New Phytologist Supporting Information

Article title: Genome-wide identification of fitness determinants in the *Xanthomonas campestris* bacterial pathogen during early stages of plant infection
Authors: Julien S. Luneau, Maël Baudin, Thomas Quiroz-Monnens, Sébastien Carrère, Olivier Bouchez, Marie-Françoise Jardinaud, Carine Gris, Jonas François, Jayashree Ray, Babil Torralba, Matthieu Arlat, Jennifer D. Lewis, Emmanuelle Lauber, Adam M. Deutschbauer, Laurent D. Noël, Alice Boulanger
Article acceptance date: 07 June 2022

The following Supporting Information is available for this article:

Fig. S1 Biological reproducibility of RB-TnSeq assays.

Fig. S2 Growth of the RB-TnSeq library *in vitro* identifies genes contributing to multiplication in plant independent conditions.

Fig. S3 *Xanthomonas campestris pathovar campestris (Xcc)* fitness measured by RB-TnSeq for virulence-associated genes (as defined by Y-W. He et *al*., 2007) in plant-associated conditions.

Fig. S4 Correlation between RB-TnSeq fitness results, competitiveness and pathogenicity during hydathode infection with *Xanthomonas campestris pathovar campestris (Xcc)* mutant strains in genes identified by RB-TnSeq during infection of cauliflower hydathodes.

Fig. S5 *In vitro* phenotypes of the *ΔXC_3253* mutant strain.

Fig. S6 Importance of XC_3388 for *Xanthomonas campestris pathovar campestris (Xcc)* metabolism.

Table S1 Strains, plasmids and oligos used in this study

Table S2 List of essential genes of *Xanthomonas campestris pathovar campestris (Xcc)*. Essentiality of genes was determined as descibes in Price et al., 2018.

Table S3 Genes affected in fitness in each tested conditions**.** This supplementary table provides the complete list of genes affected in fitness of *Xanthomonas campestris pathovar campestris*

(Xcc) 8004 in all conditions assayed: Xylem sap, hydathodes at 6 and 3 days post-inoculation, *in vitro* and genes commonly affected in all plant conditions assayed.

Table S4 RNAseq Data. This supplementary table provides the complete list of differentially expressed genes of *Xanthomonas campestris pathovar campestris (Xcc)* 8004 in all conditions assayed.

Methods S1 Supplementary material and methods for Fitness screening *in vitro;* Carbon and nitrogen substrates phenotyping; Growth measurements *in vitro.*

