Xanthomonas sontii sp. nov., a non-pathogenic bacterium isolated from healthy basmati rice (*Oryza sativa*) seeds from India.

Kanika Bansal[1, ¶], Amandeep Kaur[1, ¶], Samriti Midha[1, §], Sanjeet Kumar[1], Suresh Korpole[1] and Prabhu B. Patil[1, *]

[1]Bacterial Genomics and Evolution Laboratory, CSIR-Institute of Microbial Technology, Chandigarh, India.

[§]Present address: Institute of Infection and Global Health, University of Liverpool, Liverpool, UK.

[¶] **Equal contribution**

[*] Corresponding author

Running title: Non-pathogenic *Xanthomonas* from rice plants.

Data submission: Whole-genome sequences of PPL1[T], PPL2 and PPL3 isolates are submitted to NCBI with accession numbers NQYO00000000, NQYP00000000, NMPO00000000 respectively.

Address correspondence to Prabhu B. Patil, pbpatil@imtech.res.in

Abbreviations: OrthoANI, Orthologous average nucleotide identity; dDDH, digital DNA-DNA hybridization; NA, nutrient agar; PSA, peptone sucrose agar; GYCA, glucose yeast extract calcium carbonate agar; PBS, phosphate buffer saline; TSBA, tryptic soy broth agar; MCS, MiSeq control software.

Abstract

We report three yellow-pigmented, Gram-negative, aerobic, rod-shaped, motile bacterial isolates designated as PPL1^T, PPL2, and PPL3 from healthy basmati rice seeds. Phenotypic and 16S rRNA gene sequence analysis assigned these isolates to the genus *Xanthomonas*. The 16S rRNA showed a 99.59% similarity with *X. sacchari* CFBP 4641^T, a sugarcane pathogen. Further, biochemical and fatty acid analysis revealed it to be closer to *X. sacchari*. Still, it differed from other species in general and known rice associated species such as *X. oryzae* (pathogenic) and *X. maliensis* (non-pathogenic) in particular. Interestingly, the isolatess in this study were isolated from healthy rice plants but are closely related to species that is pathogenic and isolated from diseased sugarcane. Accordingly, *in planta* studies revealed that PPL1^T, PPL2, and PPL3 are non-pathogenic to rice plants upon leaf inoculation. Taxonogenomic studies based on average nucleotide identity (orthoANI) and digital DNA-DNA hybridization (dDDH) values with type strains of *Xanthomonas* species were below the recommended threshold values for species delineation. Whole genome-based phylogenomic analysis revealed that these isolates formed a distinct monophyletic clade with *X. sacchari* CFBP 4641^T as their closest neighbour. Further, pangenome analysis revealed PPL1^T, PPL2, and PPL3 isolatesto comprise NRPS cluster along with a large number of unique genes associated with the novel species. Based on polyphasic and genomic approaches, a novel lineage and species associated with healthy rice seeds for which the name *Xanthomonas sontii* sp. nov. is proposed. The type strain for the *X. sontii* sp. nov. is PPL1^T (JCM 33631^T = CFBP 8688^T = ICMP 23426^T = MTCC 12491^T) and PPL2 (JCM 33632 = CFBP 8689 = ICMP 23427 = MTCC 12492) and PPL3 (JCM 33633 = CFBP 8690 = ICMP 23428 = MTCC 12493) as other strains of the species.

Keywords: *Xanthomonas sontii*, non-pathogenic, healthy basmati rice seeds, orthoANI, dDDH

Introduction

Genus *Xanthomonas* is a large group of phytopathogenic bacteria belonging to a complex order *Lysobacterales* (Sanjeet Kumar, Bansal, Patil, & Patil, 2019). *Xanthomonas* infects a diverse array of plants ranging from 124 monocots and 268 dicots (Hayward 1993). It is currently listed to contain 32 validly named species (http://www.bacterio.net/) (Bull et al., 2014; Bull et al., 2012; Parte, 2018). High phenotypic uniformity and similarity of 16S rRNA gene sequence of diverse phytopathogenic *Xanthomonas* strains has made its classification difficult. Before the advent of molecular tools, the genus *Xanthomonas* was classified according to the "new-host-new species" concept resulting in a complex genus comprising more than 100 species (Burkholder & Starr, 1948; Luc Vauterin, Rademaker, & Swings, 2000). In the 1990s, a major reclassification was accomplished based on traditional gold-standard for taxonomy i.e., DNA-DNA hybridization resulting in 20 different species (L Vauterin, Hoste, Kersters, & Swings, 1995). The advent of next-generation sequencing technology has revolutionized the field by providing us with robust parameters for taxonomic and phylogenetic classifications. Genomotaxonomy is based on digital DNA-DNA hybridization (dDDH) and average nucleotide identity (ANI) with 70% and 96% as cut-off for species delineation, respectively (Auch, von Jan, Klenk, & Göker, 2010; Lee, Kim, Park, & Chun, 2016).

X. maliensis, isolated from healthy rice plants is the only known species of *Xanthomonas* genus following a non-pathogenic lifestyle (Triplett et al., 2015). Another species of *Xanthomonas* associated with rice is *X. oryzae*, a devastating pathogen causing rice leaf blight disease worldwide (Midha et al., 2017; NIÑO- LIU, Ronald, & Bogdanove, 2006). Rice is a staple food for more than half of the world's population. Its management is of the highest concern. There is a need for continued investigation of the rice-associated *Xanthomonas* community from both diseased and healthy tissues. As non-pathogenic species do not cause economic damage and are isolated from healthy samples, these are widely overlooked. However, they

provide important insights into the lifestyle, adaptation, and the virulence mechanisms of their pathogenic counterparts. In our previous studies, the healthy rice seed is already reported to have a diversity of associated and endophytic bacteria (Chaudhry, Sharma, Bansal, & Patil, 2017; Midha et al., 2016). In the present study, we report genome sequence and analysis of three yellow-colored novel *Xanthomonas* isolates associated with healthy basmati rice seeds from agricultural fields of the northern part of India. The isolates were not able to cause disease upon leaf inoculation of healthy plants. Taxonogenomics based polyphasic studies revealed that these rice host-associated isolates formed a distinct lineage within *Xanthomonas* genus and belonged to a novel species. The closest related species are reported to be pathogenic to sugarcane suggesting a distinct evolutionary route of this novel non-pathogenic species. Accordingly, comparative genomics revealed a distinct set of unique genes comprising NRPS and other diversified genes. At the same time inter-strain genomic variations suggest ongoing selection within the members of this species and need for further systematic cellular, genetic and molecular studies of the novel species.

Materials and methods

Bacterial strain isolation and culture conditions

All three isolates were cultured from healthy rice seeds (Pusa basmati 1121 variety) collected from a farmers rice field located at Fazilka, Punjab, India (30.4036° N, 74.0280° E) in 2012 and 2013. Three isolates, PPL1^{T}, PPL2, and PPL3 were isolated from three independent experiments. For bacterial strain isolation, seeds were washed with sterile water followed by 70% ethanol wash and washed with sterile water. After washing, seeds were partially crushed in 0.85 % NaCl (normal saline) using sterile mortar and pestle. The mixture was then suspended in 50 ml of saline solution (Cottyn et al., 2001). The solution was incubated for 2 h at 28°C and serial dilutions were performed up to 10^{-6} and different dilutions (100µl) were plated on media like nutrient agar (NA), peptone sucrose agar (PSA), and glucose yeast extract calcium

carbonate agar (GYCA). Plates were incubated at 28°C up to 6 days. Bacterial colonies were isolated from GYCA plates and then were further maintained on PSA plates.

