MPMI Vol. 36, No. 1, 2023, pp. 73–77, https://doi.org/10.1094/MPMI-09-22-0191-A

RESOURCE ANNOUNCEMENT

The Complete Genome Resource of *Xanthomonas oryzae* pv. *oryzae* CIX2779 Includes the First Sequence of a Plasmid for an African Representative of This Rice Pathogen

Coline Sciallano, Florence Auguy, Gabriel Boulard, Boris Szurek, and Sébastien Cunnac †

PHIM Plant Health Institute, Univ Montpellier, IRD, CIRAD, INRAE, Institut Agro, Montpellier, France

Abstract

The bacterial plant pathogen *Xanthomonas oryzae* pv. *oryzae* is responsible for the foliar rice bacterial blight disease. Genetically contrasted, continent-specific, sublineages of this species can cause important damages to rice production both in Asia and Africa. We report on the genome of the CIX2779 strain of this pathogen, previously named NAI1 and originating from Niger. Oxford Nanopore long reads assembly and Illumina short reads polishing produced a genome sequence composed of a 4,725,792-bp circular chromosome and a 39,798-bp-long circular plasmid designated pCIX2779_1. The chromosome structure and base-level sequence are highly related to reference strains of African *X. oryzae* pv. *oryzae* and encode identical transcription activator-like effectors for virulence. Importantly, our *in silico* analysis strongly indicates that pCIX2779_1 is a genuine conjugative plasmid, the first indigenous one sequenced from an African strain of the *X. oryzae* species.

Genome Announcement

The plant-pathogenic bacterium *Xanthomonas oryzae* pv. *oryzae* is the causative agent of bacterial leaf blight (BLB) of rice (*Oryza sativa*). This disease has been reported in most rice-growing areas of Asia and Africa and can severely harm yields and grain quality. The genetic diversity of the *X. oryzae* pv. *oryzae* pathovar can be partitioned into two groups that reflect the continent of origin of the strains. In Africa, BLB was first documented in 1979 in Mali and is being increasingly reported (Verdier et al. 2012). African *X. oryzae* pv. *oryzae* strains notably differ from their Asian counterparts at the level of their transcription activator-like effector (TALEs) repertoires, holding up to nine members that are unique to African strains (Doucouré et al. 2018; Tran et al. 2018).

In addition to bona fide chromosomes, bacterial genomes may include extrachromosomal and autonomously replicating DNA molecules that are referred to as plasmids. They vary greatly in terms of size, sequence content, or contribution to fitness. These genetic elements are often capable of shuttling between bacterial genotypes of varying relatedness and are key vehicles for horizontal gene transfer. Because they may contain antibiotic or heavy metal–resistance genes as well as virulence systems, plasmids can have a profound impact on the adaptation of bacterial pathogens to their hosts (Frost et al. 2005; Sundin 2007). Our knowledge on plasmids within *X. oryzae* strains is rather limited. Earlier molecular genetics work revealed that 14 of 17 Asian *X. oryzae* pv. *oryzae* isolates

† Corresponding author: S. Cunnac; sebastien.cunnac@ird.fr

The author(s) declare no conflict of interest.

Accepted for publication 21 October 2022.

Funding

This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation (INV-008733). Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission.

Keywords

Oxford Nanopore Technologies, plant disease, TAL effector

Table 1. Main features of the CIX2779 strain genome sequences assembly and annotation

Feature	Chromosome	pCIX2779_1
GenBank accessions	CP101719	CP101720
Size (bp)	4,725,792	39,798
Oxford Nanopore Technologies coverage	90.7×	267.9×
Illumina coverage	103.3×	381.6×
Count of Polypolish-corrected positions	398	8
G+C content (%)	63.9	61.0
RefSeq assembly accession	GCF_024981335.1	
Predicted genes	4,398	
Predicted coding sequences	4,177	
Predicted RNA genes	221	
Predicted pseudogenes	540	

carry at least a plasmid. Furthermore, one of them contributed to symptom development (Amuthan and Mahadevan 1994). It is only recently that the first indigenous *X. oryzae* plasmid sequence was determined by Sanger sequencing from a Chinese isolate of the *oryzicola* pathovar. pXOCgx01 is a 53-kb conjugative plasmid that improves tolerance to heavy metal (Niu et al. 2015). Subsequently, the first plasmid sequences from *X. oryzae* pv. *oryzae* strains were assembled with long-read sequencing of the genome of the Indian BXO1 (Kaur et al. 2019) and IX-280 (Carpenter et al. 2020) strains. However, so far, the plasmid content of African *X. oryzae* remains unexplored.