Fig. S1 Biological reproducibility of RB-TnSeq assays. Hydathodes 3 days post-inoculation (dpi) **(a)**, Hydathodes 6 dpi **(b)**, Xylem sap **(c)**, MOKA rich medium **(d)**, MME Glucose **(e)** and MME Glucose + DropOut –LTH (Leucine-Tryptophan-Histidine) **(f).** Each dot represents a gene. Red dots show the genes selected according to our criteria ($|f| > 1$ & $|t\text{-score}| > 3$). The light-gray dashed line traces the y = x ideal correlation between the two replicates. Two out of four biological replicates are shown for each condition and are representative of the pairwise correlations between all the replicates.
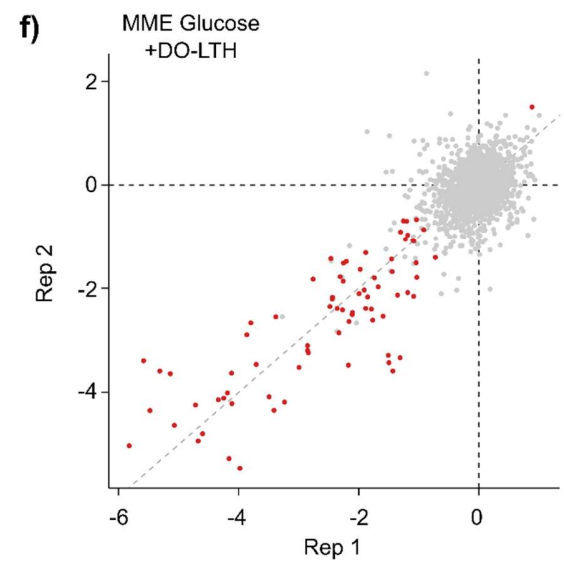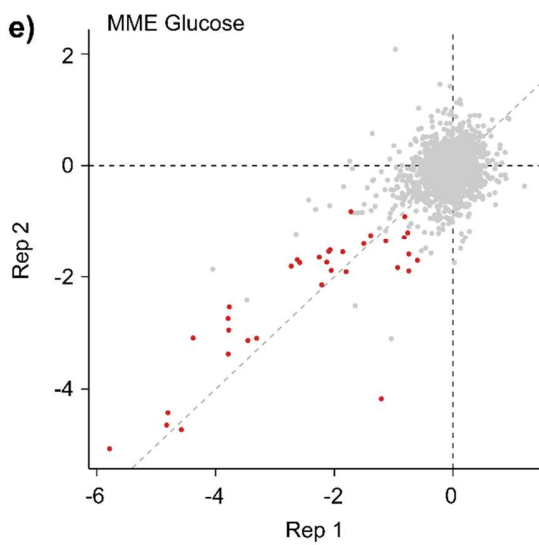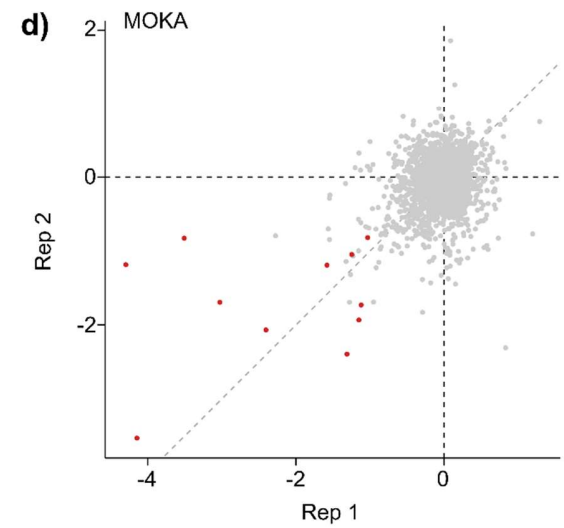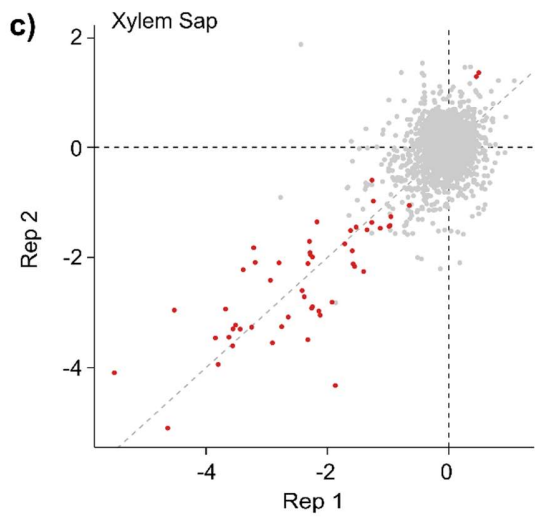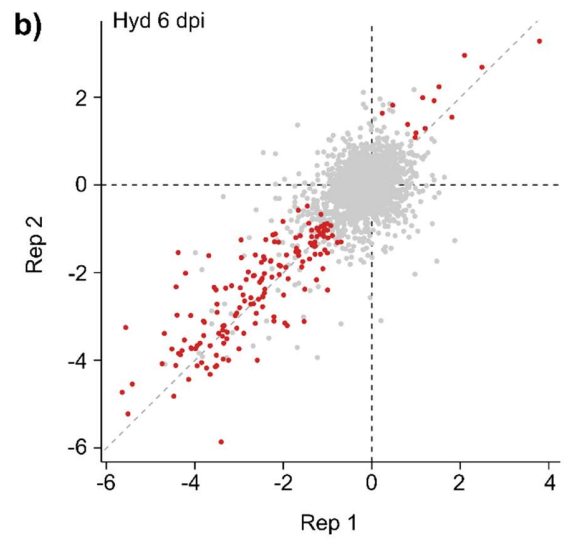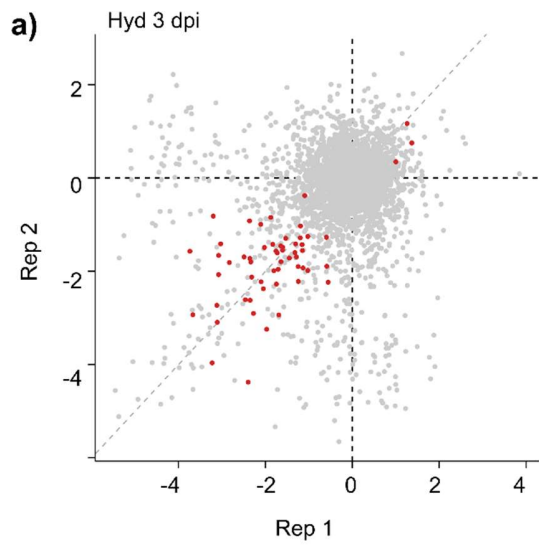