16S rRNA sequencing and phylogenetic tree construction

The 16S rRNA gene was amplified by PCR using universal primers: 27F (5'AGAGTTTGATCCTGGCTCAG 3') and 1492R (5' GGTTACCTTGTTACGACTT 3'). The Amplified product was purified using ExoSAP-IT™ PCR Product Cleanup Reagent (Thermo Fisher Scientific). Then, purified 16S rRNA gene was sequenced by the Sanger sequencing method on ABI 3130xl Genetic Analyzer (Applied Biosystems) using 1492R primer for sequencing PCR. Further, the partial sequence of 16S rRNA sequenced gene was feed in EzTaxon server (a database of type strains of prokaryotes) (https://www.ezbiocloud.net/) for identification purpose. As we performed whole genome sequencing of PPL1^{T}, PPL2, and PPL3, we used complete 16s rRNA gene sequences with contig number PPLI (NQYO01000058, locus_tag CJ027_08295), PPL2 (NQYP01000168, locus_tag CEK63_20225), and PPL3 (NMPO01000144, locus_tag CEK64_18465) for phylogenetic analysis. The 16S rRNA gene sequence of other strains was fetched from NCBI using IDs given in LPSN (Parte, 2018). Then 16S rRNA gene sequence alignment was performed using Clustal W (Thompson, Higgins, & Gibson, 1994) and tree was constructed using MEGA7 (Sudhir Kumar, Stecher, & Tamura, 2016).

Morphological assay

For TEM (transmission electron microscopy), PPL1^{T} strain was grown in nutrient broth and incubated at 28°C for 20 h. Subsequently, cells were harvested by centrifugation at 2000 rpm for 10 minutes. Cell pellet was washed twice with phosphate buffer saline (1X PBS) and finally resuspended in PBS. The bacterial suspension was placed on a carbon-coated copper grid (300 mesh, Nisshin EM Co., Ltd.) for 15 minutes. The grid was then negatively stained for 30

seconds with 2% phosphotungstic acid, dried and examined under JEM 2100 transmission electron microscope (JEOL, Tokyo, Japan) operating at 200 kV.

Biochemical characterization

Biochemical characterization such as carbohydrate utilization, acid production, and various enzymatic activities of PPL1^{T}, PPL2 and PPL3 isolates were performed using BIOLOG GEN III MICROPLATE^{TM} according to manufacturer's instructions. Isolates were cultured for 24 h on NA plates at 28°C. Bacterial cells were then resuspended in Suspension Buffer IF-A to the recommended turbidity range. Then, 100µl of suspension was transferred to each well of a BIOLOG GEN III microplate. Plates were incubated at 28°C and readings were taken after 24 h using MicroStation 2 Reader, and results were interpreted using MicroLog 3 software version 5.2.01.

For fatty acid analysis, isolates were grown on tryptic soy broth agar (TSBA) medium for 48 hrs at 28℃. Total fatty acids of cells were separated from a loopful of culture as methyl esters using the method described (Buyer, 2002). The analysis was performed using the Sherlock Microbial Identification System (MIDI version 6.1, database RTSBA 6.0) as described previously [MIS operating manual version 6.1].

DNA extraction, genome sequencing, assembly and annotation

Genomic DNA extraction of PPL1^{T}, PPL2 and PPL3 isolates was carried out using ZR Fungal/Bacterial DNA MiniPrep kit (Zymo Research, Irvine, CA, USA). Qualitative assessment of DNA was performed using NanoDrop 1000 (Thermo Fisher Scientific, Wilmington, DE, USA) and agarose gel electrophoresis. DNA was quantified using Qubit 2.0 fluorometer (Life Technologies). Nextera XT sample preparation kits (Illumina, Inc., San Diego, CA, USA) were used to prepare Illumina paired-end sequencing libraries (250 x 2 read length) with dual indexing adapters. In-house sequencing of the Illumina libraries was carried out on Illumina MiSeq platform (Illumina, Inc., San Diego, CA, USA). Adapter trimming was

performed automatically by MiSeq control software (MCS), and remaining adapters were detected by NCBI server and removed by manual trimming. Sequencing reads were *de novo* assembled into high-quality draft genome on CLC Genomics Workbench v7.5 (CLC bio, Aarhus, Denmark) using default settings. Genome annotation was performed by NCBI PGAP pipeline (http://www.ncbi.nlm.nih.gov/genome/annotation_prok).

In planta pathogenicity test

Pathogenicity tests were performed on 30-day-old greenhouse grown rice plants of PUSA-basmati 1121 variety. Briefly, PPL1^{T}, PPL2, PPL3 and *X. oryzae* strain BXO1 were grown in PS (1% peptone and 1% sucrose) media at 28°C at 200 rpm until OD at 600nm reaches 1.0. Cell pellets were then washed with PBS. Finally bacterial suspension of PPL1^{T}, PPL2, PPL3 and BXO1 were inoculated on rice plants (PUSA-basmati 1121) leaf clip method by dipping scissors in bacterial culture and clipping tips of rice leaves. After 14 days, pathogenicity was assessed by measuring the length of lesions on leaves. Here, BXO1 was positive control and plant leaves inoculated with only PBS were negative controls. Lesion length was measured from 10 inoculated leaves from two independent experiments.

Phylogenomic and taxonogenomic analysis

The core genome tree was constructed using PhyML v3.0 (Guindon et al., 2010). Briefly, core genome alignment was obtained using Roary v3.11.2 (Page et al., 2015) with an identity cut-off of 80%. The core gene alignment was converted into phylip format using SeaView v4.4.2-1 (Gouy, Guindon, & Gascuel, 2010) and then, newick tree was obtained using PhyML. *Stenotrophomonas maltophilia* ATCC13637^{T} was used as an outgroup. Taxonogenomic analysis of all type strains or representative strains of *Xanthomonas* (type strains for which genome sequence was not available we took another strain with complete genome available at NCBI including *X. theicola* CFBP 4691, *X. hyacinthi* CFBP 1156, *X. pisi* DSM 18256, *X. phaseoli* CFBP 412, *X. hortorum* MO 81, *X. perforans* 91-118, *X. euvesicatoria* LMG 27970,

X. gardneri ICMP 7383, *X. melonis* CFBP 4644, *X. populi* CFBP 1817, *X. fragariae* PD 885, and *X. codiaei* CFBP 4690) was performed using OrthoANI v1.2 (Lee et al., 2016) values calculated by using USEARCH v5.2.32 (Edgar, 2010) and dDDH were calculated using Web tool GGDC 2.0 (http://ggdc.dsmz.de/distcalc2.php)

Pangenome analysis

Pangenome analysis was performed using Roary v3.11.2 (Page et al., 2015). Briefly, gff files were generated using PROKKA (Seemann, 2014). Then, gff files were used as input for roary v3.11.2 with an identity cutoff of 90%. Unique genes obtained from pangenome analysis were further classified based on COG (cluster of orthologous groups) using EggNOG (Huerta-Cepas et al., 2017; Huerta-Cepas et al., 2016). **Results and Discussion**

Isolation, phenotypic and molecular identification of novel *Xanthomonas* isolates from healthy rice seeds

Three bacterial isolates designated as PPL1^{T}, PPL2, and PPL3 were isolated from healthy Pusa Basmati rice seeds. Morphologically colonies appeared as creamish-yellow, round, smooth, convex, and circular. All cells stained Gram-negative and as observed under transmission electron microscope, they were rod-shaped bacteria with monopolar flagella (Fig. S1). All three isolates grew well between 20°C to 37°C with an optimum temperature at 28°C (±2°C) and no growth was observed at 50°C. Isolates were able to grow at pH 6.0, whereas no growth observed at pH 5.0. Further, 16S rRNA for all the three isolates was sequenced for species-level identification and phylogenetic analysis. All three isolates had 100% 16S rRNA gene sequence identity. The sequence identities between PPL1^{T} and *X. sacchari* CFBP 4641^{T} (a sugarcane pathogen) (L Vauterin et al., 1995) and *X. maliensis* LMG 27592^{T} (a non-pathogenic *Xanthomonas* species isolated from healthy rice) (Triplett et al., 2015) were 99.59 % and 98.15% respectively based on 16S rRNA gene sequences. . 16S rRNA-based phylogenetic analysis for all the representative members of *Xanthomonas* genus is shown in (Fig. 1). 16S

rRNA gene based phylogeny revealed that PPL1^T, PPL2 and PPL3 isolates are closely related to *X. sacchari* CFBP 4641^T compared to other rice associated species i.e., *X. oryzae* or *X. maliensis* (Swings et al., 1990; Triplett et al., 2015) or other validly named species. However, unlike PPL1^T, PPL2, and PPL3 that were isolated from healthy rice seeds, ,. *X. sacchari* was isolated from infected tissues of sugarcane (L Vauterin et al., 1995).