To build genomic resources enabling a better exploration of the genetic diversity of African *X. oryzae* pv. *oryzae* strains, our group conducted long-read sequencing of several strains in our collection with the MinION device of Oxford Nanopore Technologies (ONT). Here, we report on the genome sequence of the African *X. oryzae* pv. *oryzae* CIX2779 strain, which was isolated in 2004 from a diseased rice leaf collected in the Saga locality of the Niamey area in Niger and which was originally designated as NAI1 by Gonzalez et al. (2007).

Following extraction with Qiagen Genomic-tip 100/G, the CIX2779 genomic DNA was used in multiplex library construction with the ONT rapid library preparation kit and was sequenced with a MinION Mk1C device on a R10.3 (FLO-MIN111) flowcell. DNA read sequences were obtained with the Guppy basecalling software (v6.0.1+652ffd179) and a high-accuracy model (dna_r10.3_450bps_hac.cfg). To produce a circularized and polished genome assembly, those sequences served as input to the CulebrONT pipeline (v2.0.1) (Orjuela et al. 2022) with the FLYE (2.9-b1768) and Miniasm (0.3-r179) assemblers and the Medaka polisher tool. A preliminary inspection of the resulting assemblies indicated that they are structurally highly consistent across the tested assemblers (Fig. 1A) and include two circular contigs. We anticipated the presence of extra-chromosomal elements in the CIX2779 genome because, for this strain in contrast to a typical African *X. oryzae* pv. *oryzae* strain, the profile of the purified genomic DNA in pulsed field gel electrophoresis included an additional high–molecular weight band of lower intensity that could correspond to relaxed circular plasmid DNA (Fig. 1B). In order to improve base-level accuracy, the FLYE primary assembly sequences were further polished with Polypolish (v0.5.0), as described previously (Wick and Holt 2022), with 150-bp paired end Illumina NextSeq reads derived from 350-bp inserts (library construction and sequencing was performed by FASTERIS, Plan-les-Ouates, Switzerland). This yielded a finished genome consisting of a 4,725,792-bp circular chromosome and a 39,798-bp long circular plasmid named pCIX2779_1 (Table 1). Interestingly, both ONT and Illumina read coverages for pCIX2779_1 are about three- to fourfold those of the chromosome, suggesting that, on average, there are approximately three copies of this plasmid in a CIX779 cell. As a primary sequence quality control measure, gene content was estimated first with the BUSCO software (Manni et al. 2021), which returned 98.52% complete, 0.7% fragmented and 0.78% missing metrics against the single-copy orthologs in the Xanthomonadales Odb10 reference dataset. This is in line with other complete *X. oryzae* pv. *oryzae* genomes found in sequence databases. This assembly was also annotated with the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (Tatusova et al.