Fig. S2 Growth of the RB-TnSeq library *in vitro* identifies genes contributing to multiplication in plant independent conditions. (a) Venn Diagram showing *Xanthomonas campestris pathovar campestris (Xcc)* genes contributing to in vitro fitness which are nearessential and not specific to the plant-associated environments tested. The MOKA rich medium was used to construct the mutant library. **(b)** Genome-wide gene fitness assessment in minimal medium MMX Glucose DropOut –LTH (Leucine-Tryptophan-Histidine) medium. Genes related to the biosynthesis of leucine, histidine and tryptophan are highlighted. Horizontal black lines show the gene fitness threshold ($|f| \geq 1$) we used to consider genes affected in growth. The expected loss of fitness in these auxotrophic mutants grown in a medium lacking the three amino acids confirms the utility of our RB-TnSeq library to identify conditionally-relevant genes and validates the quality of the readout for evaluating gene fitness.
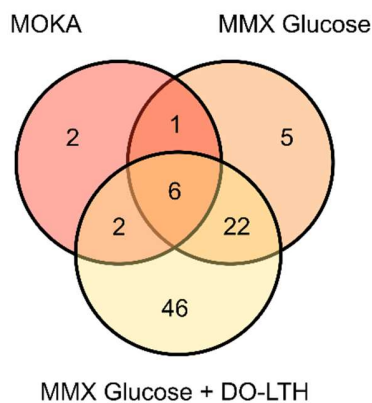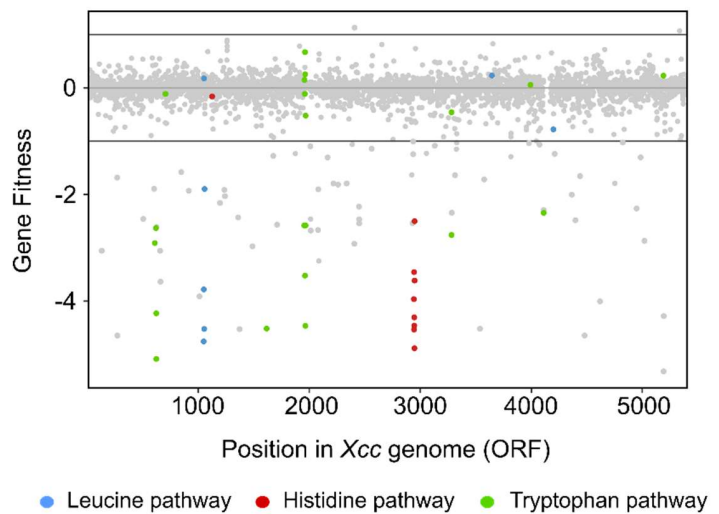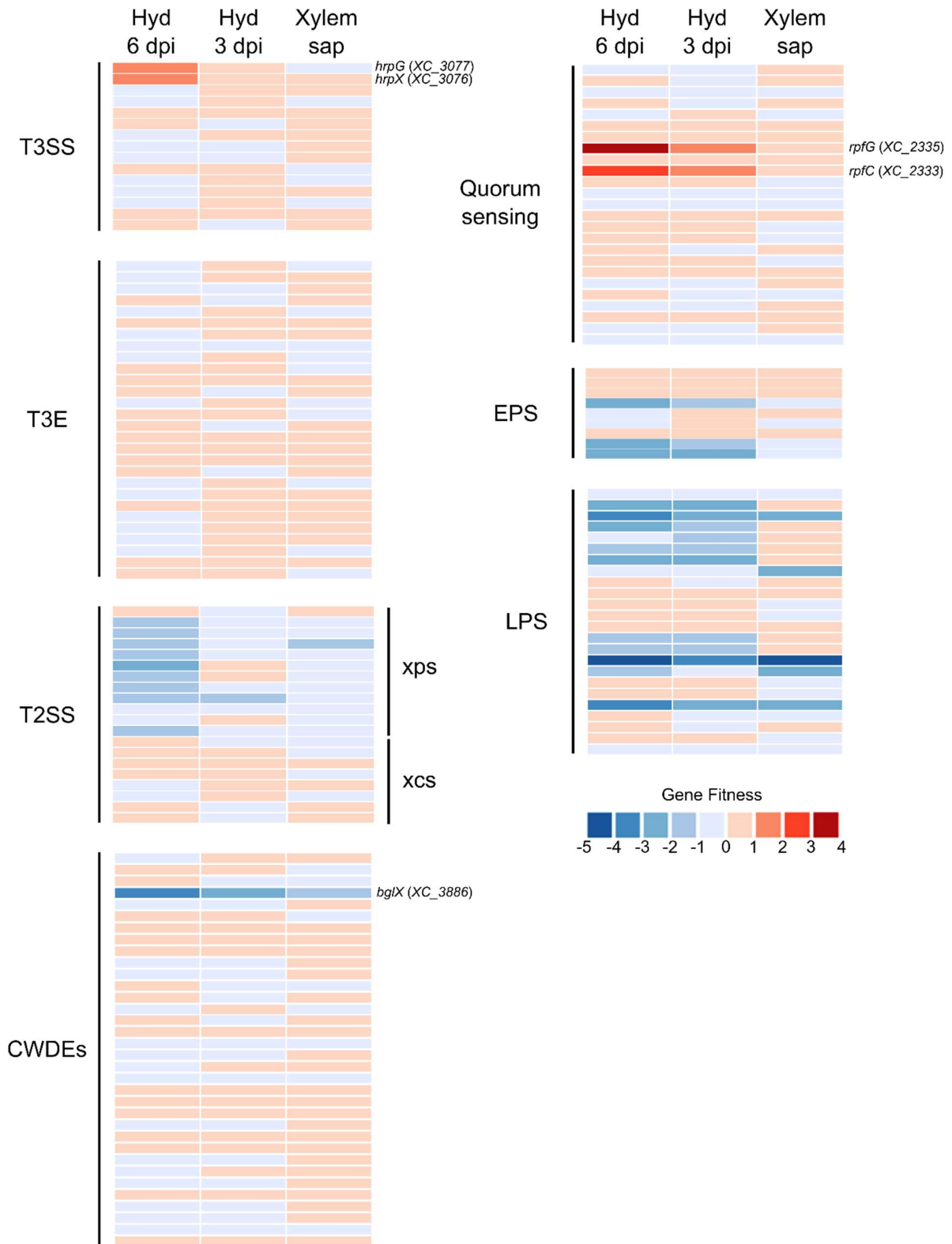
Fig. S3 *Xanthomonas campestris pathovar campestris (Xcc)* fitness measured by RB-TnSeq for virulence-associated genes (as defined by He et *al*., 2007)in plant-associated conditions**.** T3SS: Type III Secretion System; T3E: Type III Effector; T2SS: Type II Secretion System; CWDEs: Cell Wall Degrading Enzymes; EPS: Exopolysaccharide; LPS: Lipopolysaccharide; Hyd: Hydathode.

