

The complete genome sequence of *Xanthomonas theicola*, the causal agent of canker on tea plants, reveals novel secretion systems in clade-1 xanthomonads

Ralf Koebnik,¹ Daiva Burokiene,² Claude Bragard,³ Christine Chang,⁴ Marion Fischer-Le Saux,⁵ Roland Kölliker,⁶ Jillian M. Lang,⁷ Jan E. Leach,⁷ Emily K. Luna,⁷ Perrine Portier,⁵ Angeliki Sagia,^{1,8} Janet Ziegler,⁴ Stephen P. Cohen,⁹ and Jonathan M. Jacobs^{9,10}

¹ IRD, Cirad, Université Montpellier, IPME, Montpellier, France

² Nature Research Centre, Institute of Botany, Laboratory of Plant Pathology, Akademijos g. 2, Vilnius, Lithuania

³ Earth & Life Institute, Université Catholique Louvain-la-Neuve, Louvain-la-Neuve, Belgium

⁴ Pacific Biosciences, Menlo Park, CA 94025, U.S.A.

⁵ IRHS-UMR 1345, Université d'Angers, INRAE, Institut Agro, SFR 4207 QuaSaV, 49071, CIRM-CFBP, Beaucouzé, France

⁶ Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zürich, Universitätsstrasse 2, 8092 Zürich, Switzerland

⁷ Department of Agricultural Biology, Colorado State University, Fort Collins, CO 80523, U.S.A.

⁸ Department of Biology, University of Crete, Heraklion, Greece

⁹ Department of Plant Pathology, The Ohio State University, Columbus, OH 43210, U.S.A.

¹⁰ Infectious Disease Institute, The Ohio State University, Columbus, OH 43210, U.S.A.

Corresponding author: Ralf Koebnik, koebnik@gmx.de

ABSTRACT

Xanthomonas theicola is the causal agent of bacterial canker on tea plants. There is no complete genome sequence available for *X. theicola*, a close relative of the species *X. translucens* and *X. hyacinthi*, thus limiting basic research for this group of pathogens. Here we release a high-quality complete genome sequence for the *X. theicola* type strain, CFBP 4691^T. Single-molecule real-time sequencing with a mean coverage of 264X revealed two contigs of 4,744,641 bp (chromosome) and 40,955 bp (plasmid) in size. Genome mining revealed the presence of non-ribosomal peptide synthases, two CRISPR systems, the Xps type 2 secretion system, and the Hrp type 3 secretion system. Surprisingly, this strain encodes an additional type 2 secretion system and a novel type 3 secretion system with enigmatic function, hitherto undescribed for xanthomonads. Four type 3 effector genes were found on complete or partial transposons, suggesting a role of transposons in effector gene evolution and spread. This genome sequence fills an important gap to better understand the biology and evolution of the early-branching xanthomonads, also known as clade-1 xanthomonads.

KEYWORDS: Bacterial canker, Tea plant, *Xanthomonas theicola*, Evolution, Effector-encoding transposon, Type 2 secretion, Type 3 secretion

Members of the genus *Xanthomonas*, including *X. theicola*, cause diseases on hundreds of host plants, including economically important crops and ornamental plants. The list of prokaryotic names with standing in nomenclature (LPSN; www.bacterio.net) includes more than 30 different

Xanthomonas species, which can be classified into two major groups based on sequence comparisons (Hauben et al. 1997; Gonçalves and Rosato 2002; Parkinson et al. 2007, 2009; Ferreira-Tonin et al. 2012). Clade 1, also known as the clade of early-branching species (Parkinson et al. 2007), can be further divided into two subclades (Figure 1). Subclade 1A contains *X. hyacinthi*, *X. theicola* and *X. translucens*. Subclade 1B contains *X. albilineans*, *X. sacchari* and the proposed species “*X. pseudalbilineans*” (Pierretti et al. 2015), which are all pathogenic on sugarcane, and the proposed, rice-associated species “*X. sontii*” (Bansal et al. 2019a, 2019b; Mirghasempour et al. 2020). At least three more species fall into clade 1 but await formal description (Parkinson et al. 2009; Khenfous-Djebari et al. 2019).

Tea is an important commodity consumed by more than 3 billion people worldwide and has tremendous socioeconomic, cultural and medical importance. The tea plant (*Camellia sinensis* [L.] O. Kuntze) is native to the Indian Subcontinent and to East and Southeast Asia and is cultivated in tropical and subtropical regions. The genomes of the two most widely used varieties, *C. sinensis* var. *assamica* and *C. sinensis* var. *sinensis*, were released in 2017 and 2018 (Xia et al. 2017; Wei et al. 2018). As holds true for virtually all domesticated plants, bacterial diseases threaten tea plants and production. Multiple bacterial species have been reported to cause infections of tea, such as *Agrobacterium tumefaciens* (Bradbury 1986), *Mixta theicola* (Kato Tanaka et al. 2015; Palmer et al. 2018), *Pseudomonas avellanae* (syn. *Pseudomonas syringae* pv. *theae*) (Gardan et al. 1999) and *Xanthomonas theicola* (Uehara et al. 1980).