Biochemical and fatty acid composition

Biochemical tests such as carbohydrate utilization, acid production and test for various enzymatic activities of all PPL1^T, PPL2, and PPL3 isolates performed using OMNILOG GEN III system (BIOLOG) are shown in Table 1. Isolates are able to utilize D-maltose, D-trehalose, D-cellobiose, gentiobiose, sucrose, α-D-lactose, α-D-glucose, D-mannose, D-fructose, D-galactose, L-alanine, L-aspartic acid, L-glutamic acid, L-lactic acid and citric acid. All PPL1^T, PPL2, and PPL3 isolates were similar in their colony morphology and displayed similar biochemical characteristics, as shown in Table 1. Results revealed that PPL1^T, PPL2, PPL3 and *X. sacchari* NCPPB 4341^T isolates were distinguishable from *X. maliensis* LMG 27592^T, another non-pathogenic strain isolated from healthy rice leaves in some carbohydrate utilization characteristics such as L-rhamnose, D-mannitol, D-arabitol, L-arginine, L-aspartic acid and L-histidine (Triplett et al., 2015; L Vauterin et al., 1995). Further, rice associated non-pathogenic isolates i.e., all PPL1^T, PPL2, and PPL3 isolates and *X. maliensis* LMG 27592^T showed differences in utilization of α-D-lactose, D-galactose and assimilation of citric acid when compared with rice pathogenic *X. oryzae* LMG 5047^T strain as shown in Table 1.

Further, fatty acid profile of PPL1^T, PPL2 and PPL3 isolates was also investigated (Table S1). The major fatty acids iso-$C_{11:}$0, Iso $C_{11:0}$ 3-OH, Iso $C_{13:0}$ 3-OH that are characteristic features of *Xanthomonas* genus and useful for differentiating *Xanthomonas* strains from other bacteria were present in all PPL1^T, PPL2, and PPL3 isolates (L Vauterin et al., 1995). In fatty acid composition, major fatty acids among PPL1^T, PPL2, and PPL3 isolatesinclude summed feature

3 ($C_{16:1}$ w7c/$C_{16:1}$ w6c), $C_{16:0}$, iso-$C_{14:0}$, iso-$C_{15:0}$, iso-$C_{16:0}$, iso-$C_{17:0}$, and anteiso-$C_{15:0}$. PPL1^T, PPL2, and PPL3 isolates were found to be homogenous in their fatty acid composition but their global fatty acid profile was similar to *X. sacchari* CFBP 4641^T and not to already known rice associated species *X. oryzae* LMG 5047^T and *X. maliensis* LMG 27592^T (Luc Vauterin, Yang, & Swings, 1996). **PPL1^T, PPL2, and PPL3 isolates donot cause lesions on rice leaves upon leaf-clip inoculation**

As PPL1^T, PPL2, and PPL3 isolates were also isolated from healthy rice seeds we checked PPL1^T, PPL2, and PPL3 isolates pathogenic status by performing *in planta* studies by leaf clip inoculation. *X. oryzae* pv. *oryzae* strain BXO1 that causes bacterial blight disease on rice was taken as a positive control. Plant leaves inoculated with only PBS were taken as a negative control. Pathogenicity of the isolates was assessed by measuring disease symptoms/lesions length on the leaves after 14 days of infection. PPL1^T, PPL2, and PPL3 isolates were not able to cause any symptoms/lesions on rice plant leaves, whereas BXO1 inoculated leaves showed lesions (on an average 13 cm lesion in length) as shown by yellowing/wilting of rice leaves (Fig. 2). Pathogenicity assay clearly revealed that PPL1^T, PPL2, and PPL3 isolates are non-pathogenic to the host and not able to cause any lesions on leaves upon leaf clip inoculation.

Genomic features and taxonogenomics of novel *Xanthomonas* isolates

Whole-genome sequencing of PPL1^T, PPL2, and PPL3 isolates was carried out using the in-house Illumina MiSeq platform. The size of assembled genomes was approximately 5 Mb with genome coverage 78x, 109x, and 80x for PPL1^T, PPL2, and PPL3, respectively, and N50 ranging from 32 to 48 kb. Whole-genome sequences of PPL1^T, PPL2, and PPL3 isolates were submitted to NCBI with accession numbers NQYO00000000, NQYP00000000, NMPO00000000 respectively. Further, genomes were checked for the completeness and presence of contamination. Complete assembly statistics for all the isolates are given in Table 2.

To check whether the PPL1^T, PPL2, and PPL3 isolates belong to a novel species, we calculated orthoANI and dDDH values. The orthoANI and dDDH values (Table 3) of PPL1^T, PPL2, and PPL3 with type and representative strains of genus *Xanthomonas* species were below the cut-off for species delineation. These isolates have *X. sacchari* as their closest relative with ANI values of ~ 94% and around less than 79% with other representative strains. PPL1^T, PPL2, and PPL3 isolates showed dDDH values of around 55% with *X. sacchari* and less than 35% with other species of the genus *Xanthomonas*. **Genome based phylogeny investigation of the novel *Xanthomonas* isolates**

To assess genome-level phylogeny of PPL1^T, PPL2, and PPL3, we constructed a core genome phylogenetic tree using genome sequence of type strain of species reported in *Xanthomonas*. *Stenotrophomonas maltophilia* ATCC13637^T was used as an outgroup (Fig. 3). A total of 494 genes constitutes the core content of isolates and used for core genome tree construction. Interestingly, the analysis revealed 27 species formed one group as previously reported (Hauben, Vauterin, Swings, & Moore, 1997) including *X. pisi, X. vesicatoria, X. citri, X. codiaei, X. fragariae, X. bromi, X. campestris, X. dyei, X. phaseoli, X. hortorum, X. arboricola, X. cynarae, X. cucurbitae, X. vasicola, X. floridensis, X. perforans, X. euvesicatoria, X. maliensis, X. gardneri, X. axonopodis, X. cassavae, X. nasturtii, X. alfalfae, X. prunicola, X. oryzae, X. melonis,* and *X. populi.* Whereas, PPL1^T, PPL2, PPL3, *X. sacchari, X. theicola, X. translucens, X. hyacinthi, X. albilineans* formed a second group. Here, PPL1^T, PPL2 and PPL3 formed a monophyletic clade distinguishing them from other strains and *X. sacchari* is the closest neighbour of these strains.