He Y-Q, Zhang L, Jiang B-L, Zhang Z-C, Xu R-Q, Tang D-J, Qin J, Jiang W, Zhang X, Liao J, *et al.* 2007. Comparative and functional genomics reveals genetic diversity and determinants of host specificity among reference strains and a large collection of Chinese isolates of the phytopathogen *Xanthomonas campestris pv. campestris*. *Genome Biology* **8**: R218.

Hyd 6 dpi | Hyd 3 dpi | Xylem sap

T3SS — hrpG (XC_3077), hrpX (XC_3076)

T3E

T2SS — xps, xcs

CWDEs — bglX (XC_3886)

Quorum sensing — rpfG (XC_2335), rpfC (XC_2333)

EPS

LPS

Gene Fitness

-5 -4 -3 -2 -1 0 1 2 3 4

Fig. S4 Correlation between RB-TnSeq fitness results, competitiveness and pathogenicity during hydathode infection with *Xanthomonas campestris pathovar campestris (Xcc)* mutant strains in genes identified by RB-TnSeq during infection of cauliflower hydathodes. (a) Correlation between gene fitness values obtained with RB-TnSeq and 1:1 competition assays, both at 6 days post-inoculation (dpi) in hydathodes. Dots represent the mean of gene fitness values for the four biological replicates of RB-TnSeq and the median of Competitive Index (CI) obtained over at least four biological replicates. **(b)** Correlation between gene fitness values obtained with RB-TnSeq at 6 dpi in hydathodes and disease symptoms scoring at 10 dpi. Dots represent the mean of gene fitness values for the four biological replicates of RB-TnSeq and the median of Disease Index obtained over three biological replicates. Correlation strength is indicated by the Spearman correlation coefficient r and the associated p-value.
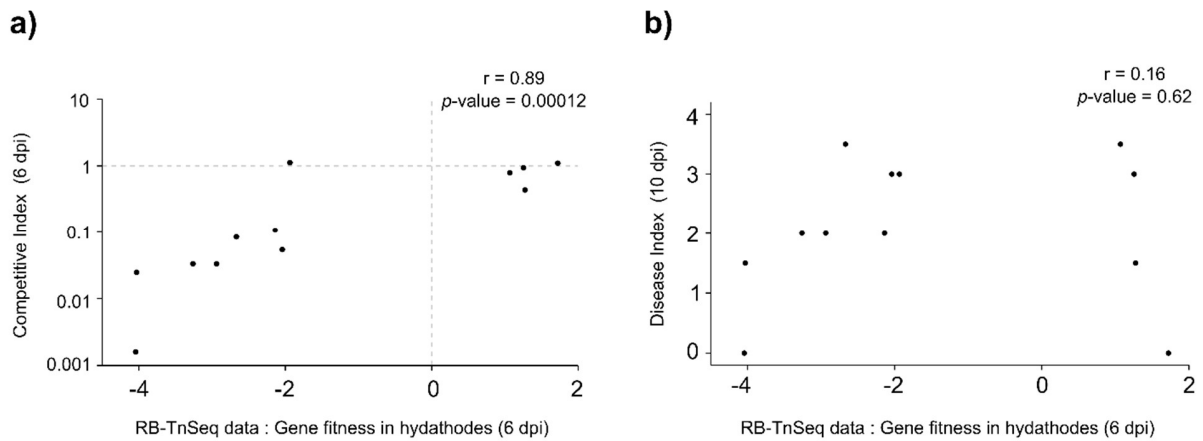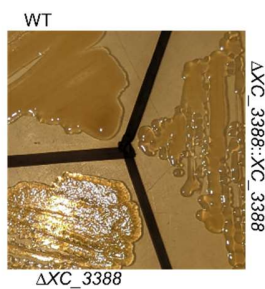