X. theicola Vauterin et al. 1995 (formerly *X. campestris* pv. *theicola* Uehara, Arai, Nonaka & Sano 1980) was first described by Uehara and colleagues in 1980 as the causal agent of bacterial canker on tea plants (Uehara et al. 1980; Vauterin et al. 1995). Only six isolates, which were isolated in Chinan-cho, Japan, from *C. sinensis* in June 1974, have been deposited in international strain collections. To expand information on clade 1 xanthomonads, we determined the complete genome sequence of the type strain TC1 (ATCC 700184^T, CFBP 4691^T, DSM 18797^T, ICMP 6774^T, LMG 8684^T).

Genome sequencing and phylogenetic relationships

From lyophilized stock available at CIRM-CFBP (https://www6.inrae.fr/cirm_eng/CFBP-Plant-Associated-Bacteria), a single colony of strain CFBP 4691 was isolated and grown on rich nutrient agar media. Genomic DNA was extracted using the Genomic DNA buffer set and Genomic-tips following the manufacturer's instructions (Qiagen, Valencia, CA, U.S.A.). DNA was sheared three times with a Covaris g-TUBE at 5,500 rpm for 2 min to 8- to 30-kb fragments and sequenced using long-read single-molecule real-time sequencing on a PacBio Sequel I. The genome was assembled with HGAP v4 (Pacific Biosciences, Menlo Park, CA, U.S.A.). QUAST 5.0.2 was used to assess genome quality (Gurevich et al. 2013). NCBI Prokaryotic Genome Annotation Pipeline version 4.9 was used for functional annotation of genes (Haft et al. 2018).

The *X. theicola* strain CFBP 4691 genome project is available at NCBI as BioProject PRJNA606814 and the biological sample is described at NCBI as BioSample SAMN14114378. Mean coverage for the CFBP 4691 genome was 264X. The genome was assembled into two contigs of 4,744,641 and

40,955 bp in size - a chromosome and a plasmid, respectively - with a total G+C content of 66.0%. The N50 and L50 were 4,744,641 and 1, respectively. Of the predicted 4,492 genes, 3,766 encoded proteins, 61 encoded functional RNAs, and 665 were considered pseudogenes.

Genome-wide comparisons with other clade-1 xanthomonads by calculating pairwise average nucleotide identities (<http://enve-omics.ce.gatech.edu>; Rodriguez-R and Konstantinidis 2016) confirmed the unique standing of strain CFBP 4691 within this clade (Supplemental Table 1). *X. theicola* is considered a sister species of *X. hyacinthi* and belongs together with *X. translucens* to subclade 1A, which is distinct to subclade 1B, which comprises *X. albilineans*, “*X. pseudalbilineans*”, *X. sacchari* and “*X. sontii*” (Hauben et al. 1997; Gonçalves and Rosato 2002; Parkinson et al. 2007, 2009; Bansal et al. 2019a; Khenfous-Djebari et al. 2019; Mirghasempour et al. 2020). To position all the clade-1 strains in the context of Parkinson’s comprehensive *gyrB*-based species evaluation, we also included three still undefined species-level clades (Slc-5 to -7) in the phylogenetic comparison for which genomic information is not available yet (Parkinson et al. 2009) Figure 1).

Type 3 secretion system and HrpX regulon

Members of both subclades have been found to be remarkably different in their genetic makeup. For instance, type 3 secretion systems (T3SS) have been described as key virulence factors for most strains of xanthomonads but in clade 1 they are restricted to subclade 1A. We therefore scrutinized the genomes of sequenced strains, preferentially type and pathotype strains, for the presence of this and other gene clusters (Figure 1).

As the other subclade-1A strains, strain CFBP 4691 has a noncanonical T3SS (Wichmann et al. 2013; Merda et al. 2017), including its alternative translocon component HpaT (Pesce et al. 2017). Type 3 effector (T3E) prediction using described T3Es from *Xanthomonas* (<http://www.xanthomonas.org>; White et al. 2009), *Pseudomonas syringae* sensu lato (<http://www.pseudomonas-syringae.org>; Lindeberg et al. 2009) and the *Ralstonia solanacearum* species complex (<https://iant.toulouse.inra.fr/bacteria/annotation/site/prj/T3Ev3/>; Peeters et al. 2013) identified members of 12 Xop classes (Table 1), most of which are widely conserved among subclade-1A strains. Similarly, strain CFBP 4691 shares the candidate T3E HgiB with most *X. translucens* strains (Pesce et al. 2017). Most subclade-1A strains, except *X. translucens* pv. *graminis*, encode transcription activator-like effectors (TALEs). Strain CFBP 4691 encodes two TALEs, which were assigned to novel classes by AnnoTALE and designated as TalHR1 and TalHS1 (Grau et al. 2016).