Pangenome analysis reveals large number of unique genes of PPL1^T, PPL2, and PPL3

To look into unique gene content of PPL1^T, PPL2, and PPL3 isolates we performed pangenome analysis by taking species from the second group in consideration i.e., *X. sacchari* CFBP 4641^T, *X. albilineans* CFBP 2523^T, *X. translucens* DSM 18974^T, *X. hyacinthi* CFBP 1156^T, *X. theicola*

CFBP 4691^{T}. The total size of pangenome was 13440 with 900 core genes and 745 genes unique to PPL1^{T}, PPL2, and PPL3 isolates(Fig. 4a). Unique genes of PPL1^{T}, PPL2, and PPL3 are provided in the table S1. Further, unique genes identified from pangenome analysis were investigated for their GC content. Overall, 60% of the unique genes had atypical GC content (i.e., 69 ± 2.5%). Genes with atypical GC content point toward the possibility of their acquisition through horizontal gene transfer and diversification of PPL1^{T}, PPL2, and PPL3 isolates from their closest relatives. Unique genes belonging to PPL1^{T}, PPL2, and PPL3 isolates were further classified into different COG classes (Fig. 4b). Among these, 11% of the genes belong to "metabolism" class, 15% of genes to "cellular processes and signalling" class and 10% of the genes to "Information storage and processing" class. 64% of the genes were assigned to a poorly characterized class with unknown or hypothetical functions. Core biosynthetic genes of non-ribosomal peptide synthetase (NRPS) gene cluster were found to be unique to PPL1^{T}, PPL2, and PPL3 (Fig. 5). This was having 15% identity with *Lysobacter capsici* and is not having its homologs in *Xanthomonas* genus. The unique gene pool probably indicates towards diversification of PPL1^{T}, PPL2, and PPL3 isolates from other species of *Xanthomonas* genus.

Further, to assess the genomic differences of PPL1^{T}, PPL2, and PPL3 isolates, we carried pangenome analysis of PPL1^{T}, PPL2 and PPL3 isolates. Number of unique genes were 78, 67, and 475 for PPL1^{T}, PPL2, and PPL3, respectively. This analysis clearly depicted the divergence of PPL3 from PPL1^{T} and PPL2, which is in concordance with the core-genome phylogeny, dDDH and ANI values. Distinct phylogenomic status along with a large number of unique genes in the species reiterates the importance of *Xanthomonas* member associated with rice microbiome. Hence, further studies are required to understand its ongoing adaptation in healthy tissues and also to its host rice using in-depth cellular, molecular, genetic and functional approaches.

Description of *Xanthomonas sontii* sp. nov.

Xanthomonas sontii (son'ti.i. N.L. masc. gen. n. *sontii* named in honour of Ramesh V. Sonti, a renowned Indian bacterial and plant molecular geneticist).

Cells are Gram-negative, aerobic, rod-shaped, motile with monopolar flagella. Colonies appear creamish-yellow, round, smooth, convex and circular after 24 hrs of growth at 28°C on peptone sucrose agar (PSA). Growth was observed at pH 6.0 and 1% NaCl and sodium lactate concentrations. Cells metabolically active in the presence of D-maltose, D-trehalose, D-cellobiose, gentiobiose, sucrose, α-D-lactose, α-D-glucose, D-mannose, D-fructose, D-galactose, L-alanine, L-aspartic acid, L-glutamic acid, L-lactic acid and citric acid whereas inactive in L-rhamnose, L-arginine, L-histidine, D-mannitol, D-arabitol.. The fatty acids profile showed presence of distinct fatty acids iso-$C_{11:0}$, iso $C_{11:0}$ 3-OH, iso $C_{13:0}$ 3-OH that are characteristic features of *Xanthomonas* genus. Major fatty acids among PPL1^T, PPL2, and PPL3 isolates include summed feature 3 ($C_{16:1}$ w7c/$C_{16:1}$ w6c), $C_{16:0}$, iso-$C_{14:0}$, iso-$C_{15:0}$, iso-$C_{16:0}$, iso-$C_{17:0}$, anteiso-$C_{15:0}$ with relatively higher proportion of $C_{10:0}$ and $C_{18:0}$. Isolated from healthy rice seeds and do not cause any disease symptoms upon leaf inoculation. The type strain is PPL1^T (JCM 33631^T = CFBP 8688^T = ICMP 23426^T = MTCC 12491^T).

Authors Contributions

SM isolated isolates, KB and AK performed strain identification characterisation. KB and SK have performed genome sequencing. AK, KB and SK did genome analysis. AK and KB drafted manuscript with inputs from PBP and SK. PBP conceived the study and participated in its design. All the authors read and approved the manuscript.

Acknowledgements

This work was supported by a project entitled "Mega-genomic insights into co-evolution of rice and its microbiome (MICRA)" MLP0020 of Council of Scientific and Industrial Research (CSIR).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial support that could be considered as a potential conflict of interest.

References

Auch, A. F., von Jan, M., Klenk, H.-P., & Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in genomic sciences, 2*(1), 117-134.

Bull, C. T., Coutinho, T., Denny, T., Firrao, G., Fischer-Le Saux, M., Li, X., . . . Takikawa, Y. (2014). List of new names of plant pathogenic bacteria (2011-2012). *Journal of Plant Pathology, 96*(2), 223-226.

Bull, C. T., De Boer, S., Denny, T., Firrao, G., Fischer-Le Saux, M., Saddler, G., . . . Takikawa, Y. (2012). LIST OF NEW NAMES OF PLANT PATHOGENIC BACTERIA (2008-2010) Prepared by the International Society of Plant Pathology Committee on the Taxonomy of Plant Pathogenic Bacteria (ISPP-CTPPB). *Journal of Plant Pathology*, 21-27.

Burkholder, W. H., & Starr, M. P. (1948). The generic and specific characters of phytopathogenic species of Pseudomonas and Xanthomonas. *Phytopathology, 38*(6), 494-502.

Buyer, J. S. (2002). Identification of bacteria from single colonies by fatty acid analysis. *Journal of microbiological methods, 48*(2-3), 259-265.

Chaudhry, V., Sharma, S., Bansal, K., & Patil, P. B. (2017). Glimpse into the genomes of rice endophytic bacteria: diversity and distribution of firmicutes. *Frontiers in microbiology, 7*, 2115.

Cottyn, B., Regalado, E., Lanoot, B., De Cleene, M., Mew, T., & Swings, J. (2001). Bacterial populations associated with rice seed in the tropical environment. *Phytopathology, 91*(3), 282-292.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics, 26*(19), 2460-2461.

Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution, 27*(2), 221-224.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology, 59*(3), 307-321.

Hauben, L., Vauterin, L., Swings, J., & Moore, E. (1997). Comparison of 16S ribosomal DNA sequences of all Xanthomonas species. *International Journal of Systematic and Evolutionary Microbiology, 47*(2), 328-335.

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., & Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular biology and evolution, 34*(8), 2115-2122.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., . . . Kuhn, M. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic acids research, 44*(D1), D286-D293.

Kumar, S., Bansal, K., Patil, P. P., & Patil, P. B. (2019). Phylogenomics insights into order and families of Lysobacterales. *Access Microbiology, 1*(2), e000015.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution, 33*(7), 1870-1874.

Lee, I., Kim, Y. O., Park, S.-C., & Chun, J. (2016). OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *International Journal of Systematic and Evolutionary Microbiology, 66*(2), 1100-1103.

Midha, S., Bansal, K., Kumar, S., Girija, A. M., Mishra, D., Brahma, K., . . . Patil, P. B. (2017). Population genomic insights into variation and evolution of Xanthomonas oryzae pv. oryzae. *Scientific reports, 7*(1), 1-13.

Midha, S., Bansal, K., Sharma, S., Kumar, N., Patil, P. P., Chaudhry, V., & Patil, P. B. (2016). Genomic resource of rice seed associated bacteria. *Frontiers in microbiology, 6*, 1551.

NIÑO-LIU, D. O., Ronald, P. C., & Bogdanove, A. J. (2006). Xanthomonas oryzae pathovars: model pathogens of a model crop. *Molecular plant pathology, 7*(5), 303-324.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., . . . Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics, 31*(22), 3691-3693.

