Genome Announcement:

The draft genome sequence of *Xanthomonas* species strain Nyagatare, isolated from diseased bean in Rwanda

Running head: Genome sequence of *Xanthomonas* species strain Nyagatare

Valente Aritua[1], Augustine Musoni[2], Fred Kato[1], Mathew M. Abang[3], Robin Buruchara[4], Melanie Sapp[5], James Harrison[6], David J. Studholme[6*] and Julian Smith[5]

[1]International Center for Tropical Agriculture, P.O. Box 6247, Kampala, Uganda.

[2]Rwanda Agriculture Board, P.O. Box 5016 Kigali, Rwanda.

[3]FAO Sub-regional Office for Eastern Africa, P.O. Box 5536, Addis Ababa, Ethiopia.

[4]International Centre for Tropical Agriculture (CIAT) P.O. Box 823-00621, Nairobi, Kenya.

[5] International Development, The Food and Environment Research Agency, Sand Hutton, York, YO41 1LZ, UK.

[6]Biosciences, University of Exeter, Exeter EX4 4QD, Devon, UK.

Corresponding author: DJS.

Phone +44 (0) 1392 724678

Fax: +44 (0) 1392 263434

Email: d.j.studholme@exeter.ac.uk

Keywords: common beans; bacterial canker; *Xanthomonas*; Rwanda;

Abstract

We announce the genome sequence for *Xanthomonas* species strain Nyagatare, isolated from beans showing unusual disease symptoms in Rwanda. This strain represents the first sequenced genome belonging to an as-yet undescribed *Xanthomonas* species known as Species-Level Clade 1. It has at least 100 kb of genomic sequence that shows little or no sequence similarity to other xanthomonads, including a unique lipopolysaccharide synthesis gene cluster. At least one genomic region appears to have been acquired from relatives of *Agrobacterium* or *Rhizobium* species. The genome encodes homologues of only three known type-three secretion system effectors: AvrBs2, XopF1 and AvrXv4. Availability of the genome sequence will facilitate development of molecular tools for detection and diagnostics for this newly discovered pathogen of beans and facilitate epidemiological investigations of a potential causal link between this pathogen and the disease outbreak.

Main text

Common bean (*Phaseolus vulgaris*) is an important subsistence and cash crop for smallholder farmers in Rwanda, providing a major source of protein and micronutrients such as iron and zinc (Larochelle & Alwang, 2014). In November 2013, farmers in Nyagatare district reported unusual disease on variety ISAR SCB 101 (RWR 2245). Leaf symptoms included curling of upper leaves, wilting, drying and dropping off. There were also brownish and white spots on affected leaves as well as brownish to dark necrosis on veins and margins. The stems and branches developed extensive white scabs, which later developed into grey gall-like structures. Green to dark-brown-black streaks and wounds that developed into cankers and necrotic tissues also developed on the stems. The pods developed grey scabs and spots coalescing into large swellings, similar to those on stems. Many of the pods were water soaked, aborted or poorly filled. On dissection, stem vascular tissues were untainted, suggesting the pathogen is intercellular. A survey by the Rwanda Agriculture Board in November 2013 found that 6 of the 14 sectors of the Nyagatare District were affected. Although the implications were serious for farmers concerned, overall the sitiation was not yet alarming with no more than 15 ha being affected but there is concernabout possible future spread.

Bacteria were isolated from diseased plant material on YDC (Yeast extract dextrose carbonate) medium at CIAT Pathology Laboratory, Uganda. Pathogenicity was demonstrated by inoculation of the isolated strain onto CAL96 beans under glasshouse conditions; symptoms are shown in the Supplementary Material. Genomic DNA was sequenced to approximately 58-fold coverage using the Illumina MiSeq with Nextera XT library preparation, generating 663,444 pairs of 300-bp reads and assembled into 91 scaffolds with

a total length of 4,885,384 bp and an N_{50} length of 101,745 bp using Velvet 1.2.10 (Zerbino & Birney, 2008) followed by gap-filling using GapCloser version 1.12-r6 (Luo *et al.*, 2012). Data are available at GenBank under accession numbers GCA_000764855.1 and JRQI00000000.1.