Beyond clade-2 xanthomonads' T3Es, we identified one candidate T3E with similarity to a hypothetical effector from *R. solanacearum* (T3E_Hyp7) and another one with weak similarity to HopO, an ADP-ribosyl transferase from *P. syringae*. Bidirectional BLAST analyses support the similarity between HopO1 and G4Q83_14055 although the *X. theicola* sequence is much longer (968 aa) than the *P. syringae* sequences (between 283 and 301 aa). The homology is limited to the C-terminal region of G4Q83_14055 (29% sequence identity), which contains a predicted ADP-ribosyltransferase domain (<http://pfam.xfam.org/>), as found in HopO1. Among all the analyzed subclade-1A strains, three candidate T3Es appear to be restricted to the species *X. theicola*, which

include XopD, T3E_Hyp7 and the HopO-related protein. In the absence of biochemical or genetic support that the T3E_Hyp7- and HopO-related proteins are indeed we prefer not to assign them a name according to the Xop nomenclature (White et al. 2009).

Seven candidate T3E genes contain a canonical plant-induced promoter (PIP) box and a properly spaced -10 motif, which are the hallmarks of HrpX-induced genes (Koebnik et al. 2006), and are therefore likely to be co-expressed with the T3SS: *xopB*, *xopE2*, *xopG*, two paralogs of *xopAD*, *xopAF* and T3E_Hyp7. In addition to these T3E and the T3SS genes, a gene coding for a secreted serine protease that is conserved in clade-2 xanthomonads and in *Clavibacter michiganensis* was found to possess a PIP box/-10 motif. A putative role of this gene in pathogenicity awaits experimental proof.

Effector-encoding transposons

While annotating T3E genes we realized that strain CFBP 4691 has two identical copies of the *xopG* gene. We first considered this as an artifact produced during the genome assembly process. However, careful analysis revealed that the *xopG* gene is part of a 6781-bp sequence that is duplicated in the chromosome and shows features of a transposon, such as 28-bp inverted repeats at the ends, flanked by identical 5-bp motifs resulting from target site duplication, *tnpA*, *tnpS* and *tnpT* genes with a resolution recombination site between the *tnpS* and *tnpT* genes (Figure 2). A structurally similar transposon of the Tn3 family has been described for strains of *X. citri* and other clade-2 xanthomonads (Ferreira et al. 2015). Two related sequences, encompassing one copy of the inverted repeat and the full or partial *tnpA* gene, are present on

the plasmid. Interestingly, both of them are linked to a T3E gene (*xopAF* and *xopE2*) (Figure 2). This finding suggests that *xopG*, *xopAF* and *xopE3* were once acquired and that *xopG* may be duplicated further by transposition events.

Other protein secretion systems

In addition to the Hrp T3SS, this strain does encode a second, hitherto undescribed T3SS for xanthomonads the function of which remains enigmatic. The corresponding gene cluster contains 26 predicted genes (loci G4Q83_11445 to G4Q83_11570), including all the conserved genes of such a secretion system (eleven *sct* genes) and genes encoding components of the needle and the translocon complex and an AraC-type regulator. None of the genes contains an early stop or frameshift mutation, indicating that the system is most likely functional. Similar T3SSs are found in soil bacteria of the species *Glaciimonas immobilis* and *Solimicrobium silvestre* and in several strains of *Herbaspirillum* (all from the family *Oxalobacteraceae* in the order *Burkholderiales*) (GenBank accession numbers: CP022736, AWRY01, JACHHQ01, VFNG01, VFNI01, VIUO01) (Reinhold-Hurek and Hurek 1998; Zhang et al. 2011; Margesin et al. 2018).

The universally conserved *xps* type 2 secretion system (T2SS) was present in all of the strains, but none of them harbor the *xcs* T2SS, which is found in several clade-2 species, such as *X. campestris*, *X. citri* and *X. euvesicatoria* (Lu et al. 2008). Notably, the *X. hyacinthi* strain CFBP 1156 encodes a third T2SS, which among all xanthomonads sequenced to date is only present in the *X. arboricola* pv. *populi* strain CFBP 3122 (BioProject PRJNA339373). Interestingly, this is an atypical strain lacking a T3SS (Merda et al. 2017). Beyond *Xanthomonas*, similar T2SSs are found in

environmental strains of *Rhodanobacter* and *Pseudoxanthomonas* (both from the order *Xanthomonadales*) (NCBI WGS accession numbers: AJXW01, LMHR01, LVJS01, MEGI01, OCND01, PDWP01, PDWU01) (Weon et al. 2006; Lee et al. 2007; Li et al. 2014).