Parte, A. C. (2018). LPSN–List of Prokaryotic names with Standing in Nomenclature (bacterio. net), 20 years on. *International Journal of Systematic and Evolutionary Microbiology, 68*(6), 1825-1829.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics, 30*(14), 2068-2069.

Swings, J., Van den Mooter, M., Vauterin, L., Hoste, B., Gillis, M., Mew, T., & Kersters, K. (1990). Reclassification of the Causal Agents of Bacterial Blight (Xanthomonas campestris pv. oryzae) and Bacterial Leaf Streak (Xanthomonas campestris pv. oryzicola) of Rice as Pathovars of Xanthomonas oryzae (ex Ishiyama 1922) sp. nov., nom. rev. *International Journal of Systematic and Evolutionary Microbiology, 40*(3), 309-311.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research, 22*(22), 4673-4680.

Triplett, L. R., Verdier, V., Campillo, T., Van Malderghem, C., Cleenwerck, I., Maes, M., . . . Cottyn, B. (2015). Characterization of a novel clade of Xanthomonas isolated from rice leaves in Mali and proposal of Xanthomonas maliensis sp. nov. *Antonie van Leeuwenhoek, 107*(4), 869-881.

Vauterin, L., Hoste, B., Kersters, K., & Swings, J. (1995). Reclassification of xanthomonas. *International Journal of Systematic and Evolutionary Microbiology, 45*(3), 472-489.

Vauterin, L., Rademaker, J., & Swings, J. (2000). Synopsis on the taxonomy of the genus Xanthomonas. *Phytopathology, 90*(7), 677-682.

Vauterin, L., Yang, P., & Swings, J. (1996). Utilization of fatty acid methyl esters for the differentiation of new Xanthomonas species. *International Journal of Systematic and Evolutionary Microbiology, 46*(1), 298-304.

Figure Legends

Fig. 1 Neighbor-joining phylogenetic tree based on the 16S rRNA gene sequences showing phylogenetic relationship between PPL1^T, PPL2, and PPL3 isolates and other species of genus *Xanthomonas*. All PPL1^T, PPL2, and PPL3 isolates are highlighted in the green box. The tree was rooted using *S. maltophilia* ATCC 13637^T as an outgroup. Bootstrap values are given at the nodes. The scale bar indicates number of substitutions per site

Fig. 2 *In planta* infection of rice leaves: (a) Rice leaves were inoculated with *X. oryzae* pv. *oryzae* (BXO1), PPL1, PPL2, and PPL3. At 14 days post-inoculation, images of leaves were captured. (b) Quantitation of lesion length (in cm) after 14 days post-inoculation (dpi). Error bar indicates standard deviation of readings from 10 inoculated leaves and from two independent experiments.

Fig. 3 Whole genome based phylogeny taking all type strains and representative strains of genus *Xanthomonas*. The scale bar shows the number of nucleotide substitution per site. PPL1^{T}, PPL2 and PPL3 isolates (highlighted in coloured box) formed a distinct cluster. *S. maltophilia* ATCC13637^{T} was used as an outgroup. Bootstrap values are mentioned at the nodes

Fig. 4 (a) Venn diagram showing core genes and genes unique to PPL1^{T}, PPL2, and PPL3 and other strains. (b) Pie chart representing COG classification of unique genes identified from pangenome analysis

Fig. 5 Schematic representation of unique NRPS/cluster system in strain PPL1^{T}. Different genes are colored according to the predicted functions. The NRPS domains are shown above: A-adenylation domain; C-condensation domain; TE-thioesterase domain. Below is the heat map showing comparison of percentage gene identity between PPL1 and other strains from genus *Xanthomonas*

Table legends

Table 1 Comparison of biochemical characteristics of PPL1^{T}, PPL2, PPL3 and *X. sacchari* NCPPB 4341^{T} with type strains of rice associated species *X. oryzae* LMG 5047^{T}* and *X. maliensis* LMG 27592^{T}* (* data taken from literature) (Triplett et al., 2015; L Vauterin et al., 1995). Symbols represent '+' positive, '-' negative. **Table 2** Genome assembly statistics of PPL1^{T}, PPL2 and PPL3 isolates

Table 3 Digital DNA-DNA hybridization (dDDH) and OrthoANI pairwise comparison of PPL1^{T}, PPL2, and PPL3 isolates PPL1^{T}, PPL2 and PPL3 with the type or representative strains of other species of genus *Xanthomonas*

Supplementary figure S1: Transmission electron micrograph of strain PPL1^{T} with monopolar flagella. Bar 0.5µm

Supplementary Table S1 Comparative fatty acid profiles of PPL1^{T}, PPL2, and PPL3 isolates with other *Xanthomonas* strains. 1. PPL1^{T}; 2. PPL2; 3. PPL3; 4. *X. sacchari* LMG 471^{T}; 5. *X. theicola* LMG 8684^{T}; 6. *X. translucens* LMG 876^{T}; 7. *X. hyacinthi* LMG 739^{T}; 8. *X. albilineans* LMG 494^{T}; 9. *X. maliensis* LMG 27592^{T} 10. *X. oryzae* LMG 5047^{T}; 11. *X. arboricola* LMG 747^{T}; 12. *X. axonopodis* LMG 538^{T}*;* 13. *X. bromi* LMG 947^{T}*;* 14. *X. campestris* LMG 568^{T}; 15. *X. cassava* LMG 673^{T}; 16. *X.codiaei* LMG 8678^{T}; 17. *X. cucurbitae* LMG 690^{T}; 18. *X. faragariae* LMG 708^{T}*;* 19. *X. hortorum* LMG 733^{T}*;* 20. *X. melonis* LMG 8670^{T}*;* 21. *X. pisi* LMG 847^{T}*;* 22. *X .populi* LMG 5743^{T}*;* 23. *X. vasicola* LMG 736^{T}; 24. *X. vesicatoria* LMG 911^{T}. Symbols represent '-' data not available, 'ND' not detected. Summed feature 1 comprises 15:1 iso H/13:0 3OH, summed features 3 comprises 16:1 w7c/16:1 w6c, summed features 8 comprises 18:1 w7c and summed feature 9 comprises 16:0 10-methyl The data for PPL1, PPL2 and PPL3 isolates was generated in the present study whereas for the rest of the strains data was taken from literature (Luc Vauterin et al., 1996)

Supplementary table S2 List of genes unique to PPL1^{T}, PPL2, and PPL3 isolates

Tables

Table 1 Comparison of biochemical characteristics of PPL1^{T}, PPL2, PPL3 and *X. sacchari* NCPPB 4341^{T} with type strains of rice associated species *X. oryzae* LMG 5047^{T}* and *X. maliensis* LMG 27592^{T}* (* data taken from literature) (Triplett et al., 2015; L Vauterin et al., 1995). Symbols represent '+' positive, '-' negative.