To investigate the core and variable portions of the genome, we used *dnadiff* from the Mummer package (Delcher *et al.*, 2002) to perform pairwise sequence comparisons between the Nyagatare strain genome and all previously sequenced *Xanthomonas* genomes (results are tabulated in the Supplementary Material Figure S1). The highest degree of shared accessory genome was with *X. arboricola* 3004 (73.73% of genome shared with Nyagatare). Figure 1A also provides an overview of genomic conservation and variation. The genome with greatest sequence similarity was *X. cassavae* (Bolot *et al.*, 2013) with 89.16% nucleotide sequence identity. Average nucleotide identity (ANI) values, as calculated by JSpecies (Richter & Rosselló-Móra, 2009), between members of a single species usually exceed 95%. The ANI values between Nyagatare and *X. cassavae* were 87.38% (ANIb) and 89.12% (ANIm). Between Nyagatare and *X. arboricola* 3004, ANIb was 85.54% and ANIm was 88.84%. Between Nagatare and *X. fuscans* the respective values for ANIb and ANIm were 85.82% and 88.66%. Thus strain Nyagatare does not belong to any of the previously sequenced species and is phylogenetically distinct from previously studied pathogens of common bean (that fall within the species *X. axonopodis* and *X. fuscans*). The lack of sequenced genomes with very high sequence similarity to strain Nyagatare precluded high-resolution phylogenomic analysis (Rodriguez-R *et al.*, 2012); however, the availability of an extensive database of sequences for the phylogenetic marker gene *gyrB* (Parkinson *et al.*, 2009) allowed us to more precisely examine its phylogenetic position. As illustrated in Figure 1B, the Nyagatare strain falls within Parkinson's Sequence-Level Clade 1 (Parkinson *et al.*,

2009), along with little-studied pathogens of *Zinnia elegans*, *Hibiscus esculentus*, *Cannabis sativa*, *Helianthus annuus* and *Nicotiana tabacum* (NCPPB strains 2439, 2190, 2877, 1325 and 1068).

Commensurate with its phylogenetic distinctness from previously sequenced *Xanthomonas* species, the Nyagatare strain has at least 100 kb of genomic sequence that shows little or no sequence similarity to other xanthomonads, as judged by BLASTN searches. This includes a 16.5-kb region located between *metB* and *etfA* (JRQI01000003.1 positions 48,238-64,812) harbouring genes for lipopolysaccharide (LPS) synthesis that are quite distinct from any previously sequenced LPS synthesis gene cluster (Patil & Sonti, 2004). Another example is a 2.3-kb region (JRQI01000032.1 positions 37,278-34,915) that shares 84% nucleotide sequence identity with the large chromosome of *Agrobacterium radiobacter* K84 (GenBank: CP000628.1), and similar levels of identity with several *Rhizobium* species, but shares no detectable sequence similarity with any available *Xanthomonas* sequences in the NCBI databases.

Virulence factors described in previously sequenced *Xanthomonas* genomes include effector proteins that are substrates of the type-III secretion system (T3SS) (White *et al.*, 2009). The Nyagatare genome encodes an apparently complete T3SS (Figure S2). Based on TBLASTN searches between the genome of the Nyagatare strain and Ralf Koebnik's catalogue of known T3SS effectors (http://www.xanthomonas.org/t3e.html) there are homologues of only three: AvrBs2 (73 % identity between GenBank: CAJ21683.1 and JRQI01000008.1: 30,926 to 33,058), XopF1 (66% identity between CAJ22045.1 and NC00_3340) and an open reading frame (JRQI01000008.1 positions 38,866 to 39,942) encoding a protein with 87% amino-acid sequence identity to AvrXv4 which has only previously been reported in

genomes of *X. euvesicatoria* (Astua-Monge *et al.*, 2000) and *X. perforans* (Potnis *et al.*, 2011).

In conclusion, we present a draft-quality genome sequence for the Nyagatare strain. This is the first genome sequence representing Parkinson's Species-Level Clade 1 and as such its availability will aid the study of this as-yet undescribed candidate new species. Furthermore, this strain may be responsible for the mysterious disease emerging as a potentially serious threat to beans, an important subsistence crop. Availability of the genome sequence will facilitate development of molecular tools for detection and diagnostics thus enable researchers to test for an epidemiological link between this strain and the disease.