Other gene clusters

Type 6 secretion systems (T6SS) can play an important role in pathogenicity, microbial competition and cell-to-cell communication (Chassaing and Cascales 2018; Coulthurst 2019). Genes encoding T6SS components were identified and classified using SecReT6 (Li et al. 2015). Interestingly, this secretion system is restricted to some pathovars of *X. translucens* (pvs. *cerealis*, *translucens*) but absent in other clade-1 species and pathovars (Figure 1).

Members of the genus *Xanthomonas* were found to encode two clustered regularly interspaced short palindromic repeats (CRISPR) systems, IC and IF, which are defense systems against alien DNA/RNA (Makarova et al., 2011). We used CrisprCasFinder to search for CRISPR systems in selected clade-1 xanthomonads (Couvin et al. 2018). Most strains of *Xanthomonas* encode at most one system, but some strains from the species *X. hyacinthi*, *X. pseudalbilineans* and *X. albilineans* encode both systems (Pieretti et al. 2015; Cohen et al. 2020; Zhang et al. 2020). Likewise, both CRISPR systems were found in strain CFBP 4691, which supports an evolutionary scenario with an ancestor of *Xanthomonas* that contained both systems.

Within the genus *Xanthomonas*, non-ribosomal peptide synthases (NRPS) were first described for *X. albilineans*, where they synthesize several small peptides, among them the antibiotic albicidin

(Royer et al. 2013). A key step in non-ribosomal peptide biosynthesis is the transfer of the P-pant moiety from coenzyme A to a serine residue, which is catalyzed by the 4'-phosphopantetheinyl transferase (PPTase). We used the presence of both a PPTase gene and NRPS gene(s) as a marker for the ability of a strain to synthesize peptides non-ribosomally (Figure 1). This analysis revealed that NRPS are widespread among clade-1 strains with the exception of *X. hyacinthi*, which did not encode any NRPS.

Conclusion

The first complete genome sequence for the species *X. theicola* will stimulate further work on the underexplored clade-1 of xanthomonads. Notably, this work revealed the presence of two protein secretion systems that have not been described yet for the genus *Xanthomonas*, a rare T2SS and a novel T3SS. It will be interesting to study their contribution to pathogenicity. Scrutiny of the predicted proteomes revealed some not yet investigated T3Es in *Xanthomonas* and the presence of effector-encoding transposons in clade-1 xanthomonads. Another interesting discovery was the presence of two new TALEs. The availability of genome sequences of the host plant will be instrumental to elucidate the TALEs' contribution, and the contribution of other secreted proteins, to canker formation on tea plants.

Acknowledgements

We are grateful to Jan Grau, Martin Luther University Halle (Saale), Germany, for assistance with predicting and classifying TALE genes. SPC is supported by a United States Department of Agriculture National Institute of Food and Agriculture Postdoctoral Fellowship (2018-08122). This

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

References

1. Bansal, K., Kaur, A., Midha, S., Kumar, S., Korpole, S., and Patil, P. B. 2019a. *Xanthomonas sontii* sp. nov., a non-pathogenic bacterium isolated from healthy basmati rice (*Oryza sativa*) seeds from India. bioRxiv 738047; <https://doi.org/10.1101/738047>.
2. Bansal, K., Midha, S., Kumar, S., Kaur, A., Sonti, R. V., and Patil, P. B. 2019b. Ecological and evolutionary insights into pathogenic and non-pathogenic rice associated *Xanthomonas*. bioRxiv 453373v2; <https://doi.org/10.1101/453373>.
3. Bradbury, J.F. 1986. Guide to plant pathogenic bacteria. CAB International Farnham Royal, Slough UK, p.12 (ISBN 0851985572).
4. Chassaing, B., and Cascales, E. 2018. Antibacterial weapons: targeted destruction in the microbiota. Trends Microbiol. 26:329-338.
5. Cohen, S. P., Luna, E. K., Lang, J. M., Ziegler, J., Chang, C., Leach, J. E., Fischer Le-Saux, M., Portier, P., Koebnik, R., and Jacobs, J. M. 2020. High-quality genome resource of *Xanthomonas hyacinthi* generated via long-read sequencing. Plant Disease 104:1011-1012.
6. Coulthurst, S. 2019. The Type VI secretion system: a versatile bacterial weapon. Microbiology 165:503-515.
7. Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E. P. C., Vergnaud, G., Gautheret, D., and Pourcel, C. 2018. CRISPRCasFinder, an update of

- CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46:W246-W251.
8. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., and Gascuel, O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36:W465-W469.
 9. Falahi Charkhabi, N., Booher, N. J., Peng, Z., Wang, L., Rahimian, H., Shams-Bakhsh, M., Liu, Z., Liu, S., White, F. F., and Bogdanove, A. J. 2017. Complete genome sequencing and targeted mutagenesis reveal virulence contributions of Tal2 and Tal4b of *Xanthomonas translucens* pv. *undulosa* ICMP11055 in bacterial leaf streak of wheat. *Front. Microbiol.* 8:1488.
 10. Ferreira, R. M., de Oliveira, A. C., Moreira, L. M., Belasque Jr, J., Gourbeyre, E., Siguier, P., Ferro, M. I., Ferro, J. A., Chandler, M., and Varani, A. M. 2015. A TALE of transposition: Tn3-like transposons play a major role in the spread of pathogenicity determinants of *Xanthomonas citri* and other xanthomonads. *mBio* 6:e02505-14.
 11. Ferreira-Tonin, M., Rodrigues-Neto, J., Harakava, R., and Destéfano, S. A. 2012. Phylogenetic analysis of *Xanthomonas* based on partial *rpoB* gene sequences and species differentiation by PCR-RFLP. *Int. J. Syst. Evol. Microbiol.* 62:1419-1424.
 12. Gardan, L., Shafik, H., Belouin, S., Broch, R., Grimont, F., and Grimont, P. A. 1999. DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremae* sp. nov. and *Pseudomonas cannabina* sp. nov. (ex Sutic and Dowson 1959). *Int. J. Syst. Bacteriol.* 49:469-478.
 13. Gonçalves, E. R., and Rosato, Y. B. 2002. Phylogenetic analysis of *Xanthomonas* species based upon 16S-23S rDNA intergenic spacer sequences. *Int. J. Syst. Evol. Microbiol.* 52:355-361.

14. Grau, J., Reschke, M., Erkes, A., Streubel, J., Morgan, R. D., Wilson, G. G., Koebnik, R., and Boch, J. 2016. AnnoTALE: bioinformatics tools for identification, annotation, and nomenclature of TALEs from *Xanthomonas* genomic sequences. *Sci. Rep.* 6:21077.
15. Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. 2013. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29:1072-1075.
16. Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., Li, W., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Zheng, C., Thibaud-Nissen, F., Geer, L. Y., Marchler-Bauer, A., and Pruitt, K. D. 2018. RefSeq: An update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* 46:D851-D860.
17. Hauben, L., Vauterin, L., Swings, J., and Moore, E. R. 1997. Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *Int. J. Syst. Bacteriol.* 47:328-335.
18. Hersemann, L., Wibberg, D., Widmer, F., Vorhölter, F. J., and Kölliker, R. 2016. Draft genome sequences of three *Xanthomonas translucens* pathovar reference strains (pv. *arrhenatheri*, pv. *poae* and pv. *phlei*) with different specificities for forage grasses. *Stand. Genomic Sci.* 11:50.
19. Jaenicke, S., Bunk, B., Wibberg, D., Spröer, C., Hersemann, L., Blom, J., Winkler, A., Schatschneider, S., Albaum, S. P., Kölliker, R., Goesmann, A., Pühler, A., Overmann, J., and Vorhölter, F. J. 2016. Complete genome sequence of the barley pathogen *Xanthomonas translucens* pv. *translucens* DSM 18974^T (ATCC 19319^T). *Genome Announc.* 4:e01334-16.
20. Kato Tanaka, Y., Horie, N., Mochida, K., Yoshida, Y., Okugawa, E., and Nanjo, F. 2015. *Pantoea theicola* sp. nov., isolated from black tea. *Int. J. Syst. Evol. Microbiol.* 65:3313-3319.

21. Khenfous-Djebari, B., Kerkoud, M., Fischer-Le Saux, M., Benhassine, T., Bouznad, Z., Koebnik, R., and Bragard, C. 2019. Isolation of atypical wheat-associated xanthomonads in Algeria. *Phytopathologia Mediterranea* 58:497-506.
22. Koebnik, R., Krüger, A., Thieme, F., Urban, A., and Bonas, U. 2006. Specific binding of the *Xanthomonas campestris* pv. *vesicatoria* AraC-type transcriptional activator HrpX to plant-inducible promoter boxes. *J. Bacteriol.* 188:7652-7660.
23. Lee, C. S., Kim, K. K., Aslam, Z., and Lee, S. T. 2007. *Rhodanobacter thiooxydans* sp. nov., isolated from a biofilm on sulfur particles used in an autotrophic denitrification process. *Int. J. Syst. Evol. Microbiol.* 57:1775-1779.
24. Letunic, I., and Bork, P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47:W256-W259.
25. Li, D., Pang, H., Sun, L., Fan, J., Li, Y., and Zhang, J. 2014. *Pseudoxanthomonas wuyuanensis* sp. nov., isolated from saline-alkali soil. *Int. J. Syst. Evol. Microbiol.* 64:799-804.
26. Li, J., Yao, Y., Xu, H. H., Hao, L., Deng, Z., Rajakumar, K., and Ou, H. Y. 2015. SecReT6: a web-based resource for type VI secretion systems found in bacteria. *Environ. Microbiol.* 17:2196-2202.
27. Lindeberg, M., Biehl, B. S., Glasner, J. D., Perna, N. T., Collmer, A., and Collmer, C. W. 2009. Gene Ontology annotation highlights shared and divergent pathogenic strategies of type III effector proteins deployed by the plant pathogen *Pseudomonas syringae* pv. *tomato* DC3000 and animal pathogenic *Escherichia coli* strains. *BMC Microbiol.* 9:S4.
28. Lu, H., Patil, P., Van Sluys, M. A., White, F. F., Ryan, R. P., Dow, J. M., Rabinowicz, P., Salzberg, S. L., Leach, J. E., Sonti, R., Brendel, V., and Bogdanove, A.J. 2008. Acquisition and evolution