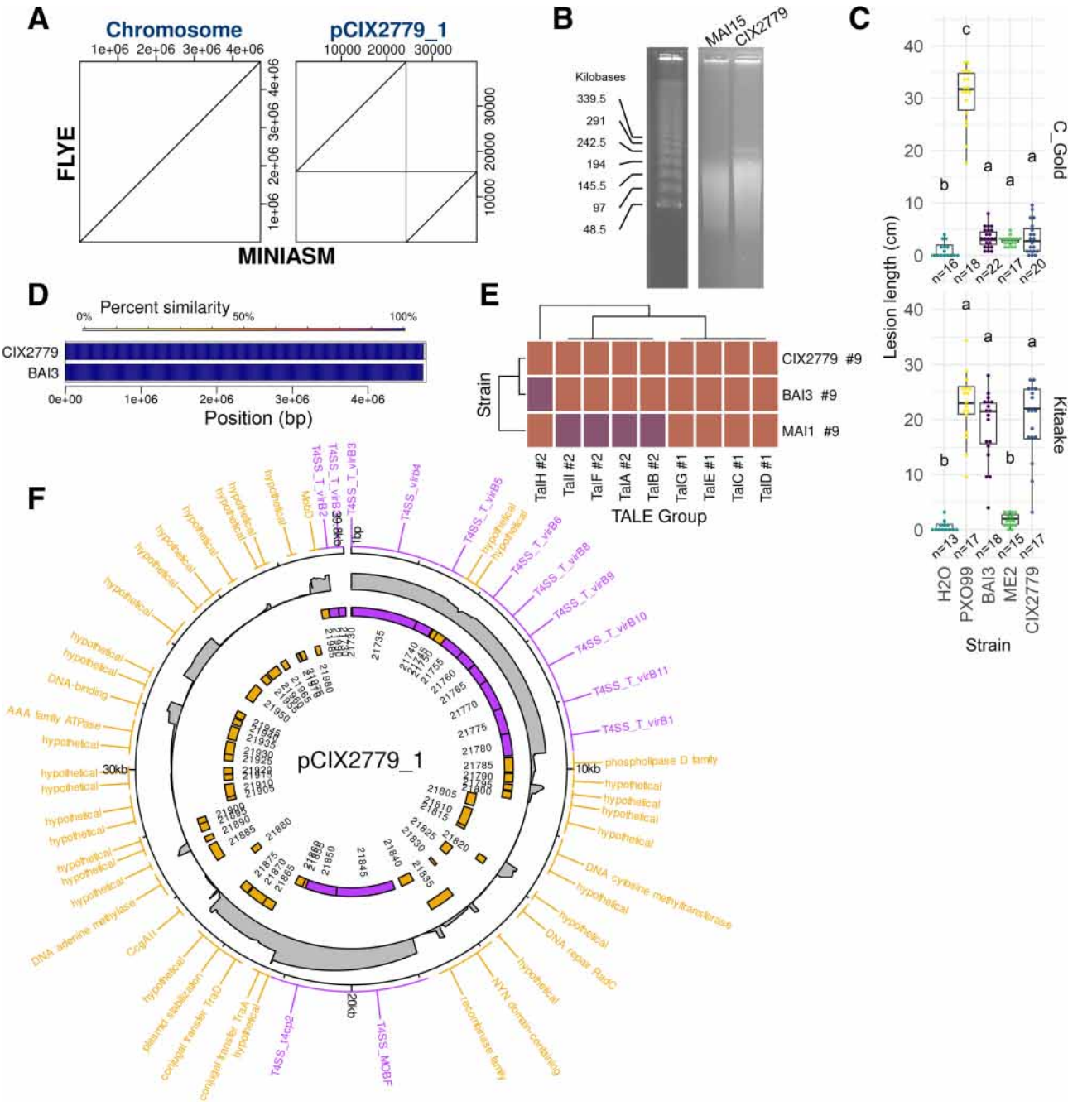

Fig. 1. The African *Xanthomonas oryzae* pv. *oryzae* CIX2779 strain harbors a approximately 40-kb conjugative plasmid. **A,** Dotplot representations of alignments between the Flye and Miniasm versions of the CIX2779 chromosome or pCIX2779_1. Alignments were obtained with the FindSynteny and AlignSynteny functions from the DECIPHER R package. **B,** A stained agarose gel after pulsed field electrophoresis of CIX2779 or MAI15 (as an African *X. oryzae* pv. *oryzae* reference) strains purified genomic DNA. Both the ladder and the samples panels derive from the same gel image (unaltered after acquisition) that was cropped to focus on the relevant areas. Samples were run for 21 h at 4.5 V cm^{-2}, with a 120-degree angle, in a Tris-borate-EDTA 0.5×, 1% agarose gel. **C,** Lesion length data measured two weeks after inoculation, from two replicate leaf-clipping experiments on the Kitaake or the Carolina Gold (C_Gold) varieties, with water (H_20), the CIX2779 strain, or control strains PXO99 (the $PXO99^A$ Asian *X. oryzae* pv. *oryzae* reference strain), BAI3 (the African *X. oryzae* pv. *oryzae* reference strain), and ME2 (a nonvirulent *pthXo1 tal* gene–inactivated mutant strain of PXO99A). Strains causing statistically different lesion length values on a host genotype, based on a Dunn's test for pairwise multiple comparisons of the ranked data (*P* values < 0.05) were labeled with different letters. **D,** Sequence alignment of the CIX2779 and BAI3 genomes performed as in A. **E,** Putative transcription activator-like effector (TALE) proteins were predicted and processed as before (Doucouré et al. 2018) to produce this summary of TALE diversity in the CIX2779 and reference African strain genomes. Columns correspond to groups of evolutionarily related sequences. Cells in the matrix with identical fill color across a column indicate that the corresponding TALEs have identical repeat variable di-residue sequences. Values after the hash sign indicate the count of TALEs in a strain (rows) or number of distinct variants in a group (columns). **F,** Circular