fucikovsky, L. 1997. Other plant pathogenic bacteria reported on wheat. In: *The bacterial diseases of wheat: concepts and methods of disease management*, eds. by E. Duveiller, L. Fucikovsky and K. Rudolph, p. 66. CIMMYT, Distrito Federal, Brazil.
- Evtushenko, L. I. and Dorofeeva, L. V. 2012. Genus XXII. *Rathayibacter*. In: *Bergey's manual of systematic bacteriology*, eds. by W. Whitman, M. Goodfellow, P. Kämpfer, H. J. Busse, M. Trujillo, W. Ludwig, K. Suzuki and A. Parte, pp. 953-964. Springer, New York, NY, USA.
- Fattah, F. A. 1988. Effects of inoculation methods on the incidence of ear-cockle and 'tundu' on wheat under field conditions. *Plant Soil* 109:195-198.
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J. and Bateman, A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44:D279-D285.
- Galperin, M. Y., Makarova, K. S., Wolf, Y. I. and Koonin, E. V. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 43:D261-D269.
- Grissa, I., Vergnaud, G. and Pourcel, C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35:W52-W57.
- Gupta, A. and Chattoo, B. B. 2008. Functional analysis of a novel ABC transporter ABC4 from *Magnaporthe grisea*. *FEMS Microbiol. Lett.* 278:22-28.
- 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 Bioinformatics* 11:119.
- Jago, M. V. and Culvenor, C. C. 1987. Tunicamycin and corynetoxin poisoning in sheep. *Aust. Vet. J.* 64:232-235.
- Jones, A. M. and Wildermuth, M. C. 2011. The phytopathogen *Pseudomonas syringae* pv. tomato DC3000 has three high-affinity iron-scavenging systems functional under iron limitation conditions but dispensable for pathogenesis. *J. Bacteriol.* 193:2767-2775.

- Kanehisa, M. and Goto, S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28:27-30.
- Kanehisa, M., Sato, Y. and Morishima, K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428:726-731.
- Kopf, M., Klähn, S., Voss, B., Stüber, K., Huettel, B., Reinhardt, R. and Hess, W. R. 2014. Finished genome sequence of the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6714. *Genome Announc.* 2:e00757-14.
- Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305:567-580.
- Kumar, S., Stecher, G. and Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33:1870-1874.
- Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T. and Ussery, D. W. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100-3108.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948.
- Lee, I. M., Bartoszyk, I. M., Gundersen-Rindal, D. E. and Davis, R. E. 1997. Phylogeny and classification of bacteria in the genera *Clavibacter* and *Rathayibacter* on the basis of 16S rRNA gene sequence analyses. *Appl. Environ. Microbiol.* 63:2631-2636.
- Lowe, T. M. and Eddy, S. R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955-964.
- Lukashin, A. V. and Borodovsky, M. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* 26:1107-1115.
- Maqbool, A., Horler, R. S., Muller, A., Wilkinson, A. J., Wilson, K. S. and Thomas, G. H. 2015. The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. *Biochem. Soc. Trans.* 43:1011-1017.
- Medini, D., Donati, C., Tettelin, H., Masignani, V. and Rappuoli, R. 2005. The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15:589-594.
- Mehta, Y. R. 2014. Spike diseases caused by bacteria. In: *Wheat diseases and their management*, ed. by Y. R. Mehta, pp. 105-121. Springer, New York, NY, USA.
- Moreira, L. M., Facincani, A. P., Ferreira, C. B., Ferreira, R. M., Ferro, M. I., Gozzo, F. C., de Oliveira, J. C., Ferro, J. A. and Soares, M. R. 2015. Chemotactic signal transduction and phosphate metabolism as adaptive strategies during citrus canker induction by *Xanthomonas citri*. *Funct. Integr. Genomics* 15:197-210.
- Paruthi, I. J. and Bhatti, D. S. 1985. Estimation of loss in yield and incidence of *Anguina tritici* on wheat in Haryana (India). *Int. Nematol. Netw. Newsl.* 2:13-16.