- of plant pathogenesis-associated gene clusters and candidate determinants of tissue-specificity in *Xanthomonas*. PLoS One 3:e3828.
29. Makarova, K. S., Aravind, L., Wolf, Y. I., and Koonin, E. V. 2011. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biol. Direct 6:38.
 30. Margesin, R., Albuquerque, L., Zhang, D. C., Froufe, H. J. C., Severino, R., Roxo, I., Egas, C., and da Costa, M. S. 2018. *Solimicrobium silvestre* gen. nov., sp. nov., isolated from alpine forest soil. Int. J. Syst. Evol. Microbiol. 68:2491-2498.
 31. Merda, D., Briand, M., Bosis, E., Rousseau, C., Portier, P., Barret, M., Jacques, M. A., and Fischer-Le Saux, M. 2017. Ancestral acquisitions, gene flow and multiple evolutionary trajectories of the type three secretion system and effectors in *Xanthomonas* plant pathogens. Mol. Ecol. 26:5939-5952.
 32. Mirghasempour, S. A., Huang, S., Studholme, D. J., and Brady, C. L. 2020. A grain rot of rice in Iran caused by a *Xanthomonas* strain closely related to *X. sacchari*. Plant Dis. 104:1581-1583.
 33. Palmer, M., Steenkamp, E. T., Coetzee, M. P. A., Avontuur, J. R., Chan, W. Y., van Zyl, E., Blom, J., and Venter, S. N. 2018. *Mixta* gen. nov., a new genus in the *Erwiniaceae*. Int. J. Syst. Evol. Microbiol. 68:1396-1407.
 34. Parkinson, N., Aritua, V., Heeney, J., Cowie, C., Bew, J., and Stead, D. 2007. Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. Int. J. Syst. Evol. Microbiol. 57:2881-2887.

35. Parkinson, N., Cowie, C., Heeney, J., and Stead, D. 2009. Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *Int. J. Syst. Evol. Microbiol.* 59:264-274.
36. Peeters, N., Carrère, S., Anisimova, M., Plener, L., Cazalé, A. C., and Genin, S. 2013. Repertoire, unified nomenclature and evolution of the Type III effector gene set in the *Ralstonia solanacearum* species complex. *BMC Genomics* 14:859.
37. Pesce, C., Bolot, S., Berthelot, E., Bragard, C., Cunnac, S., Fischer-Le Saux, M., Portier, P., Arlat, M., Gagnevin, L., Jacques, M. A., Noël, L. D., Carrère, S., and Koebnik, R. 2015a. Draft genome sequence of *Xanthomonas translucens* pv. *graminis* pathotype strain CFBP 2053. *Genome Announc.* 3:e01174-15.
38. Pesce, C., Bolot, S., Cunnac, S., Portier, P., Fischer-Le Saux, M., Jacques, M. A., Gagnevin, L., Arlat, M., Noël, L. D., Carrère, S., Bragard, C., and Koebnik, R. 2015b. High-quality draft genome sequence of the *Xanthomonas translucens* pv. *cerealis* pathotype strain CFBP 2541. *Genome Announc.* 3:e01574-14.
39. Pesce, C., Jacobs, J. M., Berthelot, E., Perret, M., Vancheva, T., Bragard, C., and Koebnik, R. 2017. Comparative genomics identifies a novel conserved protein, HpaT, in proteobacterial type III secretion systems that do not possess the putative translocon protein HrpF. *Front. Microbiol.* 8:1177.
40. Pieretti, I., Cociancich, S., Bolot, S., Carrère, S., Morisset, A., Rott, P., and Royer, M. 2015. Full genome sequence analysis of two isolates reveals a novel *Xanthomonas* species close to the sugarcane pathogen *Xanthomonas albilineans*. *Genes (Basel)* 6:714-733.