	PPL1^{T}	**PPL2**	**PPL3**	***X. sacchari* NCPPB 4341^{T}**	***X.maliensis* 97 LMG 27592^{T}***	***X.oryzae* LMG 5047^{T}***
D-maltose	+	+	+	+	+	+
D-trehalose	+	+	+	+	+	+

D-cellobiose	+	+	+	+	+	+
Gentiobiose	+	+	+	+	+	+
Sucrose	+	+	+	+	+	+
D-raffinose	-	-	-	-	-	-
α-D-lactose	+	+	+	+	+	-
α-D-glucose	+	+	+	+	+	+
D-mannose	+	+	+	+	+	+
D-fructose	+	+	+	+	+	+
D-galactose	+	+	+	+	+	-
L-rhamnose	-	-	-	-	+	-
D-sorbitol	-	-	-	-	-	-
D-mannitol	-	-	-	-	+	-
D-arabitol	-	-	-	-	+	-
L-alanine	+	+	+	+	+	+
L-arginine	-	-	-	-	+	-
L-aspartic acid	+	+	+	+	-	-
L-glutamic acid	+	+	+	+	+	+
L-histidine	-	-	-	-	+	-
D-saccharic Acid	-	-	-	-	-	-
L-lactic acid	+	+	+	+	+	+
Citric acid	+	+	+	+	+	-

Table 2 Genome assembly statistics of PPL1^{T}, PPL2 and PPL3 isolates

S. No.	Strain Name	Isolation source	Genome Size	GC (%)	Fold (X)	Contigs	N50 (kb)	CDS	rRNA + tRNA	Completeness / Contamination	Accession No.
1	PPL1^{T}	Rice seeds	4.8	69	78	332	32.9	4149	3+51	96.31/0.22	NZ_NQYO00000000
2	PPL2	Rice seeds	4.9	68.8	109	200	48.1	4105	2+50	95.79/0.07	NZ_NQYP00000000
3	PPL3	Rice seeds	4.8	68.8	80	231	40.3	4106	3+52	98.56/0.55	NZ_NMPO00000000

498

499

500

501

502 **Table 3** Digital DNA-DNA hybridization (dDDH) and OrthoANI pairwise comparison of

503 PPL1^{T}, PPL2, and PPL3 isolates with the type or representative strains of other species of

504 genus *Xanthomonas*

505

506

Strain	PPL1		PPL2		PPL3	
	dDDH	oANI	dDDH	oANI	dDDH	oANI
X. theicola CFBP 4691	32.4	86.9	32.4	86.8	32.5	87.0
X. translucens DSM 18974^{T}	33.0	87.0	33.0	87.0	33.0	87.1
X. hyacinthi CFBP 1156	33.6	87.6	33.6	87.5	33.4	87.6
X. albilineans CFBP 2523^{T}	28.5	84.3	28.4	84.4	28.4	84.4
X . sacchari CFBP 4641^{T}	55.1	94.1	55.3	94.2	55.2	94.1
PPL3	78.2	97.6	78.7	97.6	100.0	100.0
PPL1^{T}	100.0	100.0	99.5	99.9	78.2	97.6
PPL2	99.5	99.9	100.0	100.0	78.7	97.6
X. maliensis LMG 27592^{T}	23.1	79.7	23.1	79.6	23.1	79.5
X. campestris ATCC 33913^{T}	23.0	79.4	23.0	79.4	23.1	79.3
X. cucurbitae CFBP 2542^{T}	23.2	79.8	23.2	79.5	23.0	79.7
X. cassavae CFBP 4642^{T}	23.4	79.7	23.4	79.6	23.3	79.7

X. floridensis WHRI 8848^{T}	23.4	79.6	23.4	79.7	23.4	79.6
X. codiaei CFBP 4690	23.6	79.9	23.6	80.1	23.7	80.1
X. melonis CFBP 4644	23.4	79.7	23.4	79.7	23.4	79.7
X. vesicatoria LMG 911^{T}	22.9	79.1	22.8	79.1	22.8	79.0
X. pisi DSM 18956	23.1	79.1	23.0	79.4	23.3	79.4
X. dyei CFBP 7245^{T}	22.9	79.2	22.9	79.1	23.0	79.2
X. fragariae PD 885	22.5	78.6	22.4	78.6	22.4	78.5
X. arboricola CFBP 2528^{T}	23.6	79.9	23.5	79.9	23.4	79.8
X. populi CFBP 1817^{T}	22.7	78.9	22.6	79.0	22.4	79.0
X. hortorum MO 81	22.9	79.1	22.8	79.2	22.8	79.3
X. cynarae CFBP 4188^{T}	23.0	79.3	23.0	79.3	22.9	79.3
X. gardneri ICMP 7383	23.1	79.2	23.0	79.3	22.9	79.2
X. nasturtii WHRI 8853^{T}	23.1	79.2	23.1	79.4	23.1	79.3
X. bromi CFBP 1976^{T}	23.0	79.0	23.0	79.1	23.0	79.1
X. prunicola CFBP 8353^{T}	22.9	78.9	22.9	79.0	22.9	78.8
X. axonopodis DSM 3585^{T}	23.1	78.8	23.1	79.1	23.0	79.0
X. citri LMG 9322^{T}	23.3	79.1	23.3	79.1	23.3	79.1
X. phaseoli CFBP 412^{T}	23.0	79.1	23.1	79.3	23.0	79.1
X. perforans 91-118	23.0	79.2	23.0	79.3	23.0	79.4
X. euvesicatoria LMG 27970 ^{T}	23.6	79.2	23.6	79.2	23.6	79.3
X. alfalfae LMG 495^{T}	23.4	79.3	23.4	79.2	23.3	79.3
X. vasicola NCPPB 2417^{T}	22.8	78.6	22.8	78.6	22.9	78.5
X. oryzae ATCC 35933^{T}	22.9	78.9	22.8	78.9	23.0	78.8
S. maltophilia ATCC 13637^{T}	22.9	79.1	22.9	79.2	22.9	79.0