Fig. S5 *In vitro* phenotypes of the *ΔXC_3253* mutant strain. (a) Motility assay showing swimming abilities after 48h on soft 0.3% agar plates. The WT strain and the *ΔfliC* non-motile flagellin mutant are positive and negative controls, respectively. **(b)** Quantitative assay for total protease activity on skimmed milk agar plates. Activity for three independent replicates is given as an arbitrary unit calculated: PrA = [π(rhalo)2-π(rcolo)2]/ [π(rcolo)2] with "r" corresponding to the radius measured for colonies and degradation halos. Statistical significance of differences in (b) was assessed with the Wilcoxon test (*: p-value ≤ 0.05). Horizontal bars show the mean of at least three biological replicates. **(c)** Differentially-expressed genes in the *ΔXC_3253* mutant and *ΔXC_3253::XC_3253* complemented strains identified by RNAseq. Genes were considered differentially-expressed if |log2FC| ≥ 1.5 and FDR ≤ 0.05.
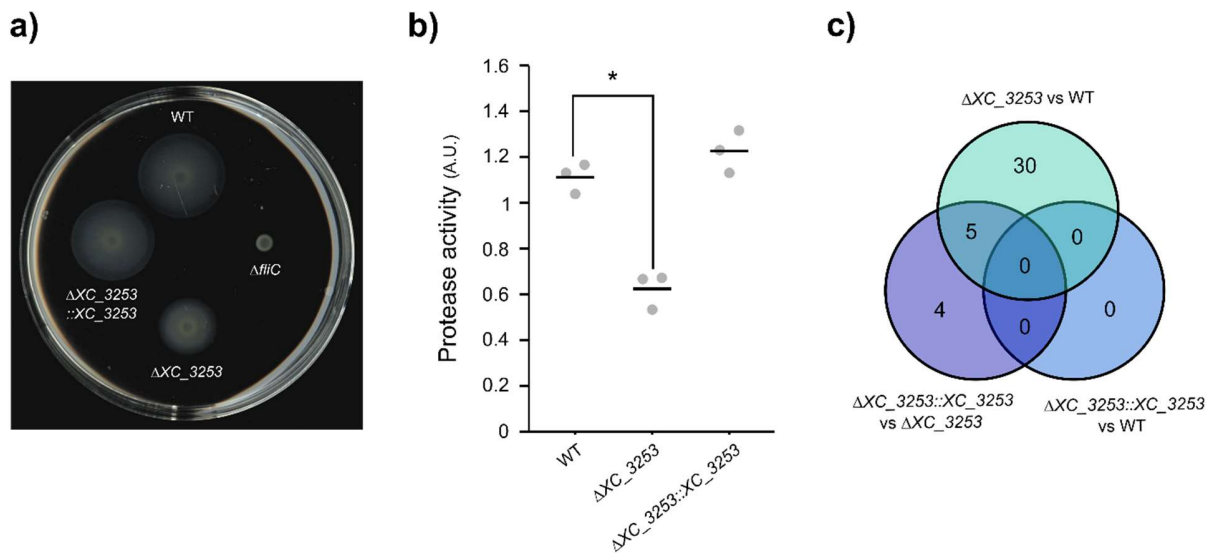
Fig. S6 Importance of XC_3388 for *Xanthomonas campestris pathovar campestris (Xcc)* metabolism. Assays were conducted on wild-type 8004 (WT), 8004 *ΔXC_3388* mutant strain (Δ*XC_3388*) and the complemented 8004 Δ*XC_3388::XC_3388* (Δ*XC_3388::XC_3388*). **(a)** Rough phenotype of the mutant strain Δ*XC_3388* observed on the MOKA plate. **(b)** Metabolic fingerprinting on Biolog plates. The carbon and nitrogen sources differentially metabolized between the Δ*XC_3388* (blue) and WT (green) strains are shown here. Two biological repetitions were performed for carbon sources (plates PM01 and PM02-A) and a third one for the assay of Δ*XC_3388* on nitrogen source (plate PM03). **(c)** Growth curves on several carbon sources (20 mM in minimal MME medium) or on MOKA rich medium confirm that XC_3388 plays a role in *Xcc* multiplication *in vitro*.