Acknowledgments

Sequencing was made possible through the Canadian International Development Agency (now incorporated into the Department of Foreign Affairs, Trade and Development) support to Pan-Africa Bean Research Alliance. V. Aritua was supported on this work by the BBSRC SCPRID Bean Grant BB/J011568/1. We are indebted for contributions from Richard Thwaites (Fera).

References

Alikhan N-F, Petty NK, Ben Zakour NL & Beatson SA (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12**: 402.

Astua-Monge G, Minsavage G V, Stall RE, Davis MJ, Bonas U & Jones JB (2000) Resistance of tomato and pepper to T3 strains of *Xanthomonas campestris* pv. *vesicatoria* is specified by a plant-inducible avirulence gene. *Mol Plant Microbe Interact* **13**: 911–921.

Bolot S, Munoz Bodnar A, Cunnac S, *et al.* (2013) Draft Genome Sequence of the *Xanthomonas cassavae* Type Strain CFBP 4642. *Genome Announc* **1**.

Darrasse A, Carrère S, Barbe V, *et al.* (2013) Genome sequence of *Xanthomonas fuscans* subsp. *fuscans* strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. *BMC Genomics* **14**: 761.

Delcher AL, Phillippy A, Carlton J & Salzberg SL (2002) Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* **30**: 2478–2483.

Larochelle C & Alwang J (2014) Impacts of Improved Bean Varieties on Food Security in Rwanda Impacts of Improved Bean Varieties on Food Security in Rwanda.

Luo R, Liu B, Xie Y, *et al.* (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**: 18.

Parkinson N, Cowie C, Heeney J & Stead D (2009) Phylogenetic structure of *Xanthomonas* determined by comparison of gyrB sequences. *Int J Syst Evol Microbiol* **59**: 264–274.

Patil P & Sonti R (2004) Variation suggestive of horizontal gene transfer at a lipopolysaccharide (lps) biosynthetic locus in *Xanthomonas oryzae* pv. *oryzae*, the bacterial leaf blight pathogen of rice. *BMC Microbiol*.

Pieretti I, Royer M, Barbe V, *et al.* (2009) The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC Genomics* **10**: 616.

Potnis N, Krasileva K, Chow V, *et al.* (2011) Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. *BMC Genomics* **12**: 146.

Richter M & Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* **106**: 19126–19131.

Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD & Perna NT (2009) Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics* **25**: 2071–2073.

Rodriguez-R LM, Grajales A, Arrieta-Ortiz M, Salazar C, Restrepo S & Bernal A (2012) Genomes-based phylogeny of the genus *Xanthomonas*. *BMC Microbiol* **12**: 43.

Da Silva ACR, Ferro JA, Reinach FC, *et al.* (2002) Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**: 459–463.

Song C & Yang B (2010) Mutagenesis of 18 type III effectors reveals virulence function of XopZ(PXO99) in *Xanthomonas oryzae* pv. *oryzae*. *Mol Plant Microbe Interact* **23**: 893–902.

Tamura K & Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**: 512–526.