507

Fig. 1

Click here to access/download;Figure;Fig. 1.tif

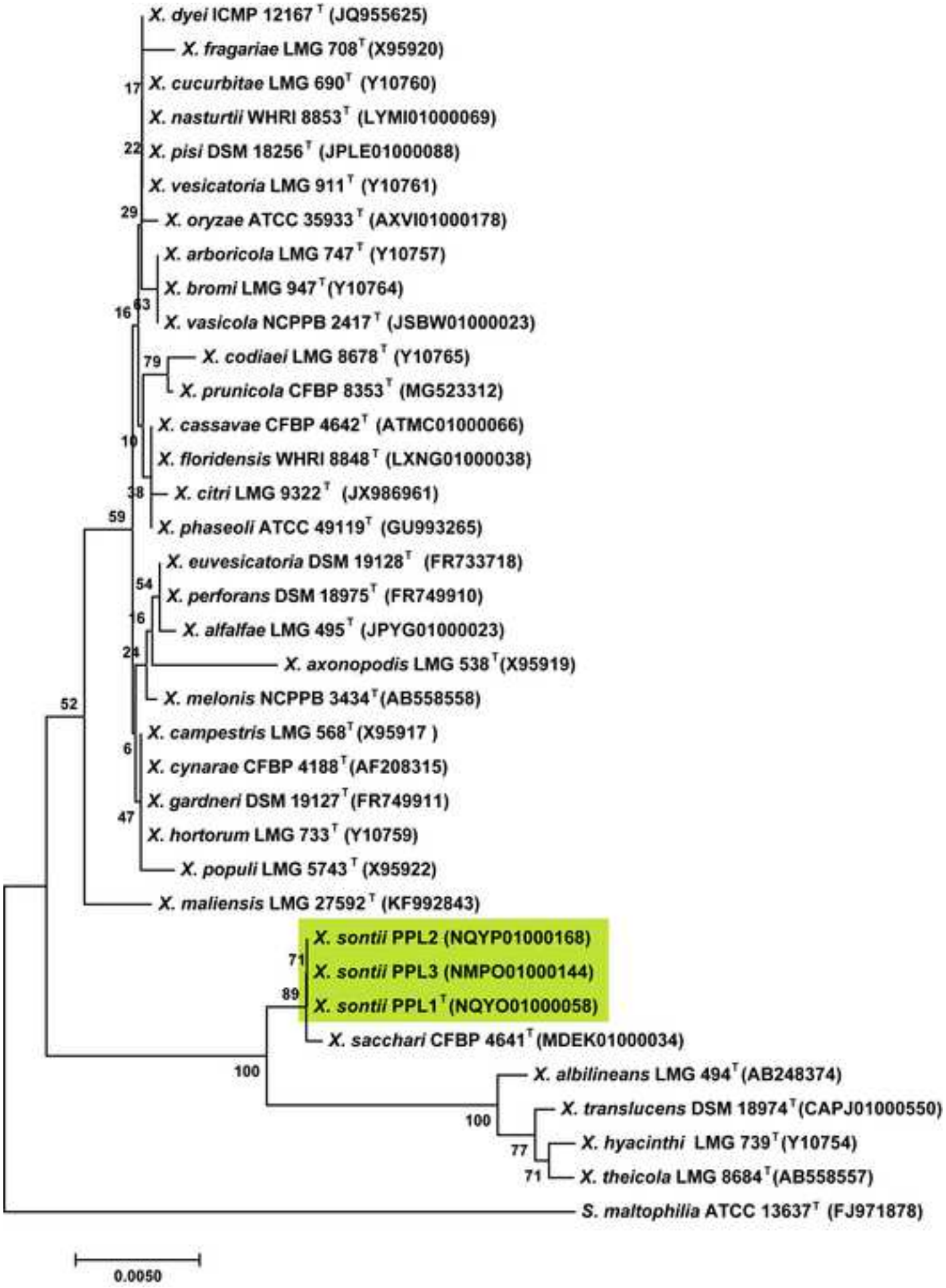

Fig. 2

Click here to access/download;Figure;Fig. 2.tif

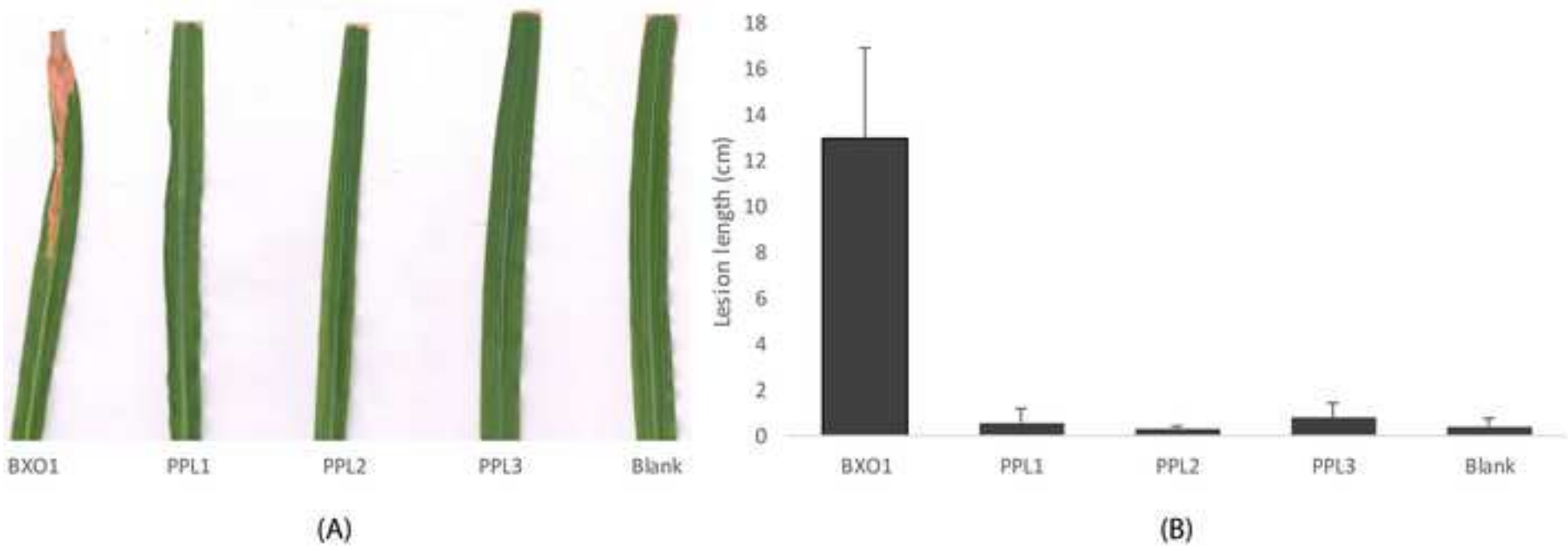

Fig. 3

