Complete genome and plasmid sequence data of three *Xanthomonas arboricola* pv. *corylina* strains, the bacterium responsible for bacterial blight of hazelnut

Joël F. Pothier,[1#*] Monika Kałużna,[2#] Andjelka Prokić,[3] Aleksa Obradović,[3] and Fabio Rezzonico[1]

[1]Environmental Genomics and Systems Biology Research Group, Institute for Natural Resource Sciences, Zurich University of Applied Sciences (ZHAW), Wädenswil, Switzerland

[2]The National Institute of Horticultural Research, Konstytucji 3 Maja 1/3, 96-100 Skierniewice, Poland

[3]University of Belgrade, Faculty of Agriculture, Belgrade, Serbia

# These two authors contributed equally to this work.

*Corresponding author: J. F. Pothier; joel.pothier@zhaw.ch

Keywords**:** bacteria, complete genome, *Xanthomonas arboricola* pv. *corylina*, hazelnut

Funding

A.P. was granted a Short-Term Scientific Mission by the COST Action CA16107 EuroXanth to conduct some experiments in Wädenswil (CH).

This work was partly financed by the National Science Centre, Poland, Grant UMO-2017/26/M/NZ9/01024 granted to M.K.

J.F.P. and F.R. acknowledge support from the Department of Life Sciences and Facility Management of the Zurich University of Applied Sciences (ZHAW) in Wädenswil.

A.O. was supported by the Ministry of Education, Science and Technological Development, Republic of Serbia and the Faculty of Agriculture contract no. 451-03-9/2021-14/200116.

Abstract

Xanthomonas arboricola pv. *corylina* is the causal agent of bacterial blight of hazelnut. The bacterium is listed as A2 quarantine pathogen in Europe since 1978 and on the Regulated Non-Quarantine Pest (RNQP) list since 2020. Three strains from various geographic regions and isolated at different times were sequenced using a hybrid approach with short- and long-read technologies to generate closed genome and plasmid sequences in order to better understand the biology of this pathogen.

Genome Announcement

Bacterial blight of hazelnuts (*Corylus* spp.) was first reported in the early twentieth century in Oregon (Barss 1913; Kałużna et al. 2021). The disease is caused by *Xanthomonas arboricola* pv. *corylina* and has since been reported in countries from all continents apart Antarctica (Kałużna et al. 2021). To limit the risk of introduction to other countries especially via planting material, this Gram-negative bacterium was listed by the European Plant Protection Organization (EPPO) as A2 quarantine pathogen in 1978 and as Regulated Non-Quarantine Pest (RNQP; Picard et al. 2018) since 2020 (European Union 2020).

The most important host for *X. arboricola* pv. *corylina* is *Corylus avellana* L. (the common hazel) but other plant species such as *Corylus pontica*, *Corylus maxima* and *Corylus colurna* were also found to be susceptible although considered as minor hosts (OEPP/EPPO 1986, 2004).

Here, we report the complete genome sequences of three strains of *X. arboricola* pv. *corylina* (Table 1). The strains were isolated between 1939 and 2007 from either *C*.

avellana or *C. maxima* and from three different countries (Table 1). These complete genomes should contribute to unveiling the ecology, evolution, and virulence of this economically relevant bacterium forhazelnut cultivation.

The CFBP 1159^{PT} and CFBP 6600 strains were initially obtained as freeze-dried cultures in glass ampoules from the international strain collection CFBP (Collection Française de Bactéries Associées aux Plantes, Beaucouzé, France). Strains were revived, stored, and handled as described previously (Dia et al. 2020). Strain Xac301 was isolated in 2007 in Poland from symptomatic leaf spots of a hazelnut (Puławska et al. 2010). Characteristic mucoid yellow colonies were obtained and a pure colony initially called RIPF X12 (=Xac301) was further grown on yeast extract nutrient agar (YNA) medium. This isolate was identified as *X. arboricola* pv. *corylina* based on cellular fatty acid content converted to methyl esters (FAMES) as well as in *gyrB* gene fragment sequence analysis (Puławska et al. 2010). The *gyrB* sequence of Xac301 was most similar to the sequence of the *X. arboricola* pv. *corylina* pathotype strain (Fig. 1A). Koch's postulates were validated with strain Xac301 using leaf inoculation of hazelnut cvs. Webb's Prize Cob., Cosford and Merveille de Bollwiller (Fig. 1B). The isolate was stored in a -80°C ultra-freezer in mixture of glycerol 20% (*v*/*v*) and PBS buffer (0.27% Na_2HPO_4; 0.04% NaH_2PO_4; 0.8% NaCl) until further use. Before extraction of DNA, Xac301 was revived and cultured on YNA medium and incubated at 26°C for 48-72h.