representation of the pCIX2779_1 plasmid with National Center for Biotechnology Information (NCBI) gene descriptions (golden color) or CONJscan model hits (violet color). The external line track (gray-filled) represents the coverage (0 to 94) of blastn hits with the set of PLSDB related plasmid sequences. The colored rectangles on the intermediate and inner circles with NCBI locus number labels delineate predicted coding sequence spans in the forward or reverse orientation, respectively.

2016). Summary metrics for this annotation are also reported in Table 1 and are comparable to those found on NCBI for other complete African *X. oryzae* pv. *oryzae* genomes. Disease assays (Fig. 1C) indicated that strain CIX2779 causes typical BLB symptoms on the susceptible Kitaake variety but is avirulent on the Carolina Gold variety, which contains the *Xo1* resistance gene. *Xo1* confers sublineage-wide resistance to African *X. oryzae* (Triplett et al. 2016). Thus, CIX2779 has virulence properties reminiscent of African *X. oryzae* pv. *oryzae* strains. To ascertain that the nature of the finished genome sequence was coherent with these observations, it was used to interrogate the LINbase (Tian et al. 2020) server for digital taxonomy assignment. The closest database match was the genome of the African strain *X. oryzae* pv. *oryzae* AXO1947 (Huguet-Tapia et al. 2016), with a FastANI value of 99.98%. In addition, apart from pCIX2779_1, the CIX2279 genome displays high similarity and colinearity with the BAI3 African *X. oryzae* pv. *oryzae* strain genome (Tran et al. 2018) (Fig. 1D). Because TALEs are key virulence determinants of *X. oryzae* pv. *oryzae* for rice colonization, we investigated the TALE coding potential in the CIX2779 genome. As summarized in Figure 1E, CIX2779 predicted TALEs are chromosomally encoded and highly similar to those of the BAI3 strain. Their sequences of repeat variable di-residues (RVDs) are identical to their counterpart in BAI3 with the exception of the TalH variant, whose sequence of RVD is identical to TalH from the MAI1 reference strain (Tran et al. 2018).