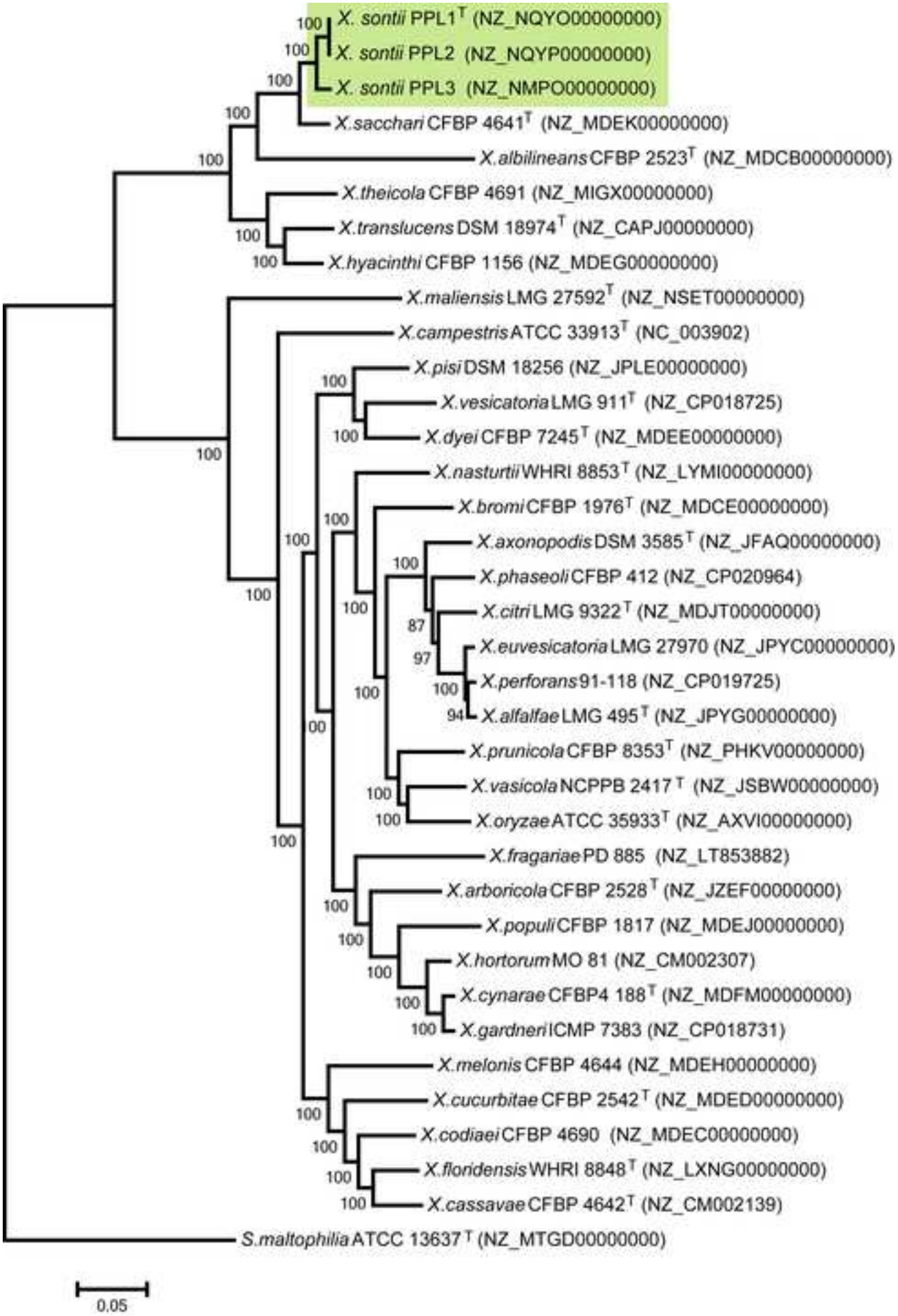

Fig. 4

Click here to access/download;Figure;Fig. 4.tif

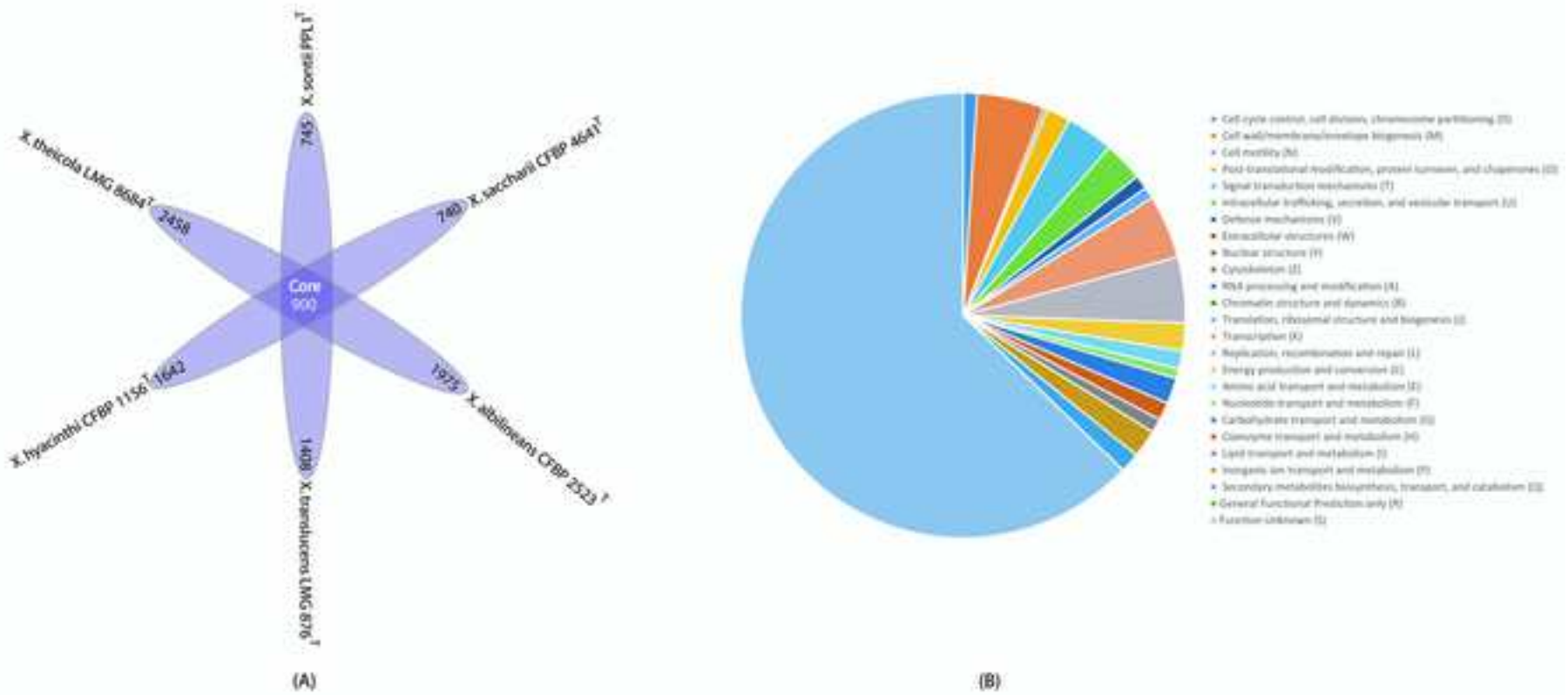

Fig. 5

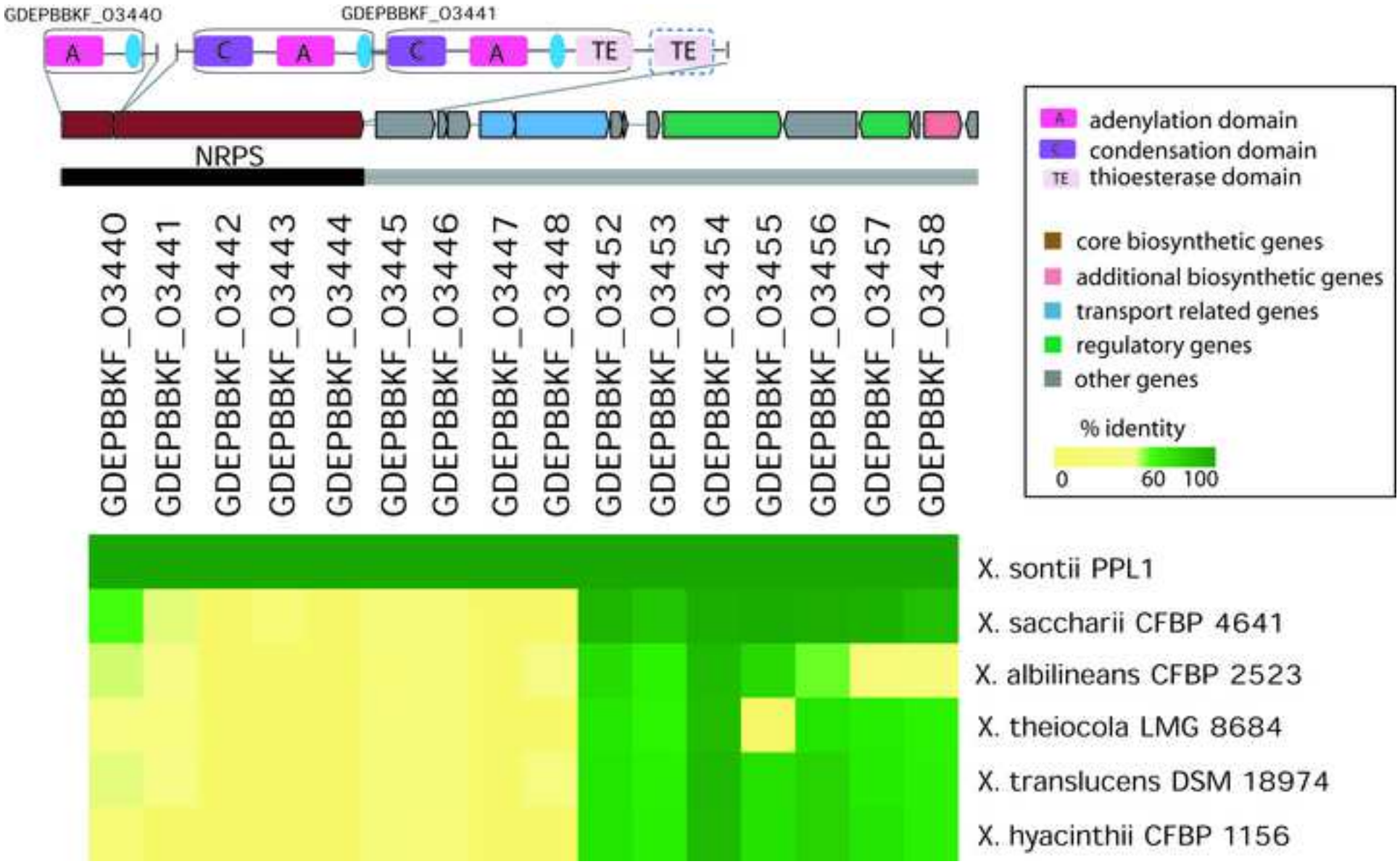

Fig. S1

Click here to access/download;Figure;Fig. S1.tif