For strain Xac301, genomic DNA (gDNA) for both short- and long-read sequencing was isolated using the modified method of Aljanabi and Martinez (1997) from cells grown overnight at 26°C on YNA as described previously (Kałużna et al. 2012). For short- read sequencing, library preparation was done using a NEBNext DNA Library Prep Master

Mix Set for Illumina (NEB, Ipswich, MA). Pooled libraries were sequenced on a MiSeq sequencer (Illumina, San Diego, CA) with 2×250-bp paired-end reads using a MiSeq reagent kit version 2 (Illumina).

For the CFBP 1159^{PT} and CFBP 6600 strains, gDNA for Illumina MiSeq short-read sequencing was extracted from cells grown overnight at 28°C in nutrient yeast extract glycerol broth using the NucleoSpin tissue kit (Macherey-Nagel, Düren, Germany), according to the manufacturer's protocol. The quality of the gDNA was checked using a fragment analyzer (Advanced Analytical Technologies, Inc. Ankeny, IA) and quantified using the Quant-iT PicoGreen double-stranded DNA quantification assay (Thermo Fisher Scientific, Waltham, MA). Library preparation was done using the Nextera XT DNA library prep kit (Illumina) following the manufacturer's instructions. Sequencing of pooled libraries was performed on a MiSeq Illumina sequencer with 2×300-bp paired-end reads using a MiSeq reagent kit version 3 (Illumina) according to the manufacturer's instructions.

For long-read sequencing the gDNA of strains CFBP 1159^{PT} and CFBP6600 was extracted from overnight-grown cells using the Gentra PureGene Yeast/Bact kit protocol (Qiagen, Hilden, Germany). The gDNA quality was checked as described above and quantified using a high sensitivity double-stranded DNA quantitation kit (Allsheng, Hangzhou, China) and a Fluo-100B fluorometer (Allsheng).

For the three strains, long-read library preparation and sequencing were performed with the ligation sequencing kit (catalog no. SQK-LSK109 for CFBP 1159^{PT} and CFBP 6600, catalog no. SQK-LSK108 for Xac301; Oxford Nanopore Technologies, Oxford, United Kingdom) and run on an R9.4.1 flow cell with a MinION sequencer. The native barcoding

97 expansion kit (catalog no. XP-NBD114) was used for multiplexing. Reads were
98 basecalled and demultiplexed using Guppy version 3.3.3.

99 Short- and long-read library preparation and sequencing were outsourced at Genomed
100 S.A. (Warsaw, Poland) in the case of strain Xac301. For strain CFBP 1159^{PT} and CFBP
101 6600, these steps were outsourced at BSSE Genomics Facility (Basel, Switzerland) for
102 short-read libraries and carried out in the Environmental Genomics and Systems Biology
103 Research Group lab facilities (ZHAW) for long-reads.

104 *De novo* hybrid assemblies using the MiSeq and MinION reads were conducted with
105 Trycycler version 0.3.3 (Wick et al. 2021). A total of 10,551, 10,533 and 199 nucleotide
106 changes were performed during the first short-read polishing round using Pilon version
107 1.22 for CFBP 1159^{PT}, CFBP 6600 and Xac301, respectively. The genomes were then
108 annotated using Prokka version 1.14.5 (Seemann 2014). All tools were run with default
109 parameters unless otherwise specified.