- Paruthi, I. J. and Gupta, D. C. 1987. Incidence of 'Tundu' in barley and kanki in wheat field infested with *Anguina tritici*. *Haryana Agric. Univ. J. Res.* 17:78-79.
- Petersen, T. N., Brunak, S., von Heijne, G. and Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8:785-786.
- Price, P. C., Fisher, J. M. and Kerr, A. 1979. On *Anguina funesta* n. sp. and its association with *Corynebacterium* sp., in infecting *Lolium rigidum*. *Nematologica* 25:76-85.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F. O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41:D590-D596.
- Rainey, F., Weiss, H., Prauser, H. and Stackebrandt, E. 1994. Further evidence for the phylogenetic coherence of actinomycetes with Group B-peptidoglycan and evidence for the phylogenetic intermixing of the genera *Microbacterium* and *Aureobacterium* as determined by 16S rDNA analysis. *FEMS Microbiol. Lett.* 118:135-139.
- Riley, I. T. 1992. *Anguina tritici* is a potential vector of *Clavibacter toxicus*. *Australas. Plant Pathol.* 21:147-149.
- Riley, I. T. and Ophel, K. M. 1992. *Clavibacter toxicus* sp. nov., the bacterium responsible for annual ryegrass toxicity in Australia. *Int. J. Syst. Bacteriol.* 42:64-68.
- Riley, I. T. and Reardon, T. B. 1995. Isolation and characterization of *Clavibacter tritici* associated with *Anguina tritici* in wheat from Western Australia. *Plant Pathol.* 44:805-810.
- Riley, I. T., Reardon, T. B. and McKay, A. C. 1988. Genetic analysis of plant pathogenic bacteria in the genus *Clavibacter* using allozyme electrophoresis. *J. Gen. Microbiol.* 134:3025-3030.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. and Barrell, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944-945.
- Sasaki, J., Chijimatsu, M. and Suzuki, K. 1998. Taxonomic significance of 2,4-diaminobutyric acid isomers in the cell wall peptidoglycan of actinomycetes and reclassification of *Clavibacter toxicus* as *Rathayibacter toxicus* comb. nov. *Int. J. Syst. Bacteriol.* 48:403-410.
- Shimizu, K. 2013. Regulation systems of bacteria such as *Escherichia coli* in response to nutrient limitation and environmental stresses. *Metabolites* 4:1-35.
- Sternes, P. R. and Borneman, A. R. 2016. Consensus pan-genome assembly of the specialised wine bacterium *Oenococcus oeni*. *BMC Genomics* 17:308.
- Takeuchi, M. and Yokota, A. 1994. Phylogenetic analysis of the genus *Microbacterium* based on 16S rRNA gene sequences. *FEMS Microbiol. Lett.* 124:11-16.
- UniProt Consortium. 2013. Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.* 41:D43-D47.
- Urban, M., Pant, R., Raghunath, A., Irvine, A. G., Pedro, H. and Hammond-Kosack, K. E. 2015. The Pathogen-Host Interactions database (PHI-base): additions and future developments. *Nucleic Acids Res.* 43:D645-D655.
- Wilson, R. A. and Talbot, N. J. 2009. Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*.

- zae*. *Nat. Rev. Microbiol.* 7:185-195.
- Wooldridge, K. G. and Williams, P. H. 1993. Iron uptake mechanisms of pathogenic bacteria. *FEMS Microbiol. Rev.* 12: 325-348.
- Yadeta, K. A. and Thomma, B. P. H. J. 2013. The xylem as battleground for plant hosts and vascular wilt pathogens. *Front. Plant Sci.* 4:97.
- Zgurskaya, H. I., Evtushenko, L. I., Akimov, V. N. and Kalakoutskii, L. V. 1993. *Rathayibacter* gen. nov., including the species *Rathayibacter rathayi* comb. nov., *Rathayibacter tritici* comb. nov., *Rathayibacter iranicus* comb. nov., and six strains from annual grasses. *Int. J. Syst. Bacteriol.* 43:143-149.