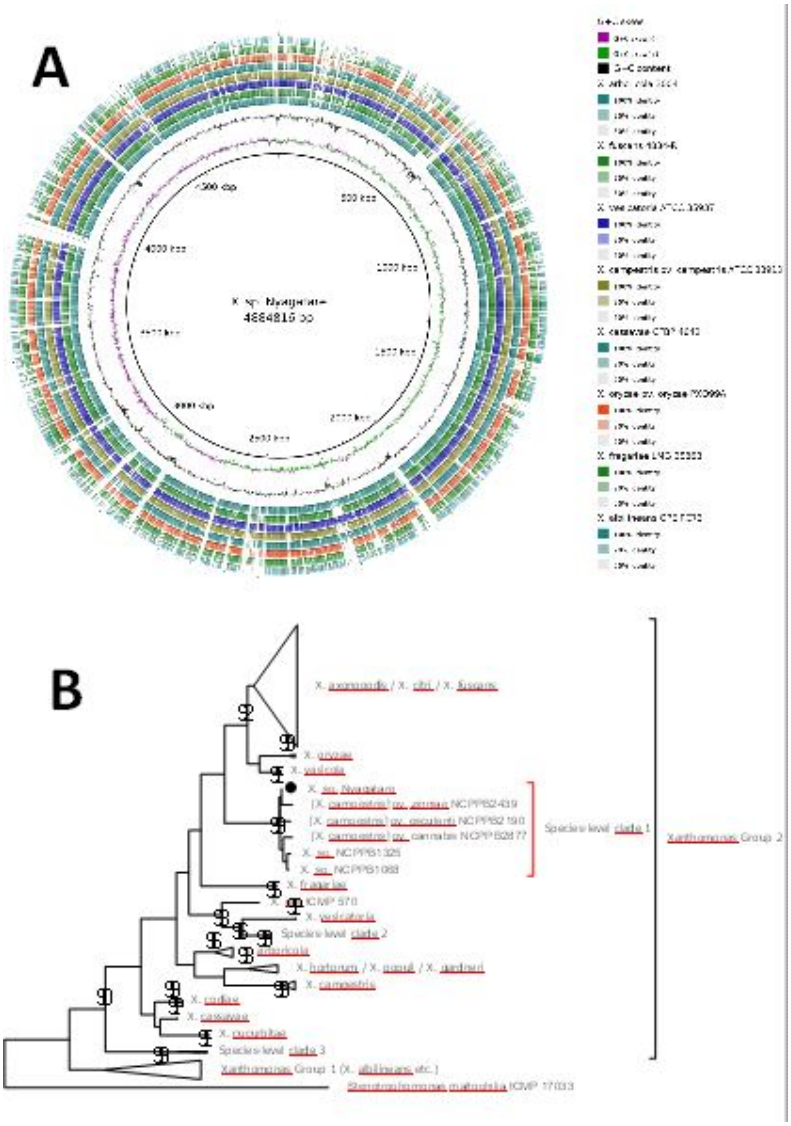
Tamura K, Stecher G, Peterson D, Filipski A & Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**: 2725–2729.

Vandroemme J, Cottyn B, Baeyen S, De Vos P & Maes M (2013) Draft genome sequence of *Xanthomonas fragariae* reveals reductive evolution and distinct virulence-related gene content. *BMC Genomics* **14**: 829.

White FF, Potnis N, Jones JB & Koebnik R (2009) The type III effectors of *Xanthomonas*. *Mol Plant Pathol* **10**: 749–766.

Young JM, Park D-C, Shearman HM & Fargier E (2008) A multilocus sequence analysis of the genus *Xanthomonas*. *Syst Appl Microbiol* **31**: 366–377.

Zerbino DR & Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**: 821–829.



177

178 **Figure 1. The genome sequence of *Xanthomonas* sp. Nyagatare**. Panel A shows a global
179 comparison of the Nyagatare genome sequence against representative previously
180 sequenced *Xanthomonas* genomes. The genome sequences (Pieretti *et al.*, 2009; Song &
181 Yang, 2010; Potnis *et al.*, 2011; Bolot *et al.*, 2013; Darrasse *et al.*, 2013; Vandroemme *et al.*,
182 2013) were aligned against the Nyagatare genome assembly using BLASTN with an E-value
183 threshold of $1x10^{-6}$. The Nyagatare assembly had first been re-ordered against the *X.*
184 *axonopodis* pv. *citri* 306 (da Silva *et al.*, 2002) reference sequence using the contig re-
185 ordering function in Mauve (Rissman *et al.*, 2009). The alignments are visualised using BLAST

Ring Image Generator (BRIG) (Alikhan *et al.*, 2011). Panel B shows the phylogenetic position of the Nyagatare strain based on comparison to previously sequenced *gyrB* genes (Parkinson *et al.*, 2009). Evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura & Nei, 1993). The tree with the highest log likelihood (-8634.7961) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 438 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 524 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura *et al.*, 2013). *Xanthomonas* group 1 and group 2, as defined by Young and colleagues (Young *et al.*, 2008) are indicated by square brackets as is also species-level clade 1 as defined by Parkinson and colleagues (Parkinson *et al.*, 2009).