110 The size of the hybrid assemblies ranged from 5,080,866 to 5,294,219 bp, a size range
111 typically found in *Xanthomonas* genomes (Table 1). The G+C contents of the genomes
112 varied from 65.37% to 65.56%, also comparable to other *Xanthomonas* spp. G+C
113 contents. Whole-genome comparison based on average nucleotide identity using
114 BLASTN (ANIb) implemented in pyANI version 0.2.10 (Pritchard et al. 2016) confirmed
115 that the three strains had high degree of synteny between them (Table 1) and to other *X.*
116 *arboricola* genomes (data not shown). Genome completeness varied between 99.7%
117 and 99.8% (Table 1) when assessed using the Benchmarking Universal Single-Copy
118 Ortholog (BUSCO) version 5.2.1 (Manni et al. 2021) and the xanthomonodales_odb10
119 (2020-03-06) lineage dataset.

Since an assembly was already existing for the pathotype strain CFBP 1159^{PT}, a comparison was performed versus the hybrid assembly presented here which revealed some improvements and few minor differences (Table 2).

A single 24 kb plasmid was present in the final assemblies of CFBP 6600 and Xac301. This plasmid contains the type three effectors (T3E) XopAG (HopG1) and the avirulence protein XopE2. The presence of XopAG in two of the three strains and the plasmid-borne localization of this T3E agrees with previous observations from a draft genome sequence of this same pathovar (Ibarra Caballero et al. 2013). In the genome of CFBP 1159^{PT}, XopE2 was detected on the chromosome. The T3E AvrBs3 was also detected in the genome of the CFBP 1159^{PT} pathotype strain whereas it is absent from the two other genomes presented in this work as previously reported from another draft genome of this same pathovar (Ibarra Caballero et al. 2013). The presence of the *copAB* operon and *copL* gene whose products are involved in copper resistance in this pathogenic bacteria (Kałużna et al. 2021) was detected in the chromosome of all three strains. Similarly, *cutC* and *pCuAC* genes whose product could be implicated in the survival of this bacterium at high copper concentration (Nuñez Cerda et al. 2021) were also found in all three strains.

The sequenced genomes discussed here will be used for further analysis of evolution within the species *X. arboricola*, better understanding of the pathogenicity and virulence as well as development of improved tools for diagnostics of this relevant pathogen for the worldwide production of hazelnut.

Data availability

The raw data and assembled/annotated genome sequences have been deposited in the European Nucleotide Archive (ENA) under BioProject no. PRJEB42844. The genome and raw read accession numbers for each strain are shown in Table1.

The author(s) declare no conflict of interest.

Acknowledgements

The authors would like to thank the HPC team of the School for Life Sciences and Facility Management at ZHAW for computer resources and support.

This article is based upon work from COST Action CA16107 EuroXanth, supported by COST (European Cooperation in Science and Technology).

Literature Cited:

Aljanabi, S. M., and Martinez, I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Res. 25:4692-4693.

Barss, H. P. 1913. A new filbert disease in Oregon. Oregon Agricultural Experiment Station Biennial Crop Pest and Horticulture Report 14:213-223.

Dia, N. C., Rezzonico, F., Smits, T. H. M., and Pothier, J. F. 2020. Complete or high-quality draft genome sequences of six *Xanthomonas hortorum* strains sequenced with short- and long-read technologies. Microbiol. Resour. Announc. 9:e00828-00820.

European Union. 2020. Commission implementing directive (EU) 2020/177 of 11 february 2020 amending council directives 66/401/EEC, 66/402/EEC, 68/193/EEC, 2002/55/EC, 2002/56/EC and 2002/57/EC, commission directives 93/49/EEC and 93/61/EEC and implementing directives 2014/21/EU and

2014/98/EU as regards pests of plants on seeds and other plant reproductive material. Official Journal of the European Union L41:1-77.

Ibarra Caballero, J., Zerillo, M. M., Snelling, J., Boucher, C., and Tisserat, N. 2013. Genome sequence of *Xanthomonas arboricola* pv. *corylina*, isolated from turkish filbert in Colorado. Genome Announc. 1:e00246-00213.

Kałużna, M., Janse, J. D., and Young, J. M. 2012. Detection and identification methods and new tests as used and developped in the framework of COST 873 for bacteria pathogenic to stone fruits and nuts: *Pseudomonas syringae* pathovars J. Plant Pathol. 94:S1.117-S111.126.