a) WT, ΔXC_3388::XC_3388, ΔXC_3388

b)

— ΔXC_3388
- - - ΔXC_3388
— WT
- - - WT

Nitrogen Source
N-Acetyl-D-glucosamine

Carbon Source

D-Mannose

Maltose

α-Hydroxyglutaric acid-γ-lactone

L-Glutamic acid

Hydroxy-L-Proline

L-Alaninamide

Dextrin

Glycogen

Arbutin

c)

MOKA rich medium
WT
ΔXC_3388
ΔXC_3388::XC_3388

MME + 20 mM Glutamate
WT
ΔXC_3388
ΔXC_3388::XC_3388

MME + 20 mM Maltose
WT
ΔXC_3388
ΔXC_3388::XC_3388

MME minimal medium
WT
ΔXC_3388
ΔXC_3388::XC_3388

MME + 20 mM Glucose
WT
ΔXC_3388
ΔXC_3388::XC_3388

MME + 20 mM Arbutin
WT
ΔXC_3388
ΔXC_3388::XC_3388
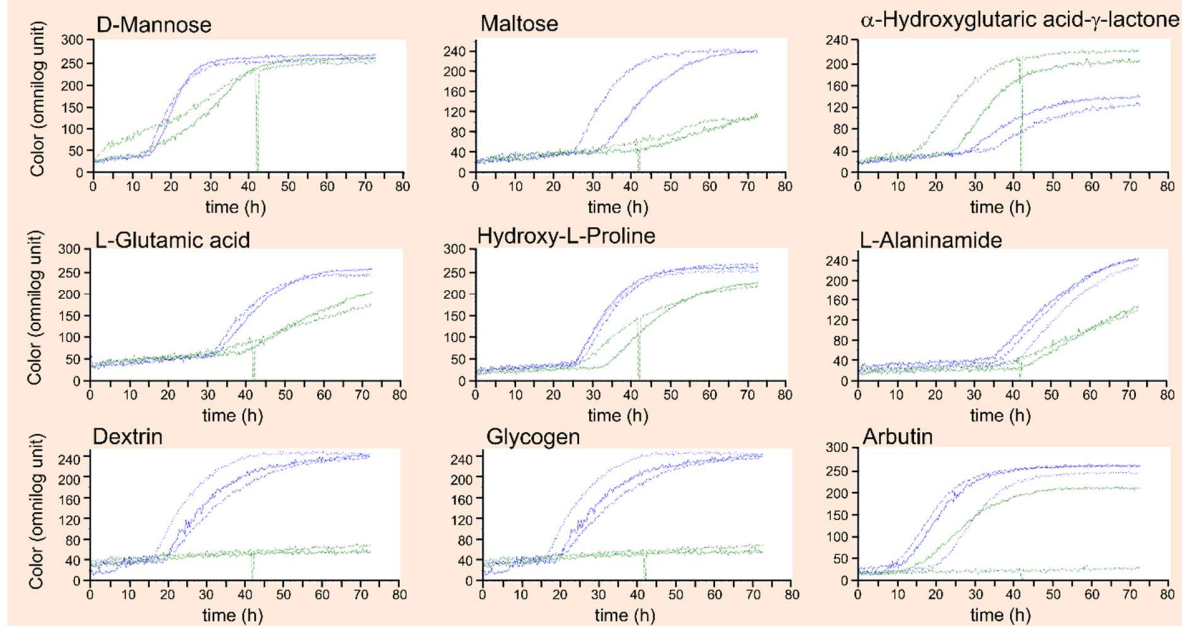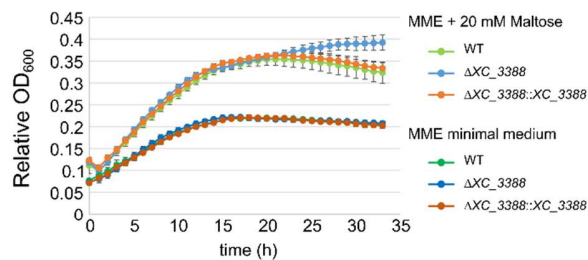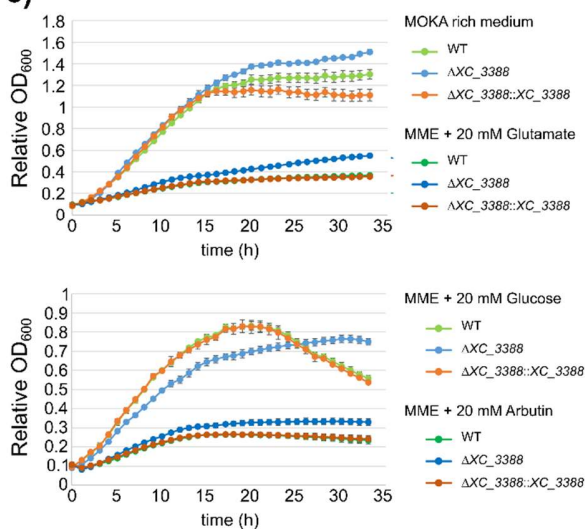
Table S1 Strains, plasmids and oligos used in this study.

Table S2 List of essential genes of *Xcc*. Essentiality of genes was determined as descibed in Price et al., 2018.
a. locus_tag corresponds to identification tag attributed when the new annotation (https://bbric.toulouse.inra.fr/reference/web/index.html?collection=reference&docid=7b2a96f 6e222f9599fceb186b2bcc2c1) ; NoA corresponds to CDS that were not annotated in Xcc 8004 annotation done by Qian et al. 2005 or by recent automatic reannotation done by NCBI.
b. ntLenNoOverlap correspond to the gene length without an overlap with another gene. **c.** number of different positions with a mapped transposon insertion.
d. total number of mapped transposon insertion reads.
e. number of different positions within the central part of the gene (10-90%) with a mapped transposon insertion.
f. total number of mapped transposon insertion reads within the central part of the gene (10-90%).
g. gene's percentage of GC nucleotides.
h. numer of TA sites in gene.
i. the density of insertion locations, scaled so that the median value is 1.
j. reads per nucleotide.
k. reads per nucleotide, normalized by GC content.

Table S3 Genes affected in fitness in each tested conditions. This supplementary table provides the complete list of genes affected in fitness of Xcc 8004 in all conditions assayed: Xylem sap, hydathodes at 6 and 3 dpi, *in vitro* and genes commonly affected in all plant conditions assayed.
a. locus_tag corresponds to identification tag attributed when the new annotation was performed on RNAseq libraries; NoA corresponds to CDS that were not annotated in Xcc 8004 annotation done by Qian et al. 2005 or by recent automatic reannotation done by NCBI.
b. mean of the fitness values obtained for 4 biological replicates.
c. standar deviation of the fitness values obtained for 4 biological replicates.
d. *t* score is computed based on the consistency of the strain fitness values for each gene (Wetmore et *al*., 2015).
e. Fitness value corresponding to the weighted average of the strain fitness for each biological replicate.