Next, because pCIX2779_1 would be the first example of a plasmid in African *X. oryzae* genomes, we investigated in greater detail whether its sequence exhibits canonical features of this type of genetic elements. To identify similarities with reference plasmids, we used the mash dist strategy of PLSDB (Galata et al. 2019) with default parameters. The pCIX2779_1 sequence elicited hits with 48 plasmids (Fig. 1F), all of them originating from strains of the *Xanthomonas* genus. The best two hits corresponded to plasmids hosted by a *X. euvesicatoria* pv. *alfalfae* and a *X. vasicola* pv. *vasculorum* strain, respectively, both originating from Africa. This list also included three indigenous plasmids from Asian *X. oryzae* hosts (GenBank accessions NZ_CP019227.1, NZ_CP011963.1, and NZ_KR071788.1). To gain a better sense of the classification of pCIX2779_1, we ran the typer tool of the MOB-suite (Robertson and Nash 2018). In this analysis, pCIX2779_1 was predicted to be a conjugative plasmid because it harbors a relaxase of the MOB_F type and a mate-pair formation marker of type MPF_T (epitomized by the *vir* system of *Agrobacterium* spp.). However, no origin of transfer nor replication type could be ascribed, potentially because of the inadequate coverage of the underlying databases (Robertson and Nash 2018). The detection of a mating pair formation complex, which is a form of type 4 secretion system (T4SS) in pCIX2779_1 prompted us to examine secretion system coding potential more broadly. For this, the NCBI annotation predicted proteins served as input of the Macromolecular System Finder (MacSyFinder) tool together with the TXSScan models bundle (Abby et al. 2016). This analysis did not predict secretion systems other than a T4SS. The CONJscan module (Guglielmini et al. 2014) of MacSyFinder detected 12 components (all mandatory and accessory genes) for the T4SS_typeT model (VirB-like systems) which, in agreement with the typing of the MOB-suite, evaluated as the best solution. The corresponding hits were highlighted in the pCIX2779_1 map of Figure 1F. All together, these results support the hypothesis that pCIX2778_1 encodes a functional conjugation system. However, the majority of the description field for other annotated coding sequences reports on 'hypothetical' proteins and, besides additional elements related to conjugation or DNA metabolism, does not inform on potential other biological functions encoded on this plasmid (Fig. 1F).

To date about 30 long-read sequencing-assembled genomes of African *X. oryzae* pv. *oryzae* strains are publicly available. Yet, the pCIX2779_1 sequence assembled here with Nanopore reads corresponds to the first indigenous plasmid discovered in a genome with a chromosomal genetic content typical of African strains. This sequence resource for *X. oryzae* pv. *oryzae* CIX2779 will serve as a foundation for future questions about the prevalence, diversity, and evolution of plasmids in African *X. oryzae* pv. *oryzae* host strains and, overall, expands the genomic data on the species for the design of new disease management strategies.

Data Availability

All sequence resources have been submitted to NCBI under the BioProject accession number PRJNA861160, which links out to the assembly (RefSeq assembly ac-

cession GCF_024981335.1) and primary ONT (Short Read Archive [SRA] accession SRX17276343) and Illumina (SRA accession SRX17276344) reads sequences. The CIX2779 strain is available upon request from the corresponding author.

Acknowledgments

The authors acknowledge the ISO 9001 certified IRD i-Trop HPC (South Green Platform) at IRD Montpellier for providing high performance computing resources.

Literature Cited

Abby, S. S., Cury, J., Guglielmini, J., Néron, B., Touchon, M., and Rocha, E. P. C. 2016. Identification of protein secretion systems in bacterial genomes. Sci Rep. 6:23080.

Amuthan, G., and Mahadevan, A. 1994. Plasmid and pathogenicity in *Xanthomonas oryzae* pathovar *oryzae*, the bacterial blight pathogen of *Oryza sativa*. J. Appl. Bacteriol. 76:529-538.

Carpenter, S. C. D., Mishra, P., Ghoshal, C., Dash, P. K., Wang, L., Midha, S., Laha, G. S., Lore, J. S., Kositratana, W., Singh, N. K., Singh, K., Patil, P. B., Oliva, R., Patarapuwadol, S., Bogdanove, A. J., and Rai, R. 2020. An *xa5* resistance gene-breaking indian strain of the rice bacterial blight pathogen *Xanthomonas oryzae* pv. *oryzae* is nearly identical to a Thai strain. Front. Microbiol. 11:579504.

Doucouré, H., Pérez-Quintero, A. L., Reshetnyak, G., Tekete, C., Auguy, F., Thomas, E., Koebnik, R., Szurek, B., Koita, O., Verdier, V., and Cunnac, S. 2018. Functional and genome sequence-driven characterization of *tal* effector gene repertoires reveals novel variants with altered specificities in closely related Malian *Xanthomonas oryzae* pv. *oryzae* strains. Front. Microbiol. 9:1657.

Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. 2005. Mobile genetic elements: The agents of open source evolution. Nat. Rev. Microbiol. 3:722-732.

Galata, V., Fehlmann, T., Backes, C., and Keller, A. 2019. PLSDB: A resource of complete bacterial plasmids. Nucleic Acids Res. 47:D195-D202.

Gonzalez, C., Szurek, B., Manceau, C., Mathieu, T., Séré, Y., and Verdier, V. 2007. Molecular and pathotypic characterization of new *Xanthomonas oryzae* strains from West Africa. Mol. Plant-Microbe Interact. 20:534-546.

Guglielmini, J., Néron, B., Abby, S. S., Garcillán-Barcia, M. P., de la Cruz, F., and Rocha, E. P. C. 2014. Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. Nucleic Acids Res. 42:5715-5727.

Huguet-Tapia, J. C., Peng, Z., Yang, B., Yin, Z., Liu, S., and White, F. F. 2016. Complete genome sequence of the African strain AXO1947 of *Xanthomonas oryzae* pv. *oryzae*. Microbiol. Genome Announcements 4:e01730-15.

Kaur, A., Bansal, K., Kumar, S., Sonti, R. V., and Patil, P. B. 2019. Complete genome dynamics of a dominant-lineage strain of *Xanthomonas oryzae* pv. *oryzae* harbouring a novel plasmid encoding a type IV secretion system. Access Microbiol. 1:e000063.

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. 2021. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol. Biol. Evol. 38:4647-4654.

Niu, X.-N., Wei, Z.-Q., Zou, H.-F., Xie, G.-G., Wu, F., Li, K.-J., Jiang, W., Tang, J.-L., and He, Y.-Q. 2015. Complete sequence and detailed analysis of the first indigenous plasmid from *Xanthomonas oryzae* pv. *oryzicola*. BMC Microbiol. 15:233.

Orjuela, J., Comte, A., Ravel, S., Charriat, F., Vi, T., Sabot, F., and Cunnac, S. 2022. CulebrONT: A streamlined long reads multi-assembler pipeline for prokaryotic and eukaryotic genomes. Peer Community J. 2.

Robertson, J., and Nash, J. H. E. 2018. MOB-suite: Software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microb. Genom. 4.

Sundin, G. W. 2007. Genomic insights into the contribution of phytopathogenic bacterial plasmids to the evolutionary history of their hosts. Annu. Rev. Phytopathol. 45:129-151.

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M., and Ostell, J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 44:6614-6624.

Tian, L., Huang, C., Mazloom, R., Heath, L. S., and Vinatzer, B. A. 2020. LINbase: A web server for genome-based identification of prokaryotes as members of crowdsourced taxa. Nucleic Acids Res. 48:W529-W537.

Tran, T. T., Pérez-Quintero, A. L., Wonni, I., Carpenter, S. C. D., Yu, Y., Wang, L., Leach, J. E., Verdier, V., Cunnac, S., Bogdanove, A. J., Koebnik, R., Hutin, M., and Szurek, B. 2018. Functional analysis of African *Xanthomonas oryzae* pv. *oryzae* TALomes reveals a new susceptibility gene in bacterial leaf blight of rice. PLoS Pathog. 14:e1007092.

Triplett, L. R., Cohen, S. P., Heffelfinger, C., Schmidt, C. L., Huerta, A. I., Tekete, C., Verdier, V., Bogdanove, A. J., and Leach, J. E. 2016. A resistance locus in the American heirloom rice variety Carolina Gold Select is triggered by TAL effectors with diverse predicted targets and is effective against African strains of *Xanthomonas oryzae* pv. *oryzicola*. Plant J. 87:472-483.

Verdier, V., Vera Cruz, C., and Leach, J. E. 2012. Controlling rice bacterial blight in Africa: Needs and prospects. J. Biotechnol. 159:320-328.

Wick, R. R., and Holt, K. E. 2022. Polypolish: Short-read polishing of long-read bacterial genome assemblies. PLoS Comput. Biol. 18:e1009802.