Kałużna, M., Fischer-Le Saux, M., Pothier, J. F., Jacques, M.-A., Obradović, A., Tavares, F., and Stefani, E. 2021. *Xanthomonas arboricola* pv. *juglandis* and pv. *corylina*: brothers or distant relatives? Genetic clues, epidemiology, and insights for disease management. Mol. Plant Pathol. 00:1-19.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35:1547-1549.

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.

Nuñez Cerda, P. D., Muster, C., Lisperguer, M. J., Vargas, E., and Bustos, S. 2021. Complete genome of *Xanthomonas arboricola* pv. *corylina* strain A7 isolated from southern Chile. Mol. Plant-Microbe Interact. DOI:10.1094/mpmi-12-20-0363-a.

OEPP/EPPO. 1986. Data sheets on quarantine organisms No. 134, *Xanthomonas campestris* pv. *corylina*. Bulletin OEPP/EPPO Bulletin 16:13-16.

OEPP/EPPO. 2004. *Xanthomonas arboricol*a pv. *corylina*. EPPO Bulletin 34:179-181.

Picard, C., Afonso, T., Benko-Beloglavec, A., Karadjova, O., Matthews-Berry, S., Paunovic, S. A., Pietsch, M., Reed, P., van der Gaag, D. J., and Ward, M. 2018. Recommended regulated non-quarantine pests (RNQPs), associated thresholds and risk management measures in the European and Mediterranean region. EPPO Bulletin 48:552-568.

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Analytical Methods 8:12-24.

Puławska, J., Kałużna, M., Kołodziejska, A., and Sobiczewski, P. 2010. Identification and characterization of *Xanthomonas arboricola* pv. *corylina* causing bacterial blight of hazelnut: a new disease in Poland. J. Plant Pathol. 92:803-806.

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

Wick, R. R., Judd, L. M., Cerdeira, L. T., Hawkey, J., Méric, G., Vezina, B., Wyres, K. L., and Holt, K. E. 2021. Trycycler: consensus long-read assemblies for bacterial genomes. bioRxiv DOI:10.1101/2021.07.04.451066.

Table 1. Genome metrics and accession numbers of the newly sequenced *Xanthomonas arboricola* pv. *corylina* genomes

							Illumina data			Oxford Nanopore data						
Strain[a]	**Origin (yr)**	**Host**	**Genome size (bp)**	**G+C content (%)**	**Total no. of genes**	**No. of plasmids**	**Total no. of reads**	**Avg read length (bp)**	**Avg coverage (×)**	**Total no. of reads**	**Read length N_{50} (bp)**	**Avg coverage (×)**	**SRA accession no. (MinION/MiSeq)**	**ENA accession no.[b]**	**ANI[c]**	**BUSCO score (%)**
CFBP 1159[PT]	USA (1939)	*Corylus maxima*	5,080,866	65.56	4,279	0	1,511,890	301	77	50,946	32,779	44	ERR5260054 ERR5260059	HG992341 (chr.)	100	99.8
CFBP 6600	France (1977)	*Corylus avellana*	5,234,232	65.42	4,499	1	1,952,794	301	100	27,923	38,810	20	ERR5260055 ERR5260060	HG992342 (chr.), HG992343 (p24)	99.70	99.7
Xac301	Poland (2007)	*Corylus avellana*	5,294,219	65.37	4,461	1	2,516,854	251	110	129,031	22,574	295	ERR5260056 ERR5260083	HG992338 (chr.), HG992339 (p24)	99.93	99.7

[a] The culture collection providing strains is abbreviated in the strain name as CFBP (Collection Française de Bactéries Associées aux Plantes, Beaucouzé, France). Superscript ^{PT} following the strain name indicates the pathotype strain for the pathovar.

[b] chr., chromosome.

[c] Average nucleotide indentity (ANI) using BLAST (ANIb) is relative to CFBP 1159^{PT}.