41. Reinhold-Hurek, B., and Hurek, T. 1998. Life in grasses: diazotrophic endophytes. *Trends Microbiol.* 6:139-144.
42. Rodriguez-R, L. M., and Konstantinidis, K. T. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Preprints* 4:e1900v1.
43. Royer, M., Koebnik, R., Marguerettaz, M., Barbe, V., Robin, G. P., Brin, C., Carrere, S., Gomez, C., Hügelland, M., Völler, G. H., Noëll, J., Pieretti, I., Rausch, S., Verdier, V., Poussier, S., Rott, P., Süssmuth, R. D., and Cociancich, S. 2013. Genome mining reveals the genus *Xanthomonas* to be a promising reservoir for new bioactive non-ribosomally synthesized peptides. *BMC Genomics* 14:658.
44. Uehara, K., Arai, K., Nonaka, T., and Sano, I. 1980. Canker of tea, a new disease, and its causal bacterium *Xanthomonas campestris* pv. *theaecola* Uehara et Arai pv. nov. *Bull. Faculty Agric.* 30:17-21.
45. Vauterin, L., Hoste, B., Kersters, K., and Swings, J. 1995. Reclassification of *Xanthomonas*. *Int. J. Syst. Bacteriol.* 45:472-489.
46. Wei, C., Yang, H., Wang, S., Zhao, J., Liu, C., Gao, L., Xia, E., Lu, Y., Tai, Y., She, G., Sun, J., Cao, H., Tong, W., Gao, Q., Li, Y., Deng, W., Jiang, X., Wang, W., Chen, Q., Zhang, S., Li, H., Wu, J., Wang, P., Li, P., Shi, C., Zheng, F., Jian, J., Huang, B., Shan, D., Shi, M., Fang, C., Yue, Y., Li, F., Li, D., Wei, S., Han, B., Jiang, C., Yin, Y., Xia, T., Zhang, Z., Bennetzen, J. L., Zhao, S., and Wan, X. 2018. Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality. *Proc. Natl. Acad. Sci. U. S. A.* 115:E4151-E4158.

47. Weon, H. Y., Kim, B. Y., Kim, J. S., Lee, S. Y., Cho, Y. H., Go, S. J., Hong, S. B., Im, W. T., and Kwon, S. W. 2006. *Pseudoxanthomonas suwonensis* sp. nov., isolated from cotton waste composts. Int. J. Syst. Evol. Microbiol. 56:659-662.
48. White, F. F., Potnis, N., Jones, J. B., and Koebnik, R. 2009. The type III effectors of *Xanthomonas*. Mol. Plant Pathol. 10:749-766.
49. Wichmann, F., Vorhölter, F. J., Hersemann, L., Widmer, F., Blom, J., Niehaus, K., Reinhard, S., Conradin, C., and Kölliker, R. 2013. The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. Mol. Plant Pathol. 14:576-88.
50. Xia, E. H., Zhang, H. B., Sheng, J., Li, K., Zhang, Q. J., Kim, C., Zhang, Y., Liu, Y., Zhu, T., Li, W., Huang, H., Tong, Y., Nan, H., Shi, C., Shi, C., Jiang, J. J., Mao, S. Y., Jiao, J. Y., Zhang, D., Zhao, Y., Zhao, Y. J., Zhang, L. P., Liu, Y. L., Liu, B. Y., Yu, Y., Shao, S. F., Ni, D. J., Eichler, E. E., and Gao, L. Z. 2017. The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. Mol. Plant 10:866-877.
51. Zhang, H. L., Ntambo, M. S., Rott, P. C., Chen, G., Chen, L. L., Huang, M. T., and Gao, S. J. 2020. Complete genome sequence reveals evolutionary and comparative genomic features of *Xanthomonas albilineans* causing sugarcane leaf scald. Microorganisms 8:E182.
52. Zhang, D. C., Redzic, M., Schinner, F., and Margesin, R. 2011. *Glaciimonas immobilis* gen. nov., sp. nov., a member of the family *Oxalobacteraceae* isolated from alpine glacier cryoconite. Int. J. Syst. Evol. Microbiol. 61: 2186-2190.

Table 1. Type 3 effectors and avirulence proteins in *Xanthomonas theicola* and their conservation in type and pathotype strains of subclade 1A.

T3E class	Locus in strain CFBP 4691	Size (aa)	PIP/ -10	XT	XH	XTA	XTPo	XTPI	XTG	XTC	XTT	XTU
				CFBP 4691	CFBP 1156	LMG 727	LMG 728	LMG 730	CFBP 2053	CFBP 2541	DSM 18974	ICMP 11055
TALE	G4Q83_03320	1318		+	+	+	+	+	-	+	+	+
TALE	G4Q83_02340	1142		+	+	+	+	+	-	+	+	+
XopB	G4Q83_12275	610	YES	+	+	+	+	+	+	+	+	+
XopD	G4Q83_12480	906		+	-	-	-	-	-	-	-	-
XopE	G4Q83_22315	358	YES	+	-	-	-	+	+	+	-	+
XopF	G4Q83_01280	691		+	+	+	+	+	+	+	+	+
XopG	G4Q83_07020, G4Q83_08310	215	YES	+	+	+	+	+	+	+	+	+
XopL	G4Q83_12185	293		Ψ	+	+	+	+	+	+	+	+
XopM	G4Q83_01275	535		+	+	Ψ	+	Ψ	Ψ	Ψ	+	Ψ
XopX	G4Q83_01820	705		+	+	+	+	+	+	+	+	+
XopZ	G4Q83_06945	1355		+	+	+	+	+	+	+	+	+
XopAD	G4Q83_11590	2644	YES	+	-	+	+	+	+	+	+	+
XopAD	G4Q83_01520	2438	YES	+	-	+	+	+	+	+	+	+
XopAF	G4Q83_22295	282	YES	+	+	-	+	Ψ	Ψ	+	+	+
XopAZ	G4Q83_01795	159		+	+	+	+	+	+	+	+	+
HgiB	G4Q83_01270	147		+	-	Ψ	+	+	+	+	+	+
HopO1	G4Q83_14055	968		+	-	-	-	-	-	-	-	-