Table S4 This supplementary table provides the complete list of differentially expressed genes of *Xcc* 8004 in all conditions assayed. DEGs are defined by 1.5< logFC > 1.5 with a FDR < 0.05**.** Genomic information about annotations as well as Log of fold change and FDR are indicated.
a. locus_tag corresponds to identification tag attributed when the new annotation was performed on RNAseq libraries.
b. corresponds to annotation of Xcc 8004 done by Qian et al. 2005.
c. corresponds to the gene name used for fitness analysis.

Methods S1 Supplementary material and methods for Fitness screening in vitro; Carbon and nitrogen substrates phenotyping; Growth measurements in vitro.

Fitness screening *in vitro*

The RB-TnSeq library was grown in MOKA for 4 hours at 28°C with 200 rpm shaking and washed twice in MgCl$_2$ 1mM. We pelleted 10^9 *Xcc* cells by centrifugation at 13,000 x g for 5 min and stored them at -80°C to be used as T$_0$ samples. The *Xcc* inoculum was adjusted to OD$_{600}$ = 0.05 (5.10^7 cfu/mL) in 10 mL of either 0.22 μm-filtrated cauliflower xylem sap, MOKA rich medium, MME poor medium + 20 mM Glucose or MME + 20 mM Glucose + 0.15% DropOut -LTH. We incubated all cultures at 28°C under agitation at 200 rpm. Because the differences in fitness become more noticeable over time, the clarity of RB-TnSeq results depends on the number of generations. Therefore, we allowed bacterial growth for up to 6-7 generations in all media. However, except in MOKA rich medium, *Xcc* growth could not attain this number of generations before reaching the stationary phase. Therefore, we restarted cultures at the end of the exponential phase by centrifugation (6,000 x g, 5 min) and resuspension of the whole population in a larger volume of fresh medium at OD$_{600}$ = 0.05. At the end of growth, we pelleted 10^9 *Xcc* cells by centrifugation at 13,000 x g for 5 min and stored the pellets at -80°C until DNA extraction. The MME + 20 mM Glucose + 0.15% DropOut -LTH condition was used to control that genes involved in known biosynthesis pathways were affected in fitness as expected (Fig. S2B).

Carbon and nitrogen substrates phenotyping

Utilization of carbon and nitrogen sources by *Xcc* were assayed using the Biolog phenotype microarray plates PM1, PM2A and PM3B (PM; Biolog, Hayward CA) according to the manufacturer's protocol. *Xcc* cells were grown in rich MOKA medium overnight at 28°C. Cells were collected and resuspended in IF-0

(for PM1 and PM2A) or IF-0 20 mM glucose (for PM3B) supplemented with 1× Biolog Redox Dye Mix H, at initial OD_{600nm} = 0.08 for plates inoculation (100 µl/well). Plates were incubated for 72 h in the Omnilog device (Biolog, Hayward, CA) at 28°C. At least two independent replicates were performed. Results were analyzed using the Scilab software.

Growth measurements *in vitro*

In vitro growth curves were obtained by growing 200 µL of *Xcc* suspensions in 96-well flat-bottom microtiter plates (Greiner) within a FLUOStar Omega apparatus (BMG Labtech, Offenburg, Germany) at 28°C. After an overnight preculture in MOKA rich medium, cells were harvested by centrifugation at 9500 x g for four minutes, washed and resuspended in MME minimal medium. Bacterial suspensions inoculated at an optical density at 600 nm (OD_{600}) of 0.15 were prepared in MME or MME supplemented with either glucose, glutamate, maltose or arbutin at 20 mM and MOKA media. For each experiment, we performed four replicates coming from two independent growth. The microplates were shaken continuously at 700 rpm using the linear-shaking mode. Each experiment was repeated three times from which one representative experiment is shown.

Protein sequence conservation and structure predictions

Conservation of protein sequences was analyzed by BlastP using default parameters except for "Max target sequences" that was set to 5000 sequences. The analysis was performed first on all organisms and then excluding *Xanthomonadales*. For structure prediction, the protein sequence was submitted to the online structure prediction sofwares, AlphaFold2 (Jumper *et al.*, 2021; https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scrollTo= kOblAo-xetgx; default parameters) and RoseTTAfold (Baek *et al.*, 2021; https://robetta.bakerlab.org/; optional parameters RoseTTAFold and predict domains).

Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, *et al.* 2021. Accurate prediction of protein structures and interactions using a three-track neural network. *Science (New York, N.Y.)* **373**: 871–876.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.* 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**: 583–589.