Table 2. Comparison of the improved hybrid assembly versus the existing assembly of *Xanthomonas arboricola* pv. *corylina* CFBP 1159^{PT}

Assembly name	GCA_905220785	ASM293984v1
Sequencing technology	Illumina MiSeq + Oxford Nanopore MinION	Illumina HiSeq
Assembler	Trycycler v.0.3.3 + Pilon v.1.22	Velvet v.1.2.07 + SOAPdenovo v.2.04
Coverage	121×	100×
Total sequence length (bp)	5,080,866	5,105,973
No. of contigs	1	124
N_{50}	5,080,866	135,548
G+C content (%)	65.56	65.50
Annotation pipeline	Prokka v.1.14.5	NCBI PGAP v.4.2
No. of CDS	4,214	4,394
No. of rRNAs (5S, 16S, 23S)	2, 2, 2	1, 1, 1
No. of tRNAs	57	51
BUSCO score (%)	99.8	99.8

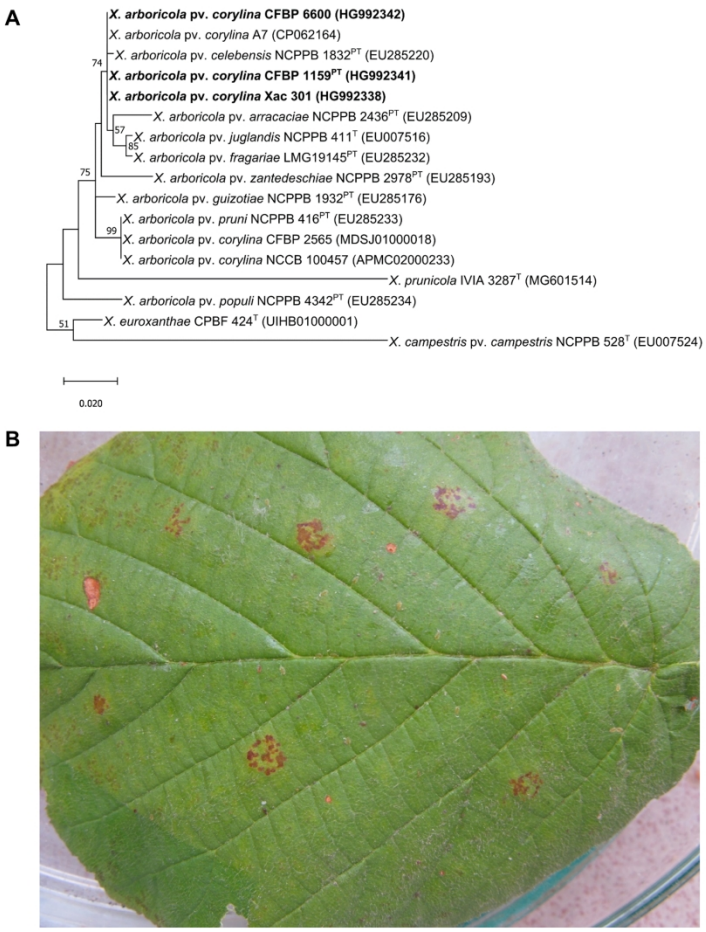

Fig. 1. **A,** Maximum-likelihood unrooted phylogenetic tree based on a 530 bp *gyrB* partial sequences of *Xanthomonas* strains. Phylogenetic and molecular evolutionary analyses were conducted using MEGA X version 10.0.5 (Kumar et al. 2018). The alignment was obtained using the MUSCLE algorithm. The tree was constructed using the JTT matrix-based model. Percent bootstrap values calculated for 1,000 iterations are indicated near nodes and displayed only when over 50. Accession numbers or source for *gyrB* sequences are indicated within parentheses next to the species name, with strains sequenced in this study marked in bold. Bar represents the expected number of substitutions per site. Superscripts following strain names indicate ^{T} the type strain of a species and ^{PT} the pathotype strain for a pathovar; **B,** Leaf spots symptoms developing on *Corylus avellana* cv. Cosford after syringe infiltration with *Xanthomonas arboricola* pv. *corylina* Xac 301 and kept for 10 weeks in the greenhouse under natural daylight conditions. Typical symptoms were already observed after five weeks post-inoculation.

190x275mm (300 x 300 DPI)