RsHyp7	G4Q83_19965	176	YES	+	-	-	-	-	-	-	-	-
AvrXccA1	G4Q83_06360	613		Ψ	-	-	+	+	-	+	-	-

Type and pathotype strains used for comparison of the T3E profile of *X. hyacinthi* CFBP 1156 with other subclade-1A xanthomonads: XT, *X. theicola* CFBP 4641; XT, *X. hyacinthi* CFBP 1156 (CP043476, CP043477; Cohen et al. 2020); XTA, *X. translucens* pv. *arrhenatheri* LMG 727 (CXOI01; Hersemann et al. 2016); XTPo, *X. translucens* pv. *poae* LMG 728 (CXOK01; Hersemann et al. 2016); XTPI, *X. translucens* pv. *phlei* LMG 730 (CXOJ01; Hersemann et al. 2016); XTG, *X. translucens* pv. *graminis* CFBP 2053 (LHSI01; Pesce et al. 2015a); XTC, *X. translucens* pv. *cerealis* CFBP 2541 (JWHD01; Pesce et al. 2015b); XTT, *X. translucens* pv. *translucens* DSM 18974 (LT604072; Jaenicke et al. 2016); XTU, *X. translucens* pv. *undulosa* ICMP 11055 (CP009750; Falahi Charkhabi et al. 2017).

YES, presence of PIP box and -10 motif; +, presence of T3E or avirulence protein; -, absence of T3E or Avr protein; Ψ, presence of a truncated version due to early stop codon(s), frame shift(s) or IS element insertion.

Legends to the figures and e-Xtra files

Figure 1. Presence of selected gene clusters in clade-1 xanthomonads. Phylogenetic tree constructed from partial *gyrB* sequences, either extracted from genome sequences or taken from Parkinson et al. 2009, using default settings at <http://phylogeny.fr> (Dereeper et al. 2008), and modified at <https://itol.embl.de/> for better visualisation (Letunic and Bork 2019). The blue circles indicate bootstrap values above 0.5 and their sizes are proportional to the bootstrap values. Presence of gene clusters was elucidated using BLAST, CrisprCasFinder and SecReT6. Green squares indicate presence and white squares indicate absence of a gene cluster. “Slc” stand for “species-level clade” (Parkinson et al. 2009).

Figure 2. Genetic organization of complete and partial effector-encoding transposons in *X. theicola* strain CFBP 4691. Genes are indicated by colored arrows. Transposition-related genes are shown in orange, effector genes in green, and other predicted genes in light blue. The terminal inverted repeats (IRL and IRR; sequences given in the red boxes) are shown as red triangles, and sequences of target site duplications are shown next to them. HrpX-regulated promoters, which consist of a PIP box and a properly spaced -10 element (Koebnik et al. 2006), are indicated as yellow circles. Presumed resolution sites, as indicated by the Ω symbol, are located upstream of the *tnpT* genes, and their sequences are given at the bottom of the figure, with sequence variations highlighted in red colour. Resolution sites include two palindromes: IR1 and IR2 are probably part of the core site at which recombination occurs, recognized by the TnpS recombinase, whereas IRa and IRb are potential binding sites for TnpT (Ferreira et al. 2015).

e-Xtra Supplemental Table 1. Genome-wide average nucleotide identities among reference strains of clade 1. Type, pathotype and reference strains used for the calculation of genome-wide average nucleotide identities: XT, *X. theicola* CFBP 4641; XT, *X. hyacinthi* CFBP 1156 (CP043476, CP043477; Cohen et al. 2020); XTT, *X. translucens* pv. *translucens* DSM 18974 (LT604072; Jaenicke et al. 2016); XTC, *X. translucens* pv. *cerealis* CFBP 2541 (JWHD01; Pesce et al. 2015b); XTG, *X. translucens* pv. *graminis* CFBP 2053 (LHSI01; Pesce et al. 2015a); XSa, *X. sacchari* CFBP 4641 (MDEK01); XSo, “*X. sontii*” PPL1 (NQYO01; Bansal et al. 2019a); XA, *X. albilineans* CFBP 2523 (MDCB01); and XP, “*X. pseudalbilineans*” GPE39 (JZHZ01; Pierretti et al. 2015).

